Proc. R. Soc. Lond. B **227**, 249–265 (1986)
Printed in Great Britain

Seeing objects in motion

By D. C. Burr, J. Ross and M. C. Morrone
Department of Psychology, University of Western Australia, Nedlands, Western Australia 6009, *Australia*

(*Communicated by H. B. Barlow, F.R.S. – Received* 13 *September* 1985)

This paper reports estimates of the conjoint spatiotemporal tuning functions of the neural mechanisms of the human vision system which detect image motion. The functions were derived from measurements of the minimum contrast necessary to detect the direction of drift of a sinusoidal grating, in the presence of phase-reversed masking gratings of various spatial and temporal frequencies. A mask of similar spatial and temporal frequencies to the test grating reduces sensitivity considerably, whereas one differing greatly in spatial or temporal frequency has little or no effect. The results show that for test gratings drifting at 8 Hz, the tuning function is bandpass in both space and time, peaked at the temporal and spatial frequency (SF) of the test (SFs were 0.1, 1 or 5 c deg^{-1}; c represents cycles throughout). For a grating of 5 c deg^{-1} drifting at 0.3 Hz, the function is bandpass in space but lowpass in time. Fourier transform of the frequency results yields a function in space-time which we term the 'spatiotemporal receptive field'. For movement detectors (bandpass in space and time) the fields comprise alternating ridges of opposing polarity, elongated in space-time along the preferred velocity axis of the detector. We suggest that this organization explains how detectors analyse form and motion concurrently and accounts, at least in part, for a variety of perceptual phenomena, including summation, reduction of motion smear, metacontrast, stroboscopic motion and spatiotemporal interpolation.

Introduction

The eye gathers light over time and space so as to guarantee satisfactory signal:noise ratios in the face of photon fluctuations and internal noise. So long as images are stationary on the retina, temporal integration is no impediment to clarity of vision; if there is motion, a long integration time (like a long camera exposure) means smear and impaired resolution if we consider only mechanism capable of resolving static images. We suggest here, by considering a range of experimental evidence, that the visual system employs motion detectors with receptive fields extended over time and space to analyse both the form and the motion of moving objects. Detectors of this type readily explain many standard motion phenomena, such as cine-motion, reverse phi-motion and spatiotemporal interpolation.

At image speeds up to 16° s^{-1}, light from a moving spot is summated completely for about 100 ms (Burr 1981), during which time the spot traverses a distance of up to 1.6°. Yet summation for a stationary disc of light is complete for diameters below 0.5° (e.g. Barlow 1958). The clear implication is that, when objects move,

summating mechanisms of a different order are brought into play. At short exposures, a moving spot appears to be smeared as one would expect by analogy with a camera (Burr, 1980). However, at longer exposures the spot no longer appears smeared. These results suggest that the mechanisms responsible for the detection of objects in motion function to remove smear, provided that there is sufficient time for them to operate.

It is now well established that the labour of the visual system is divided between mechanisms (sometimes termed channels), each sensitive to a subrange of the visible spatial frequencies. Most previous work has concentrated on stationary stimuli. Recent measurements (Anderson & Burr 1985) show that detectors responsive to motion are also selective for spatial frequency, similarly dividing the spatial frequency range between them, down to much lower spatial frequencies than hitherto observed. Mechanisms sensitive to motion, even those which respond to components of very low spatial frequency, are equipped to give information about spatial structure by virtue of their spatial tuning.

A different pattern is obtained in the temporal frequency domain. There does not exist a large number of mechanisms each tuned to a different sub-range of temporal frequencies. Whatever the spatial frequency, only two types of temporal tuning appear: a bandpass mechanism with a centre frequency near 10 Hz, and a lowpass mechanism sensitive up to about 10 Hz (e.g. Kulikowski & Tolhurst, 1973; Burr, 1981; Watson & Robson, 1981; Anderson & Burr 1985). At low spatial frequencies it is the bandpass mechanism which is dominant, and at high spatial frequencies it is the lowpass. The mechanisms which analyse moving objects have a channel-like structure, but whereas the spatial frequency spectrum is divided into many overlapping ranges the temporal frequency spectrum is not.

In this study we seek to describe the *conjoint* spatiotemporal tuning function of motion-sensitive mechanisms and to explain how these mechanisms permit the simultaneous analysis of spatial form and of motion; we then go on to consider how the tuning may account for many of the phenomena of motion perception. We chose vertical sinusoidal gratings, drifting horizontally, and the psychophysical method of *masking* to measure the human tuning functions. The rationale behind this method is fully discussed by Graham (1980), Legge & Foley (1980) and others.

Some of the concepts to be presented here have been published previously in abstract form (Burr 1983, 1982*a*; Ross & Burr 1983; see also Burr & Ross 1985), and are similar to some later elaborations of these ideas (Whatson & Ahumada 1983, 1985; Adelson & Bergen, 1985).

Methods

(*a*) *Stimuli and procedure*

We measured the detectability of drifting vertical gratings displayed together with vertical 'mask' gratings caused to reverse contrast sinusoidally. For each set of measurements the test had a fixed spatial and temporal frequency (and hence a fixed drift velocity), while the mask varied over a wide range of both parameters. The contrast of the mask was variable for the first experiment, and fixed at 0.33 for the others. The tests used were 0.1 c deg^{-1}, 8 Hz (80° s^{-1}), 1 c deg^{-1}, 8 Hz (8° s^{-1}), 5 c deg^{-1}, 8 Hz (1.6° s^{-1}) and 5 c deg^{-1}, 0.35 Hz (0.07° s^{-1}). The

temporal frequencies of 8 Hz and near 0 Hz were chosen to probe the two temporal channels, and the spatial frequencies to extend over much of the visible range of spatial frequencies.

All waveforms were generated by computer (Cromemco Z-2D) and displayed on the face of an oscilloscope (Joyce Electronics) at 100 frames per second, 1000 lines per frame, at a mean luminance of 200 cd m^{-2}. Contrast of the test grating was varied under subject control by means of a digital attenuator driven by computer. Contrast thresholds for detection of the test were determined for the test alone and for the test in the presence of the counterphase masking grating. The quantitative measure of masking was defined as the ratio of the visibility threshold in the presence of the mask to that with no mask.

Thresholds were set by the observers using the method of adjustment, with the added precaution that the computer introduced a random attenuation offset on each trial to minimize response stereotyping. In all cases the criterion for threshold was the minimum contrast at which the direction of drift could be discerned. Two observers were used for all measurements, both with vision corrected to 6/6.

(*b*) *Data manipulation*

The data were first smoothed (on a linear scale) with a two-dimensional cubic polynomial interpolation function, then elevated to the power 1.4 to yield the amplitude function of the linear filter (see result section for rationale). This function was considered as the positive–positive quadrant of the two-dimensional frequency domain. The negative–negative quadrant was filled by symmetry. For the measurements with test grating drifting at 0.35 Hz (virtually stationary), the other two quadrants were also filled by symmetry, to reflect the lack of directionality tuning. All the results were back transformed by fast Fourier transform to yield spatiotemporal impulse response functions, which we term receptive fields.

