Proc. R. Soc. Lond. B **213**, 451–477 (1981)
Printed in Great Britain

Visual hyperacuity: spatiotemporal interpolation in human vision

BY M. FAHLE AND T. POGGIO†

Max-Planck-Institut für biologische Kybernetik, Spemannstrasse 38, *D* 7400 *Tübingen, F.R.G.*

(*Communicated by H. B. Barlow, F.R.S. – Received* 13 *April* 1981)

Stroboscopic presentation of a moving object can be interpolated by our visual system into the perception of continuous motion. The precision of this interpolation process has been explored by measuring the vernier discrimination threshold for targets displayed stroboscopically at a sequence of stations. The vernier targets, moving at constant velocity, were presented either with a spatial offset or with a temporal offset or with both. The main results are:

(1) vernier acuity for spatial offset is rather invariant over a wide range of velocities and separations between the stations (see Westheimer & McKee 1975);

(2) vernier acuity for temporal offset depends on spatial separation and velocity. At each separation there is an optimal velocity such that the strobe interval is roughly constant at about 30 ms; optimal acuity decreases with increasing separation;

(3) blur of the vernier pattern decreases acuity for spatial offsets, but improves acuity for temporal offsets (at high velocities and large separations);

(4) a temporal offset exactly compensates the equivalent (at the given velocity) spatial offset only for a small separation and optimal velocity; otherwise the spatial offset dominates.

A theoretical analysis of the interpolation problem suggests a computational scheme based on the assumption of constant velocity motion. This assumption reflects a constraint satisfied in normal vision over the short times and small distances normally relevant for the interpolation process. A reasonable implementation of this scheme only requires a set of independent, direction selective spatiotemporal channels, that is receptive fields with the different sizes and temporal properties revealed by psychophysical experiments. It is concluded that sophisticated mechanisms are not required to account for the main properties of vernier acuity with moving targets. It is furthermore suggested that the spatiotemporal channels of human vision may be the interpolation filters themselves. Possible neurophysiological implications are briefly discussed.

SECTION 1

1.0. *Introduction*

Since Wülfing (1892) first described vernier acuity the extraordinary accuracy with which the human eye can estimate the relative positions of lines or other

† Present address: Department of Psychology and Artificial Intelligence Laboratory, M.I.T., Cambridge, Massachusetts 02139, U.S.A.

features in the visual field has represented a long-standing puzzle in vision research. Acuity of this type, also called hyperacuity, can be measured in a variety of situations (Westheimer & McKee 1977*b*). A typical example is the acuity found in reading a vernier (see inset of figure 8*a*). This can be as small as 5″ (Westheimer & McKee 1975), that is 0.02 mm at 1 m distance. The astonishing precision of this performance can be seen when the optical properties of the human eye are considered. In the fovea the hexagonal grid of cones samples the visual image with a sampling interval no finer than 25″, well matched to the optical point spread function of the eye (its Gaussian core has a half width of about 45″, corresponding to a spatial frequency cutoff of about 60 cycle/deg). The spacing of the smallest retinal ganglion cells, which connect the eye to the brain, is at best as fine as this; moreover, their receptive fields are probably quite large, with centre diameters of no less than 1′ (Marr *et al.* 1980). Yet hyperacuity shows that the brain can interpolate the position of a feature in the visual field with an accuracy of better than one-fifth of the distance between two neighbouring photoreceptors and one-tenth of the smallest receptive field centre of the ganglion cells.

Most remarkable of all, vernier acuity is not affected by movement of the target in a velocity range from 0°/s to at least 4°/s (Westheimer & McKee 1975). This means that a subject can detect the relative position of two lines to within a fraction of receptor diameter (and spacing) while the whole pattern is moving across 70 receptors in 150 ms. Recently, evidence has been accumulating which suggests that the visual system is able to perform a very precise temporal interpolation as well, by reconstructing the spatial pattern of activity at moments intermediate between discrete temporal presentations (Barlow 1979). The most telling demonstration, apart from cinematography, was introduced by Burr (1979*a*; see also Morgan 1976) and is shown in the top inset of figure 8*c*. Vernier line segments are displayed stroboscopically at a series of stations to portray a moving vernier; an illusory displacement occurs if the line segments are accurately aligned in space but are displayed with a few milliseconds delay in one sequence relative to the other. Not only do the segments appear to move smoothly from one station to the next but also, between the strobes, they are seen to occupy positions between those where they are actually exposed. The accuracy of detecting the equivalent displacement is again in the vernier acuity range, provided that the target moves at appropriate constant speed and elicits a clear sensation of motion. One is forced to conclude that not only spatial but also temporal interpolation is performed in the visual system to preserve acuity (and resolution) for objects in motion (see Barlow 1979).