As there were no measurements of phase, the phase spectrum was assumed. Linear phase was assumed for the results shown in figure 6, and minimum phase for the result of figure 7. For the assumption of linear phase advance, the imaginary component of the filter matrix was set to zero.

For the assumption of minimum phase we used the algorithm of Read & Treitel (1973). The phase was computed from the logarithm of the amplitude, by the following equation:

$$P(i, k) = -\mathrm{j}\ \mathrm{DFT}\ [\mathrm{sgn}(i, k) + \mathrm{bdy}(i, k)]\ \mathrm{IDFT}\ [\lg\ [A(i, k)] \qquad (1)$$

where i and k are the index of the matrix, $A(i, k)$ is the amplitude matrix, j is the imaginary unit, DFT and IDFT are the forward and inverse fast Fourier transforms, and sgn and bdy are two functions of the following equations:

$$\mathrm{sgn} = \begin{cases} 1 & 0 < i < \frac{1}{2}N \quad \text{and} \quad 0 < k < \frac{1}{2}N \\ -1 & \frac{1}{2}N < i < N \quad \text{and} \quad \frac{1}{2}N < k < N, \\ 0 & \text{elsewhere} \end{cases}$$

$$\mathrm{bdy} = \begin{cases} 1 & i = 0 \text{ and } 0 < k < \frac{1}{2}N \quad \text{or} \quad k = 0 \text{ and } 0 < i < \frac{1}{2}N \\ -1 & i = 0 \text{ and } \frac{1}{2}N < k < N \quad \text{or} \quad k = 0 \text{ and } \frac{1}{2}N < i < N \\ 0 & \text{elsewhere} \end{cases}$$

where N is the dimension of the matrix.

After the computation of the phase, the real and imaginary components of the spectra were calculated and the resulting matrix was back-transformed to obtain the impulse response function.

Results

(a) *Mask contrast*

The first stage of the study was to measure reduction in sensitivity as a function of mask contast for masks of different spatial and temporal frequencies. Each of the observers was required to adjust the contrast of the test until its direction of motion could just be discerned, in the presence of mask gratings varying from 0.06 to 0.7 contrast, and also with no mask. Both observers made adjustments for test spatial frequencies of 0.1, 1 and 5 c deg^{-1}, and temporal frequency of 8 Hz. Mask spatial frequencies varied considerably about those of the test.

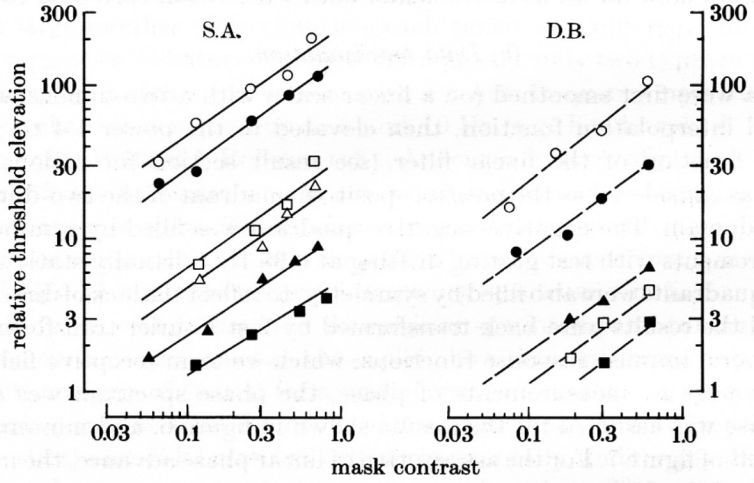

FIGURE 1. Elevation of the detectability threshold of the test grating as a function of the contrast of the mask. Threshold elevation is defined as the ratio of the detection threshold measured with the mask to that measured without it. Plotted here are a few representative results for S.A. at test frequencies 0.1 c deg^{-1}, 8 Hz, and D.B. at 5 c deg^{-1}, 8 Hz. The symbols refer to different spatial and temporal frequencies of the masks. For S.A. they are: ○, 0.1 c deg^{-1}, 8 Hz; ●, 0.1 c deg^{-1}, 24 Hz; □, 0.4 c deg^{-1}, 24 Hz; △, 0.025 c deg^{-1}, 8 Hz; ▲, 0.2 c deg^{-1}, 2.8 Hz; and ■, 0.1 c deg^{-1}, 0.7 Hz. For DB they are: ○, 5 c deg^{-1}, 8 Hz; ●, 2 c deg^{-1}, 8 Hz; ▲, 10 c deg^{-1}, 0.7 Hz; □, 20 c deg^{-1}, 4 Hz; and ■ 5 c deg^{-1}, 32 Hz. The lines, which pass through the data reasonably well, all have slope of 0.7.

Figure 1 plots representative results for the two observers: those of S.J.A. at 0.1 c deg^{-1} and D.C.B. at 1 c deg^{-1}. The ordinate shows the amount of reduction in sensitivity, calculated by dividing the threshold measured with the mask by that measured with no mask. For all test and mask frequencies the results, plotted on log–log axes, can be fairly well fitted with a line of slope 0.7 (those results not displayed followed a similar trend). In other words, irrespective of the spatial frequency of the test and of the spatial and temporal frequency of the mask, the reduction in sensitivity is proportional to the contrast of the mask raised to the

power 0.7. This result is similar to that reported by Legge & Foley (1980) for stationary stimuli, at the contrasts used here.

The results reported in figure 1 establish a relation between mask contrast and reduction in sensitivity, and show that the relation is invariant with the frequency parameters of both test and mask. This implies that the general shape of the masking curves should be relatively independent of the contrast of the mask. The relation between mask contrast and test threshold elevation will be used to infer frequency tuning functions from the masking data.

(*b*) *Masking as a function of temporal and spatial frequency*

With the contrast of the mask fixed at 0.33, reduction in sensitivity was measured for test gratings of spatial frequency 0.1, 1 and 5 c deg^{-1}, drift rate 8 Hz (velocities of 80, 8 and 1.6 deg/s), and also for 5 c deg^{-1}, 0.3 Hz (0.06° s^{-1}). For each test, the mask grating varied over 15 spatial and 11 temporal frequencies, giving 165 data points.

A few representative results for one observer (S.J.A.) are shown in figures 2 and

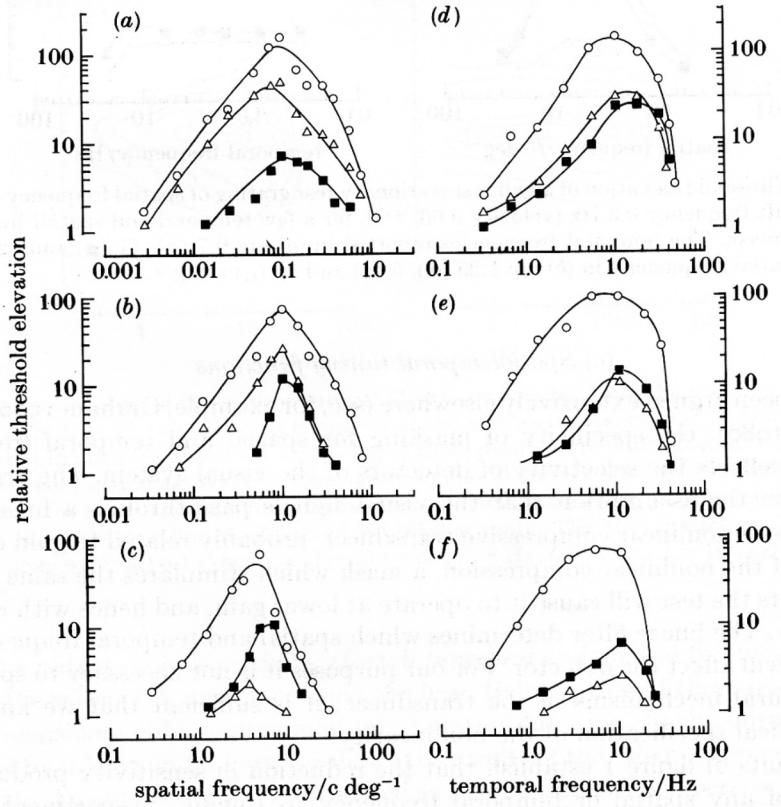

FIGURE 2. Threshold expressed as the ratio of thresholds measured with the mask to that measured without it, for a few representative spatial and temporal mask frequencies (for S.J.A.). The test gratings had spatial frequencies of 0.1 (*a*, *d*), 1 (*b*, *e*) and 5 (*c*, *f*) c deg^{-1}, and were caused to drift at 8 Hz. For (*a*) the temporal frequencies were 2.8 (■), 11 (△) and 24 (○) Hz, and for (*b*) and (*c*) 2.8 (■), 11 (△) and 32 (○) Hz. For (*d*)–(*f*) the spatial frequencies were 0.25 *f*, (■) *f* (△) and 4 *f* (○), where *f* is the spatial frequency of the test.

3 (those of D.C.B. were very similar). Figure 2 shows results for test gratings of (*a*) 0.1, (*b*) 1 and (*c*) 5 c deg^{-1}, drifted at 8 Hz, and figure 3 shows the results for a test grating of 0.3 Hz, 5 c deg^{-1}. For each test grating, there is maximum masking (reduction in sensitivity) when the spatial and temporal frequency of the mask coincide with that of the test. The amount of masking declines as the mask temporal and/or spatial frequency moves away from that of the test. The functions shown in figure 2 can all be described as bandpass, showing attenuation at both higher and lower frequencies. That of figure 3, however, where the test grating is virtually stationary, is bandpass in space, but lowpass in time, showing no low temporal frequency attenuation.

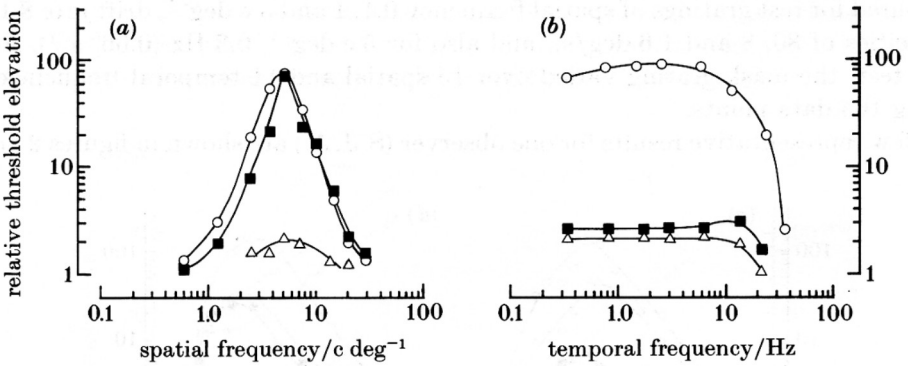

FIGURE 3. Threshold elevation of an almost stationary test grating of spatial frequency 5 c deg^{-1} and drift frequency 0.3 Hz (velocity 0.06° s^{-1}), for a few temporal and spatial frequencies of the mask. The temporal frequencies reported in (*a*) are 0.7 (○), 11 (■) and 32 (△) Hz. The spatial frequencies in (*b*) are 1.25 (■), 5 (○) and 20 (△) c deg^{-1}.

(*c*) *Spatiotemporal tuning functions*

As has been argued extensively elsewhere (see, for example, Graham 1980; Legge & Foley 1980), the specificity of masking for spatial and temporal frequency probably reflects the selectivity of detectors of the visual system. The argument is based on the assumption that the visual signals pass through a linear filter followed by a nonlinear compressive transducer, probably related to gain control. Because of the nonlinear compression, a mask which stimulates the same neuron that detects the test will cause it to operate at lower gain, and hence with reduced sensitivity. The linear filter determines which spatial and temporal frequencies of the mask will affect the detector. For our purposes it is not necessary to speculate on the neural mechanisms of the transducer; it is sufficient that we know the mathematical specification of the nonlinearity.

The results of figure 1 establish that the reduction in sensitivity produced by a mask (of any spatial or temporal frequency) is roughly proportional to the contrast of the mask raised to the power 0.7. Alternatively, mask contrast is proportional to the reduction in sensitivity raised to the power 1.4. From this relation one can calculate the relative effective contrast of a mask at the stage of the nonlinear transformer, after it has passed through the linear filter, simply by

raising the measured threshold elevation to the power 1.4. The relative effective contrast of the mask at the nonlinear transform stage reflects the gain of the linear filter for the various spatial and temporal frequencies.

Figure 4 shows contour plots of the spatiotemporal tuning functions for one observer (S.J.A.) for the four velocities of the test grating. They are calculated by