It is clear that the attainment of such spatiotemporal accuracy does not break any physical law (see Westheimer 1976). As pointed out by Barlow (1979) and by Crick *et al.* (1981), the classical sampling theorem allows a correct reconstruction of the visual input from a set of discrete samples in space and time since the lateral geniculate nucleus (l.g.n.) signal is bandlimited in temporal and spatial frequency by the photoreceptor kinetics and the eye's optics respectively. In particular, Crick *et al.* have suggested, similarly to Barlow, that the fine grid of granule cells in layer IVc of the striate cortex performs an interpolation on the output of the l.g.n. fibres, with the goal of representing the position of zero-crossings (the boundaries between

activity in an ON and OFF ganglion cell layer) with a very high accuracy (see also: Marr & Hildreth 1980; Marr *et al.* 1979, 1980).

Although spatiotemporal interpolation can be well understood in terms of information theory, the astonishing performance of the visual system seems to require an algorithm and corresponding mechanisms of great ingenuity and precision. Here we investigate some properties of this spatiotemporal interpolation. In particular, we examine its performance for a range of 'sampling intervals' in space and time. As is to be expected, temporal and spatial interpolation depend on the various parameters in characteristically different ways. Our data constrain the way that interpolation is done; in particular, they imply that the basic properties of the spatiotemporal interpolation can be understood in terms of the known properties of the visual system. We present first the experimental data an then a theoretical framework for interpreting them.

1.1. *Methods*

The vernier target used in these experiments consisted of a thin vertical bar made up of two segments. The stimuli were generated on a Tektronix 604 display equipped with P31 phosphor (fading to 1 % within 250 μs) under the control of analogue electronics built by our workshop. Each bar was intensified for 0.1 ms at Δt ms intervals at n successive stations horizontally displaced by a separation Δx. The vertical position was generated by a triangular waveform with a frequency of 3 MHz, to which a d.c. signal could be added. The horizontal position of the beam was controlled by a staircase signal: the height of the steps determined the separation between stations, and their duration the temporal frequency. The accuracy of relative positioning was limited by noise in the electronics to about ± 0.002 mm (on the screen). Each of the two segments making up the bar was 24′ high and 1.5′ wide intensified to a luminance of about 50 times detection threshold on a background of 10 cd/m^2. Moderate room illumination was supplied by overhead lighting. Subjects sat at a distance of 2.3 m from the cathode ray tube display and viewed the screen binocularly with natural pupils. Additional experiments were made at distances of 0.6, 1.1 or 6.8 m. Changes in observation distance did not affect the results. During an experimental run, a target was presented every 3 s. Brief displays of $n\Delta t = 150$ ms, with randomized direction of motion (terminating at the central fixation point) were used to prevent effective eye movements (Westheimer 1954). To be certain of this we recorded eye movements during some of the experiments. The position of the eye, which was illuminated with infrared light, was monitored by a videocamera (Shibaden). Analysis of the records showed that during the experiments there were no pursuit movements of amplitude above 10′. The experiments measured

(*a*) the acuity for detection of real vernier offsets of the two segments by δx seconds of arc,

(*b*) the acuity for detection of apparent vernier offsets produced by delaying the presentation of the lower or upper segment, displayed at the same sequence of stations, by δt ms,

(*c*) the acuity for detection of mixed vernier offsets produced by a real spatial offsét δx together with a temporal delay δt of opposite sign.

In a forced choice task the subject was required to signal whether the bottom segment was displaced to the right or to the left of the top segment by setting a binary switch. In some experiments a buzzer signalled to the subject his incorrect decision. Acuity was determined by the standard criterion of 75% correct identification. Responses (at least 150 for each data point) were collected from three observers who performed all the experiments reported here. The three subjects (T.V., A.K., and H.W.) were male students aged between 19 and 25. One of them (H.W.) has normal sight, the two others attain normal vision with correcting spectacles. They were naïve concerning the goals of this study. The main effects have been verified with several other observers, including the two authors.

Since we wanted to study how vernier acuity depends on separation between the stations (Δx) and on velocity (v) we had to choose whether to keep constant the number of stations (n) or the total presentation time ($T = n\Delta t$). The presentation time T had to be kept below 150 ms to avoid pursuit eye movements (Westheimer 1954). In all experiments reported here T is therefore constant ($T = 150$ ms) and, as a consequence, the number of stations n is variable ($n = 2$ to 95).