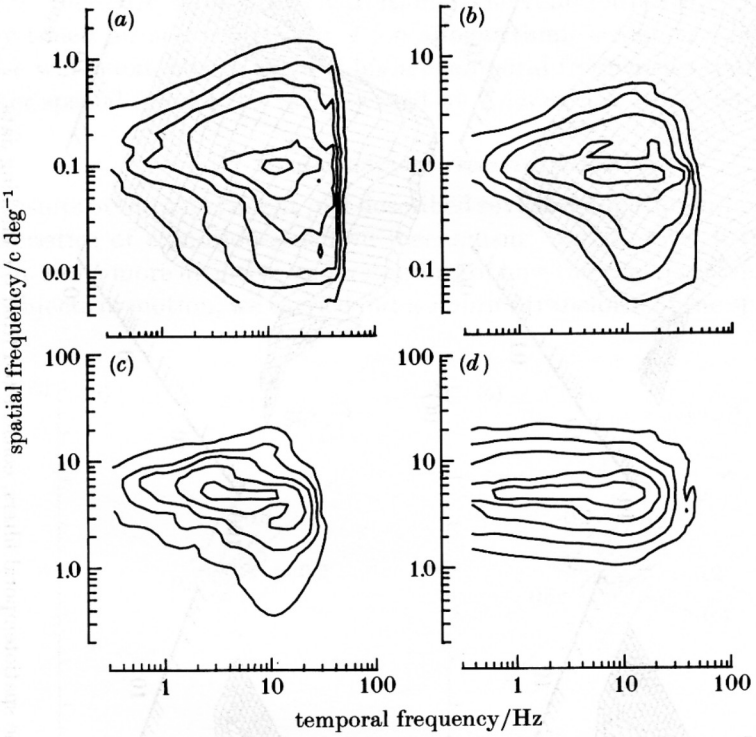

FIGURE 4. Contour plots of the spatiotemporal filters, constructed from the masking results of observer S.J.A. The elevation in threshold of the test was measured for 11 temporal and 15 spatial frequencies of the mask. The log of these results were multiplied by 1.4 (see text for rationale) and smoothed over the spatio-temporal surface with a cubic interpolation function. The results were then normalized to a maximum gain of 1, and plotted as contour maps. Each contour line is separated by 0.5 log units. The test spatial and temporal frequencies were: (*a*) 0.1 c deg^{-1}, 8 Hz; (*b*) 1 c deg^{-1}, 8 Hz; (*c*) 5 c deg^{-1}, 8 Hz; (*d*) 5 c deg^{-1}, 0.3 Hz.

raising the reduction in sensitivity at each temporal and spatial frequency of the mask to the power 1.4, and smoothing the base 10 logarithm of these results with a two-dimensional cubic polynomial interpolation function. Each contour line is separated by 0.5 log units. Figure 5 shows the results of the other observer, D.C.B., analysed in the same way, and represented as a three-dimensional perspective plot.

The tuning functions obtained with test gratings drifting at 8 Hz (figures 4*a–c* and 5*a–c*) are all bandpass in both space and time: they are peaked at the spatial and temporal frequency of the test, and roll off as the temporal or spatial frequency increases or decreases. The functions obtained with the 0.3 Hz (effectively

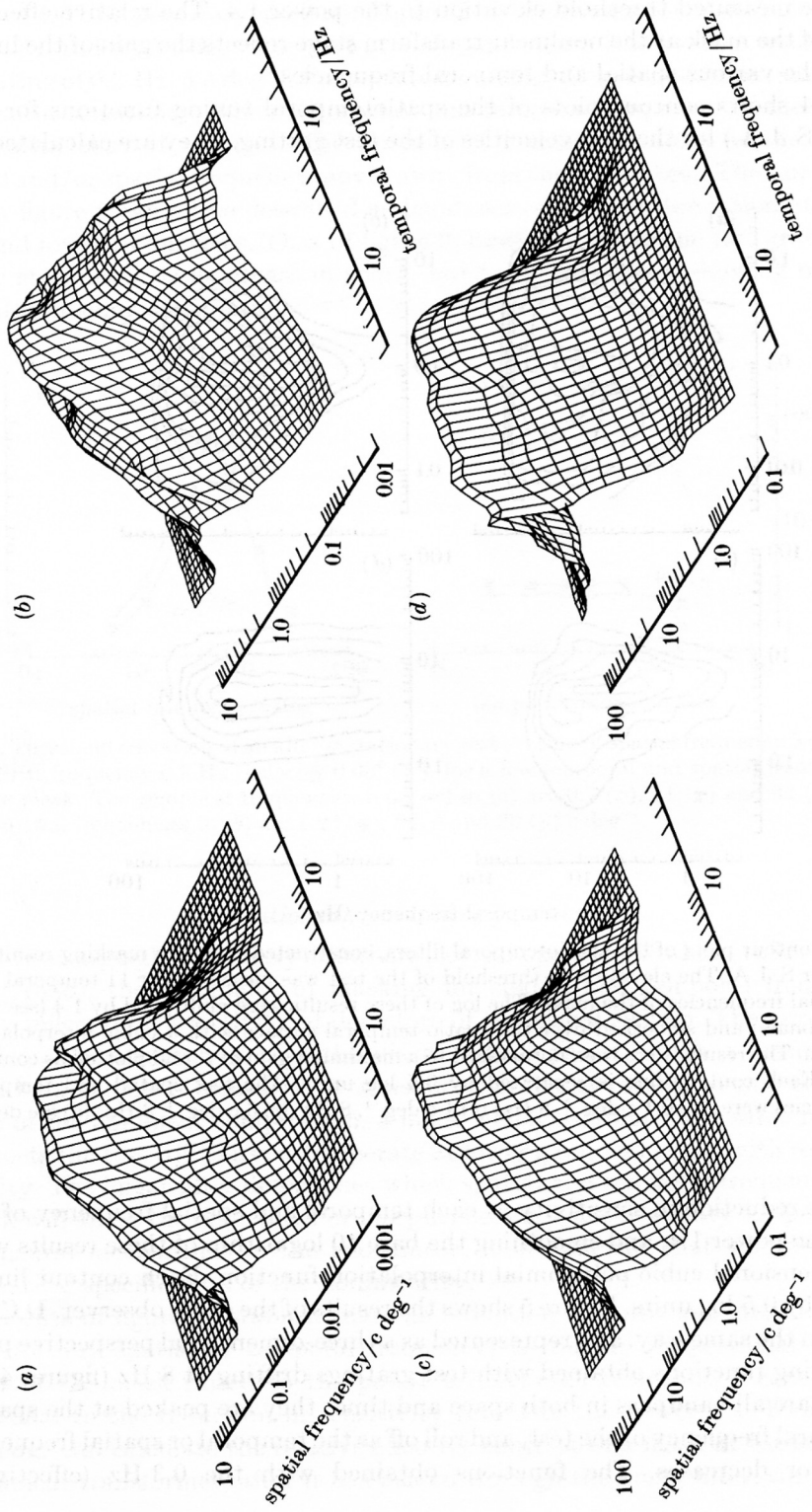

FIGURE 5. Perspective plots of the spatiotemporal filters, constructed from the masking data of observer D.C.B., following the procedure outlined in figure 4.

stationary) grating (figures 4*d* and 5*d*) are bandpass in space but not in time. In time they are lowpass, having equal gain at all temporal frequencies up to about 10 Hz. This is further support for the existence of two types of detectors, motion dependent and motion independent, as proposed by Tolhurst (1973).

The motion dependent detectors preferring different spatial frequencies (and velocities) have similar frequency response profiles (when plotted on logarithmic axes) but there are some minor variations. The function for 0.1 c deg^{-1} is more broadly tuned for spatial frequency (on a logarithmic scale), and has a temporal response which extends to slightly higher temporal frequencies than those tuned to higher spatial frequencies (as observed by Anderson & Burr 1985*a*).