Notice that the strobe interval Δt, the distance between stations Δx and the velocity v are not independent variables. Any two of them suffice to determine the third since $\Delta x = v\Delta t$. Since the total presentation time was constant, the total distance over which the pattern travelled changed as a function of velocity v. Although in most cases the stimulus remained restricted to the fovea, at high velocities the stimulus was to a large extent non-foveal (at $v = 80°/s$, the total distance was about 12°).

1.2. *The spatial type of acuity: dependence on velocity* (v) *and separation* (Δx)

The results for spatial offsets (with simultaneous presentation of the two segments at each station) are shown in figure 1*a*, *b*. The main result is that spatial acuity is relatively independent of the separation between the stations and of the velocity of the target up to rather large velocities. These data confirm and extend the results of Westheimer's & McKee (1975), which showed that vernier acuity is unaffected by rate of movement from 0°/s up to 4°/s. Our results imply that this type of vernier acuity is relatively independent of Δt, the strobe interval.

We emphasize that the high acuities at high velocities of figure 1 should not be achieved with real motion. Acuity for continuous motion (not studied in this paper) is expected to deteriorate at high velocities. We were able to support this expectation with several subjects at the limit of the velocity range available to our set-up. Figure 1*a* for instance shows that at separation $\Delta x = 1'$ and velocity $v = 10°/s$, acuity is worse. This value thus represents an upper bound to the limit velocity expected for continuous motion. As indicated in the legend of figure 1, we found that it was important in some of these high velocity, small separation measurements to mask offset information at the start and at the end of the sequence. Without this precaution subjects are sometimes able to adjust their criterion to the ending configuration of the (smeared) pattern and perform significantly better. The procedure that we used was similar to the 'masked edge condition' of Westheimer & McKee (1975). Consistently with their results we found

no significant difference between the thresholds obtained with the two techniques for small and medium velocities (at $\Delta x = 2.5'$). Only at the highest velocities some subjects showed, after a training period, a significant difference. Our data and informal experiments suggest that at large separations (certainly for $\Delta x \geqslant 7.5'$) acuity is invariant over a very wide range of velocities.

In this experiment we measured pure vernier acuity. The appearance of the spatiotemporal pattern changed considerably for too low or too high strobe rates.

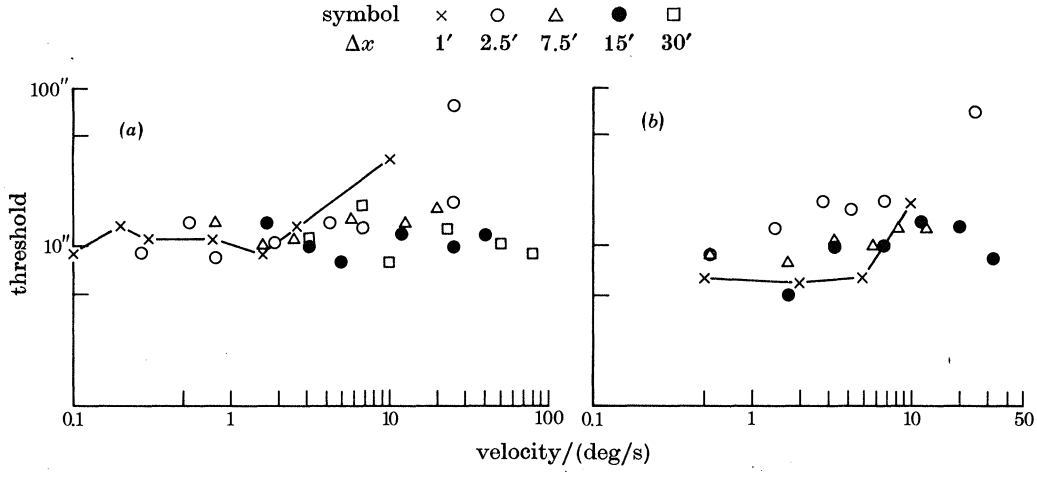

FIGURE 1. Vernier resolution threshold of spatial offset for different separations Δx between the stations as a function of velocity: (*a*) shows the data from subject A.K., (*b*) those from subject T.V. The standard deviation of the data is about 25% of the threshold value for (*a*) and 20% for (*b*). In (*a*) the point for $\Delta x = 1'$ and $v = 10°/s$ was measured masking the beginning and the ending of the trajectory; the same procedure did not change the threshold for the point at $v = 2.6°/s$ ($\Delta x = 1'$). Of the two points at $\Delta x = 2.5'$ and $v = 25°/s$ in figure 1*a*, the worse value has been measured under the 'masking' condition whereas the better one was measured in the standard way. In (*b*) also the point at $\Delta x = 2.5'$ and $v = 25°/s$ was measured with zero offset at the first and last station.