(*d*) *Spatiotemporal receptive fields*

The results of the experiment just described reveal the spatiotemporal frequency characteristics of a motion-sensitive mechanism; but they do little to aid our intuition. For a more intuitive understanding of how the visual system will respond to real objects in motion, we carried out a Fourier transform of the spatiotemporal

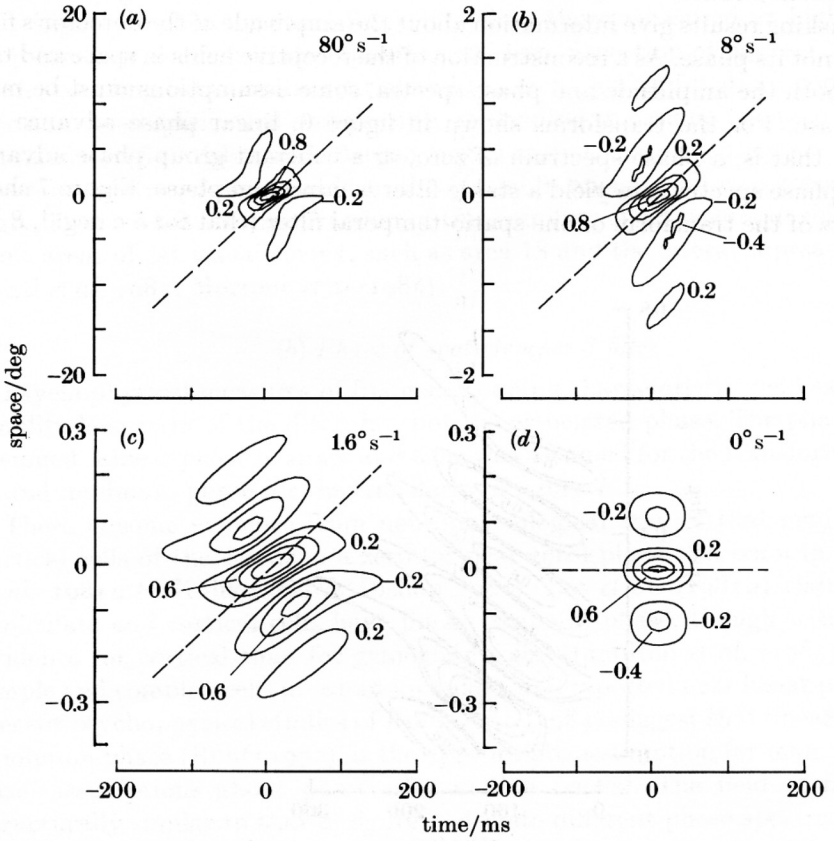

FIGURE 6. Receptive fields calculated by Fourier transform of the spatiotemporal filters of figure 4 (observer S. J. A.). Zero phase was assumed for the transform. The broken lines show the velocity of the test grating.

tuning functions of figure 4. This yields spatiotemporal weighting functions, or spatiotemporal 'receptive fields' (following the nomenclature of neurophysiology), shown in figure 6. One axis represents the single spatial dimension along which the test grating was free to move, the other represents time.

The form of all the spatiotemporal weighting functions for detectors tuned to 8 Hz (figure 6*a–c*) is basically similar. They are elongated, and oriented in space–time, at an orientation which corresponds to the drift velocity of the test grating. Each field is made up of several subregions of alternating polarity, produced by the modulation in space and time. This modulation follows a similar peridocity to the spatiotemporal modulation of the test grading. There are some variations in the shape of the fields of different spatial frequency preference: the higher spatial frequency detector has a field with more ripples than that at lower frequency, reflecting its tighter tuning for spatial frequency.

The receptive field for the motion independent detector (figure 6*d*) is quite different from the others. It is not oriented in space–time, but has a preferred velocity of 0° s^{-1}. Along the spatial axis it has subregions of alternating polarity, but along the temporal axis it does not. For any spatial location, the temporal profile is monophasic.

The masking results give information about the amplitude of the detector's filter function, not its phase. As a reconstruction of the receptive fields in space and time requires both the amplitude and phase spectra, some assumptions must be made about phase. For the transforms shown in figure 6, linear phase advance was assumed; that is, a phase spectrum of zero, or a constant group phase advance. Another phase spectrum to yield a stable filter is minimum phase. Figure 7 shows the results of the transform of one spatio-temporal filter (that for 5 c deg^{-1}, 8 Hz)

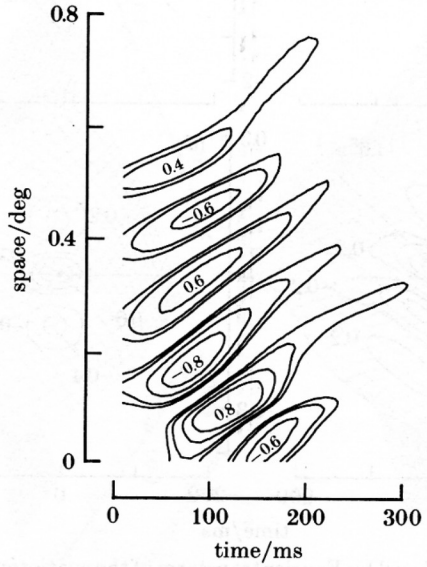

FIGURE 7. Receptive field calculated by Fourier transform of figure 4*c* assuming minimum phase in space and time (see Methods section for details).

assuming minimum phase (see Methods for mathematical details). Despite the fact that the receptive field does not have energy in negative space or negative time, the field is very similar to that which assumes linear phase (figure 5*c*). It is elongated in space-time, has a similar spread in space and time, and a similar modulation of alternating polarity.

Discussion of results

(*a*) *Spatiotemporal coupling*

The results of this experiment yield an estimate of the conjoint spatiotemporal tuning characteristics of human motion detectors. Close inspection of figures 2 and 4 reveals a small but consistent effect of spatiotemporal coupling. The spatial frequency functions measured with various temporal frequencies (left-hand curves of figure 2) do not all peak at exactly the same spatial frequency: those measured at low temporal frequencies peak at a slightly higher spatial frequency than those measured at higher temporal frequencies. Similarly, the temporal frequency functions (right-hand curves of figure 2) measured at low spatial frequencies peak at a slightly higher temporal frequency than those measured at higher spatial frequencies. The effect is also apparent on close inspection of figure 4.

The coupling between the spatial and temporal tuning functions confirms that it is an oversimplification and possibly misleading to assume (as have Watson & Ahumada (1985), Adelson & Bergen (1985) and others) that they are separable. The temporal response depends on the spatial frequency and vice-versa. Interestingly, a somewhat stronger version of this coupling effect occurs in single cells of some areas of cat visual cortex, such as area 18 and the lateral suprasylvian area (Bisti *et al.* 1985; Morrone *et al.* 1985).