The range of perceived smooth motion depends in fact on Δt: it is for $\Delta x = 15'$ between about 2.5 and 18°/s, for $\Delta x = 7.5'$ between 1.5 and 9°/s and for $\Delta x = 2.5'$ between 0.7 and 5°/s. Whereas a sharp bar smoothly moving was perceived at intermediate velocities, at very low strobe rates (*ca.* 10 Hz) the display began to flicker and a sequence of flashes of isolated vernier was perceived. For high strobe rates (*ca.* 60 Hz) and large separations, many of the stations seemed to be simultaneously illuminated, giving the impression of a stationary grating of parallel verniers. Under these conditions, vernier resolution remained essentially unchanged (see figure 1), although the spatiotemporal pattern was perceived in a distorted form. At high strobe rates and a separation of $\Delta x = 1'$, motion perception was strong but the vernier pattern was considerably smeared: acuity was then lower.

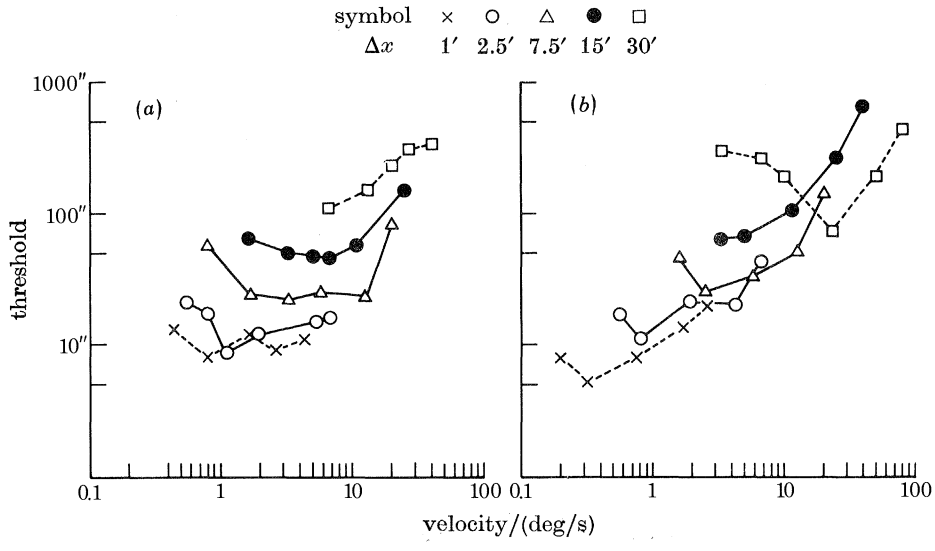

FIGURE 2. Vernier resolution thresholds of temporal offset for different separations between the stations as a function of velocity. Figure 2*a* shows the data from subject A.K., figure 2*b* those from subject T.V. The standard deviation is about 20% of the threshold values for subject A.K. and 18% for subject T.V.