(*b*) *Phase of spatiotemporal filter*

Psychophysical measures of frequency tuning characteristic yield estimate the amplitude spectra of the filter, but not the associated phase. The phase must be assumed. Linear phase in space and time was assumed for the transforms of figure 6, and minimum phase for the transform of figure 7.

There is some evidence from neurophysiological studies that geniculate and cortical cells of the cat have a zero (or near zero) phase spectrum in space (Lee *et al.* 1981*a*,*b*; Kulikowski & Bishop 1981). Lee *et al.* (1981*a*) claim that cat geniculate and cortical cells have linear temporal phase, though with less clear evidence for cortical than for geniculate cells. Hamilton *et al.* (1985) examined simple and complex cells in cat and monkey, and reported near linear phase maps. Recent psychophysical studies of Roufs *et al.* (1984) suggest that linear phase, not minimum phase (Roufs 1972), is the appropriate assumption for man. But in any case, assumptions about phase may not be crucial. The field of figure 6*c* is structurally similar to that of figure 7, despite different phase spectra.

The recent electrophysiological studies of Emerson *et al.* (1985) also support our results. They measured the response of complex cortical neurons of cat to pairs of flashed bars separated in space and time. Analysis of the suppression or

enhancement of the cell response due to the interaction of the two bars yields a spatiotemporal response profile very similar to the spatiotemporal receptive fields shown in figure 6*a*–*c*. They have ridges of alternating sign oriented in time–space. The spatial extent is variable (depending on the size of the cell receptive field, as determined by standard techniques), but all extend about 100 ms in time. It is encouraging that a technique which differs so widely from ours should yield results so similar. The spatiotemporal receptive fields which Hamilton *et al.* (1985) construct from direct amplitude and phase measurements of cat and monkey cortical cells also bear a striking similarity to our receptive fields.

General discussion

Image motion inevitably means loss of some information because components of an image with a temporal frequency above 30–40 Hz are filtered out. Rapid motion shifts the contrast sensitivity function well down the spatial frequency axis (Burr & Ross 1982), leaving visible only components of low spatial frequency. Fine detail of form is therefore lost. The receptive fields reported here, oriented in space–time, explain why it is that there is no corresponding loss in information about relative position. For maximal response, each field requires form on the scale to which it is tuned to move along a precise trajectory. It will register departures from the optimal trajectory, and thus deviation in position over time, regardless of the spatial frequency to which it is tuned. The concept of receptive fields oriented in space–time also helps to explain velocity selectivity and many of the known but still puzzling phenomena of motion perception (see also Burr 1983, 1984*a*; Watson & Ahumada 1985; Adelson & Bergen 1985).

(*a*) *Velocity selectivity*

Figure 8 is a schematized contour plot of one of the spatiotemporal receptive fields (the one tuned to 1 c deg^{-1}), taken from figure 4*b*. The field is elongated along, and symmetric about, a principal axis lying parallel neither to the spatial nor to the temporal axis. This axis is an axis of velocity, the sign and magnitude of which is given by the slope. In this case, the axis describes a velocity of 8° s^{-1} in a positive direction. Axes with negative slope describe negative velocities. The field has a central ON ridge flanked by OFF regions, giving it a selectivity to velocity. A spot moving at constant velocity is described by a line oriented in space and time. If the orientation matches that of the receptive field then it will stimulate only the ON (or only the OFF) regions of the field. If, however, the velocity varies considerably from 8° s^{-1}, it will stimulate the antagonistic flanks, reducing the absolute average value of the response. Motion in the wrong direction will stimulate ON and OFF regions equally, resulting in a null response.

Detection of velocity is probably achieved by more than one motion detector's cooperating. Figure 6 shows the receptive fields of several classes of motion detectors of differing velocity preference, and also of a motion-independent detector. These fields will all have symmetric partners tuned to velocities of opposite direction. The final analysis of motion and of the spatial structure of the object in motion will obviously rely on interactions between all classes of detectors,

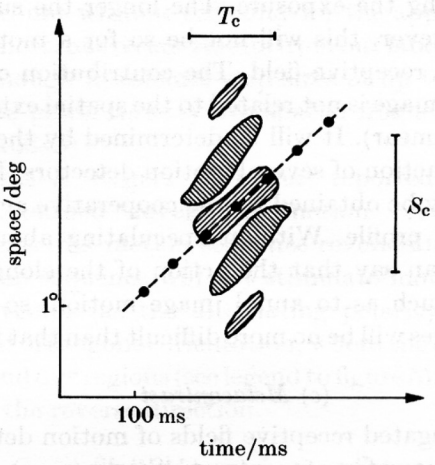

FIGURE 8. A schematized version of figure 6*b*, showing how the receptive field is orientated in space–time. The field is made up of separate opponent regions (indicated by orthogonal cross-hatching) which, following the convention established in electrophysiology, we term ON and OFF. The central ON region responds positively to light (positive contrast) and negatively to dark; the flanking OFF regions positively to dark (negative contrast) and negatively to light. Linear summation is assumed. The orientation (velocity) which is optimal for the motion detector (8° s^{-1}) is shown passing through the receptive field. Images of this velocity will be favoured, causing maximum possible stimulation of the ON or OFF regions. Velocities differing markedly from this optimum will be discriminated against, as they will stimulate both ON and OFF ridges in the field. The time (T_c) and space (S_c) for which the detector summates an image on the optimal trajectory are also shown. The region of spatial summation (S_c) also represents the extent of smear expected from the camera model. The heavy dots on the trajectory illustrate stroboscopic motion at 24 frames per second (standard cinema rate).

including the inhibitory interactions which occur between detectors of opposite direction preference (Barlow & Levick 1965; Dean *et al.* 1980; Stromeyer *et al.* 1983). We do not propose here a detailed model of velocity coding but suggest that the reader consult the works of Adelson & Bergen (1985), Watson & Ahumada (1985) and particularly Santen & Sperling (1985) for detailed and elegant models of velocity perception (see also Nakayama 1985 for review). In this discussion, by considering the response of just one class of detectors, we attempt to provide some insight into how the individual detectors may code the spatial structure of objects in motion.

(*b*) *Summation and motion smear*

The temporal and spatial summation of a detector is given by the extension of its field in time and space. Figure 8 suggests that for the particular detector measured, temporal summation (T_c) should be about 100 ms, and spatial summation (S_c) about 1°. Both estimates are in fair agreement with measured temporal and spatial summation for moving spots as well as for moving gratings (Burr 1981).

Fields like those we propose explain why objects in motion may be seen without blur (Burr 1980). When a camera is left open a moving image smears the region

of film it traverses during the exposure. The longer the summation period, the greater the smear. However, this will not be so for a motion detector with its oriented spatiotemporal receptive field. The contribution of the detector to the spatial definition of the image is not related to the spatial extent of the field (which corresponds to camera smear). It will be determined by the width of the central region, and by the interaction of several motion detectors. In principle definition as precise as desired may be obtained by the cooperative action of many fields of different cross-sectional profile. Without speculating about the way in which detectors combine we can say that the action of the elongated spatiotemporal receptive field will be such as to annul image motion, so that the problem of analysis for moving images will be no more difficult than that for stationary images.