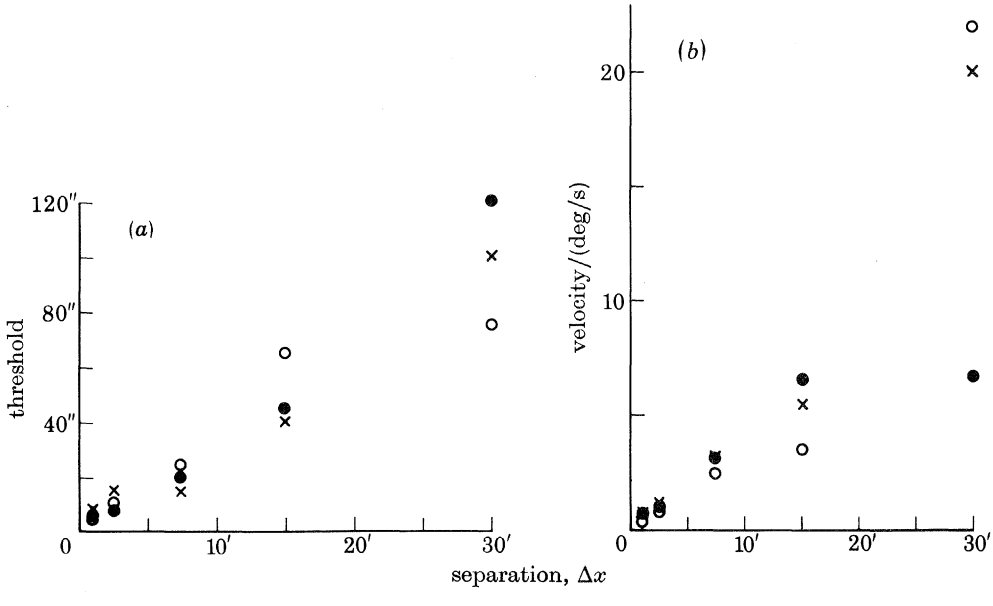

FIGURE 3. (*a*) The best vernier resolution threshold (with temporal offset) for each separation Δx. The data are from three subjects (partly from figure 2*a*, *b*): ○, A.K.; ●, T.V.; ×, H.W. (*b*) The velocity v for which optimal vernier resolution is found is plotted against the separation Δx. Same data as in figure 3*a*.

1.3. *The temporal type of acuity: dependence on v and Δx*

Figure 2*a*, *b* show the results for temporal offsets. The accuracy of detecting the equivalent displacement ($v\delta t$) is in the classical vernier acuity range (cf. Burr 1979*a*, *b*): the best value for observer A.K. was 8″ for spatial and 5″ for temporal offset at comparable separations and velocities. Our main new result is that, although acuity does not break down for large separations between the stations, at least up to half a degree, it deteriorates significantly almost in proportion to Δx (see figure 3*a*).

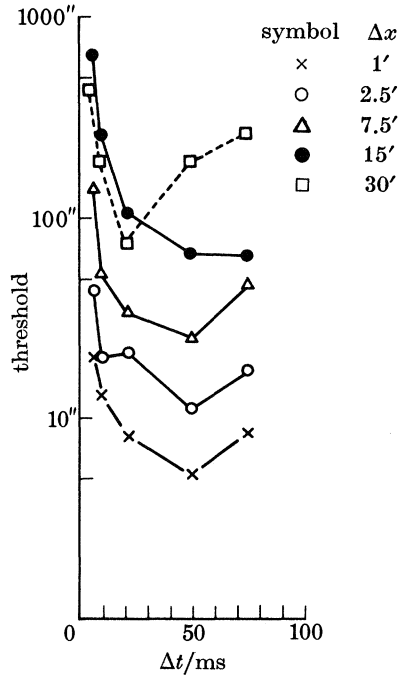

FIGURE 4. The same data of figure 2*a* are plotted as a function of the temporal interval Δt between presentations, instead of velocity.

Vernier acuity of this temporal type is bad at low and high speed. As already clearly demonstrated by Burr (1979*a*, *b*), apparent motion is necessary for temporal offsets to be seen as spatial offsets. In our experiments, deterioration of acuity at low velocities could be due to the speed *per se* as well as to the lower number of stations (because our total presentation time was constrained to $T = 150$ ms the stimulus consisted, at the lowest velocities, of two stations). It is difficult to distinguish between the two possibilities. When we kept the number of stations constant and equal to seven over the velocity range, vernier acuity was no better and sometimes worse at very low velocity than at medium velocity, despite a large presentation time T which allowed effective eye movements. In any case, deterioration of acuity at low velocities can be linked with a decreased sensation of motion (see discussion).

A second important result is that the range of velocities for which temporal interpolation is good shifts upwards for larger separations between the stations. Figure 3*b* plots the velocity for which acuity is best as a function of the corresponding separation. The fact that at higher separations higher velocities are required for good resolution suggests that a more revealing parameter is the time interval Δt between the strobes. Figure 4 shows the same data of figure 2*a* as a function of the strobe interval Δt instead of the velocity v. At any separation Δx, temporal interpolation is optimal for a temporal interval Δt between 50 and 20 ms.

1.4. *The effect of blur on spatial and temporal acuity*

Standard vernier acuity is known to be affected, as one would expect, by attenuation of the high spatial frequencies, i.e. blur, of the vernier pattern (see, for instance, Stigmar 1971). Is temporal interpolation also degraded in the same way?