(*c*) *Metacontrast*

The action of the elongated receptive fields of motion detectors may also help to explain the visual illusion of 'metacontrast' (Stigler 1910). If a stimulus is briefly displayed and followed after a suitable interval (about 50 ms) by two brief flanking stimuli, the observer sees the flanking stimuli but not the central one. Most explanations of this phenomenon presume that the suppression of the central stimulus results from some form of inhibition initiated by the flanking stimuli (see, for example, Weisstein 1968).

Our explanation requires no such inhibition. The successive presentation of stimuli at different spatial locations will stimulate receptive fields oriented in space-time. These fields do not report the separate locations of the successively presented stimuli, but the *Gestalt* of a single bar in motion. The two separate presentations become perceptually merged. Evidence supporting this notion has been provided by Burr (1984*b*), who showed the two component metacontrast stimuli summate at threshold, suggesting merger of energy. The explanation, pursued in more detail by Burr (1984*b*), is not yet complete, but it does suggest that the action of the individual motion detectors may form the basis for the phenomenon of metacontrast.

(*d*) *Apparent motion*

Motion need not be continuous to be seen as smooth. Cinematic and other stroboscopic sequences give rise to the appearance of smooth motion. Stroboscopic motion is sampled motion, indicated by the dots of figure 8. Provided that the velocity (orientation of the line joining the dots) is correct and that the dots are not too widely spaced (sufficiently high strobe rate), the detector will respond, and its response will be no different from its response to smooth motion. That the target does not appear at every instant or at every point within the range it traverses is of no account. If the strobe rate is too slow, motion-independent detectors as well as motion detectors of the wrong velocity preference will respond to the spurious temporal components introduced by sampling, producing jerky motion (Watson *et al.* 1983; Burr *et al.* 1985*a*,*b*).

Braddick (1974; 1980) has distinguished two types of apparent motion, one limited to a short range (15 min) and operating at a low level of processing, the other not so limited and operating at a higher level. Receptive fields of the kind

we propose would account without difficulty for the appearance of motion when shifts in position are short, but recent results from our laboratory suggest that some fields may be large enough to tolerate step sizes of up to eight degrees provided stimuli have adequate power at very low spatial frequencies (Burr *et al.* 1985*a*, *b*; Anderson & Burr 1985*b*).

Anstis (1970; Anstis & Rogers 1975) has reported an interesting case of stroboscopic motion, termed 'reverse phi-motion': reversing the contrast on successive presentations of a motion sequence reverses the apparent direction of motion. Such a motion sequence will not stimulate motion fields aligned to the space-time velocity vector (as the alternating polarity will average to a null response), but those of orthogonal orientation, where the alternating contrast falls appropriately on ON and OFF regions (see legend to figure 5). Fields of this orientation will signal motion in the reverse direction.

(*e*) *Spatiotemporal interpolation*

A strong test for the explanatory power of oriented spatiotemporal receptive fields is provided by an effect observed independently in Morgan's laboratory (Morgan & Thompson 1975; Morgan 1976) and ours (Burr 1975; Ross & Hogben 1975; Burr & Ross 1979): spatial offset induced by temporal delay. The initial work was with stereopsis, but the same effect holds for vernier offset (Burr 1979*a*). The best way to understand the experiment is to imagine viewing a bar with a vernier gap moving behind a picket fence. If the gaps between the pickets are small enough, the upper and lower segments will never be simultaneously in view, but will appear at the same location (fixed by the picket gap) at slightly different times. The temporal offset in the display is sufficient to create an impression of spatial offset.

Our explanation for this phenomenon is simple, and qualitatively similar to several previously advanced explanations (Burr 1979*b*; Barlow 1979, 1981; Fahle & Poggio 1981; Morgan & Watt 1983). The structure of the fields in space and time smooths a sampled motion sequence to make it effectively continuous. This operation is equivalent to interpolation of the motion sequence; thus the upper and lower bars are seen continuously as misaligned. Another way of considering the situation is that the receptive field is oriented obliquely to both the time and the space axes, so information in either dimension should affect the field's response. Experiments have shown that under a wide range of conditions, vernier misalignments introduced by temporal offset are detected with the same order of precision as those introduced by spatial offset (Burr 1979*a*). Furthermore, Swindale & Cynader (personal communication) have preliminary evidence that single cells of cat cortex signal a vernier break induced by temporal offset as they do spatially induced verniers (Swindale & Cynader 1985).

We have ignored one spatial dimension of the visual image to simplify the task of describing the spatiotemporal configuration of motion fields. Obviously, detectors must be tuned in both spatial dimensions, as well as in time. We have also ignored the problem of cooperation between detectors with different fields. Our results suggest that all motion fields have much the same stretch in time (about 100 ms) but a spatial extent increasing with velocity. There are also detectors broadly tuned about zero velocity. We can suspect that a variety of motion detectors of different

tuning will combine, as do motion-independent detectors, to give a full description of objects in motion. The receptive fields of the motion mechanisms of the visual system are structured so as to resonate to image motion, and by the pattern of units which respond, to specify both form and velocity. They do not freeze motion, but record its passage.

D. B. was supported by an Australian N.H. & M.R.C. research grant, and M.C.M. by a Fellowship from the Australian Department of Science. M.C.M. is on leave from the Scuola Normale Superiore, Pisa.

References

Adelson, E. H. & Bergen, J. R. 1985 Spatio-temporal energy models for the perception of motion. *J. opt. Soc. Am.* **A2**, 284–299.

Anderson, S. J. & Burr, D. C. 1985 Spatial and temporal selectivity of the human motion detection system. *Vision Res.* **25**, 1147–1154.

Anderson, S. J. & Burr, D. C. 1985*b* Receptive field sizes of human motion detectors. *Perception* **14**, A19.

Anstis, S. M. 1970 Phi movement as a subtractive process. *Vision Res.* **10**, 1411–1430.

Anstis, S. M. & Rogers, B. J. 1975 Illusory reversal of visual depth and movement during changes in contrast. *Vision Res.* **15**, 957–961.

Barlow, H. B. 1958 Temporal and spatial summation in human vision at different background intensities. *J. Physiol., Lond.* **141**, 337–350.

Barlow, H. B. 1979 Reconstruction of the visual image in space and time. *Nature, Lond.* **279**, 189–190.

Barlow, H. B. 1981 The Ferrier Lecture, 1980. Critical limiting factors in the design of the eye and visual cortex. *Proc. R. Soc. Lond.* B**212**, 1–34.

Barlow, H. B. & Levick, W. R. 1965 The mechanism of directionally selective units in rabbit's retina. II: Directionally selective units. *J. Physiol., Lond.* **178**, 447–504.

Bisti, S., Carmignoto, G., Galli, L. & Maffei, L. 1985 Spatial frequency characteristics of neurones of area 18 in the cat: dependence on the velocity of the visual stimulus. *J. Physiol., Lond.* **359**, 259–268.

Braddick, O. 1974 A short-range process in apparent motion. *Vision Res.* **14**, 519–527.

Braddick, O. 1980 Low-level and high-level processes in apparent motion. *Phil. Trans. R. Soc. Lond.* B**290**, 137–151.

Burr, D. C. 1975 A second binocular depth of perception system. *Thesis, University of Western Australia.*

Burr, D. C. 1979*a* Acuity for apparent vernier offset. *Vision Res.* **19**, 835–837.

Burr, D. C. 1979*b* On seeing objects in motion. *Ph.D. thesis, University of Cambridge.*

Burr, D. C. 1980 Motion smear. *Nature, Lond.* **284**, 164–165.

Burr, D. C. 1981 Temporal summation of moving images by the human visual system. *Proc. R. Soc. Lond.* B**211**, 321–339.

Burr, D. C. 1984*a* Spatial and temporal selectivity of the human visual system. *Neurosci. Lett. Suppl.* **15**, S10.

Burr, D. C. 1984*b* Summation of target and mask metacontrast stimuli. *Perception* **13**, 183–192.

Burr, D. C. & Ross, J. 1979 How does binocular delay give information about depth? *Vision Res.* **19**, 523–532.

Burr, D. C. & Ross, J. 1982 Contrast sensitivity at high velocities. *Vision Res.* **23**, 3567–3569.

Burr, D. C. & Ross, J. 1985 Visual analysis during motion. In *Vision, brain and cooperative computation* (ed. M. A. Arbib & A. R. Hanson) M.I.T. Press. (In the press.)

Burr, D. C., Ross, J. & Morrone, M. C. 1985*a* Local regulation of luminance gain. *Vision Res.* **25**, 717–728.

Burr, D. C., Ross, J. & Morrone, M. C. 1985*b* Smooth and sampled motion. *Vision Res.* (In the press.)

Dean, A. F., Hess, R. F. & Tolhurst, D. J. 1980 Divisive inhibition involved in direction selectivity. *J. Physiol. Lond.* **308**, 84.

Emerson, R. C., Citron, M. C. & Vaughn, W. J. 1985 Cortical responses to motion depend on nonlinear spatiotemporal interactions. *Invest. Ophthal. Sup.* **26**, 7.

Fahle, M. & Poggio, T. 1981 Visual hyperacuity: Spatio-temporal interpolation in human vision. *Proc. R. Soc. Lond.* B**213**, 451–477.

Hamilton, D. B., Albrecht, D. G. & Geisler, W. S. 1985 Receptive field organization of visual cortical neurons in monkey and cat: The phase and amplitude transfer functions. *Invest. Ophthal. Sci.* **26** sup, 265.

Kulikowski, J. J. & Bishop, P. O. 1981 Linear analysis of the responses of simple cells in the cat visual cortex. *Expl. Brain Res.* **44**, 386–400.

Kulikowski, J. J. & Tolhurst, D. J. 1973 Psychophysical evidence for sustained and transient neurones in the human visual system. *J. Physiol., Lond.* **232**, 149–162.

Lee, B. B., Elephandt, A. & Virsu, V. 1981*a* Phase of responses to moving sinusoidal gratings in cells of cat retina and lateral geniculate nucleus. *J. Neurophysiol.* **45**, 807–817.

Lee, B. B., Elephandt, A. & Virsu, V. 1981*b* Phase of responses to sinusoidal gratings of simple cells in cat striate cortex. *J. Neurophysiol.* **45**, 818–828.

Legge, G. E. & Foley, J. M. 1980 Contrast masking in human vision. *J. opt. Soc. Am.* **70**, 1458–1471.

Morgan, M. J. 1976 Pulfrich effect and the filling in of apparent motion. *Perception* **5**, 187–195.

Morgan, M. J. 1980 Analogue models of motion perception *Phil. Trans. R. Soc. Lond.* B**290**, 117–135.

Morgan, M. J. & Thompson, P. 1975 Apparent motion and the Pulfrich effect. *Perception* **4**, 3–18.

Morgan, M. J. & Watt, R. J. 1983 On the failure of spatiotemporal interpolation: a filtering model. *Vision Res.* **23**, 997–1004.

Morrone, M. C., Di Stefano, M. & Burr, D. 1985 Spatial and temporal selectivity of neurones of the lateral suprasylvian gyrus of the cat. *Neuroscience Letters Suppl.* **22**, 5296.

Nakyama, K. 1985 Biological image motion processing: a review. *Vision Res.* **25**, 625–660.

Read, R. R. & Treitel, S. 1973 The stabilization of two-dimensional recursive filters via the discrete Hilbert transform. *IEEE Trans. Geosci. Electron* **GE-11**, 153–207.

Ross, J. & Hogben, J. H. 1975 The Pulfrich effect and short-term memory in stereopsis. *Vision Res.* **15**, 1289–1290.

Roufs, J. A. J. 1972 Dynamic properties of vision. II. Theoretical relationships between flicker and flash thresholds. *Vision Res.* **12**, 279–292.

Roufs, J. A. J., Piceni, H. A. L. & Pellegrino van Stuyvenburg, J. A. 1984 Phase and gain of the visual transient system. In *Institute for Perception Research Annual Report*, pp. 49–56.

Santen, J. P. H. Van & Sperling, G. 1985 Elaborated Reichardt detectors. *J. opt. Soc. Am.* **A2**, 300–321.

Stigler, R. 1910 Chonophotische Studien über den Umgebungskontrast. *Pflügers Arch. ges. Physiol.* **134**, 465–435.

Stromeyer, C. F., Kronauer, R. E., Madsen, J. C. & Klein, S. A. 1984 Opponent movement mechanisms in human vision *J. opt. Soc. Am.* **A1**, 876–884.

Swindale, N. V. & Cynader, M. S. 1985 Sensitivity of neurones in cat area 17 to vernier breaks in bar stimuli. *Invest. Ophthal. Sup* **26**, 265.

Tolhurst, D. J. 1973 Separate channels for the analysis of the shape and the movement of a moving visual stimulus. *J. Physiol.*, (*Lond.*) **231**, 385–402.

Watson, A. B. & Ahumada, A. J. 1983 A look at motion in the frequency domain. *NASA tech. Memo.* no. **84352**.

Watson, A. B. & Ahumada, A. J. 1985 Model of human visual-motion sensing. *J. opt. Soc. Am.* **A2**, 322–341.

Watson, A. B., Ahumada, J. & Farrell, J. E. 1983 The window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. *NASA Tech. Pap. no.* **2211**.

Watson, A. B. & Robson, J. G. 1981 Discrimination at threshold: labelled detectors in human vision. *Vision. Res.* **21**, 1115–1122.

Weisstein, N. A. 1968 A Rashevsky–Landahl neural net: Simulation of metacontrast. *Psychol. Rev.* **75**, 494–521.