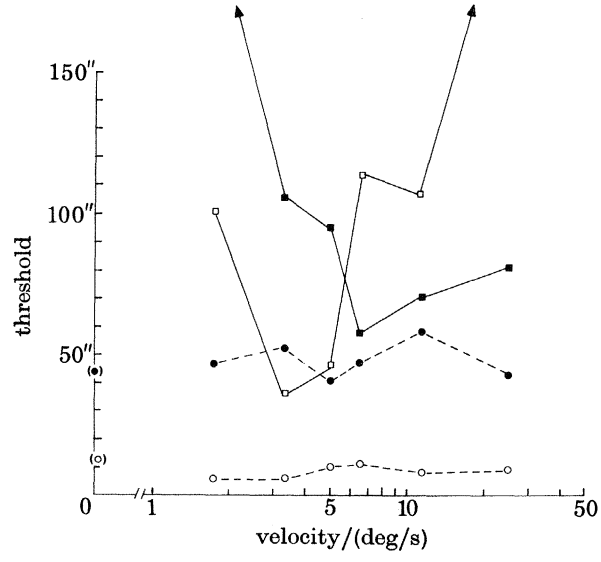

FIGURE 5. The effect of blur on spatial and temporal interpolation as a function of velocity for a separation between the stations $\Delta x = 15'$. Vernier resolution of a spatial offset is measured with (●) and without blur (○). Vernier resolution of a temporal offset is also shown with (■) and without (□) blur. The screen was blurred as described in the text. Notice that the first point for spatial offset is for $v = 0°/s$. The observer is T.V. The standard deviation is about 20% of the threshold values.

We have performed some preliminary experiments to answer this question by placing a ground glass screen at 1 cm in front of the display. When a sharp line is viewed through such a ground glass screen the resulting light distribution has an approximately Gaussian line spread function with a width at half height of between 10′ and 15′, corresponding to a cutoff frequency (at 13% amplitude) of around 3–4 cycle/deg. Our data show that, in the experimental situation of figure 2, blur of the pattern improves acuity at large separations and velocities. Figure 5 compares directly for the same observer and for the same separation the effect

of blur on spatial and temporal interpolation. Westheimer's type of acuity is degraded by blur, whereas Burr's type of acuity improves dramatically with blur (at high velocities). Out of five observers only in one did blur of the pattern cause a reduction in temporal vernier acuity at high separations and velocities. Interestingly, this particular experienced observer had achieved a very good acuity at large separations and high velocities. In all other observers blur significantly improved acuity at large Δx and high v.

These data again show that temporal hyperacuity has different characteristics from spatial hyperacuity. We shall show later that the differential effect of blur on the two kinds of acuity can be directly understood in terms of the two kinds of stimulus and known properties of the visual mechanisms.

1.5. *Spatial versus temporal offset*

With perfect interpolation the apparent offset produced by temporal delay δt would follow the ideal relation $\delta x^t = \delta t\, v$. As shown by our data the sign of the offset is indeed correctly detected. Does its size also satisfy this relation? How faithful, in other words, is temporal interpolation? To answer this question we measured the temporal delay δt needed to compensate for a given real spatial offset δx for different conditions.

Figure 6 shows that for a separation $\Delta x = 2.5'$ and a velocity $v = 1.1°/\text{s}$ the temporal offset $\delta x^t = \delta t\, v$ matches rather closely the real spatial offset δx. Under these conditions spatiotemporal interpolation is indeed rather precise (cf. Burr & Ross 1979). It is not so for higher velocities and/or larger separations (figure 6). The temporal offset needed to compensate for a real spatial offset is then much larger.

These results are somewhat similar to the data of Morgan (1979, 1980*a*, *b*). He found that the precision of the interpolation decreases as the temporal interval Δt is lengthened above about 25 ms (Morgan 1980*a*, p. 170). Our own data, obtained under different conditions, suggest, however, that the temporal interval Δt is not the only critical parameter controlling the precision of the interpolation (see figure 6). Notice that temporal interpolation must fail both at low and at too high speeds where spatial offsets should prevail, simply because movement information is then weak (compare figure 1 with figure 2); the velocities involved, however, in the experiment of figure 6 are not far from being optimal for the corresponding separations. Blur of the pattern improves temporal compensation especially at high velocities (for $\Delta x = 15'$, $v = 10°/\text{s}$ the mismatch is reduced to less than half).

A point worth stressing is that the mismatch is always in the direction of dominance of the spatial over the temporal offset. A possible and somewhat trivial explanation of the effect is based on the demonstration by Burr (1979*a*) that sensation of motion is necessary for temporal offsets to be seen as spatial offsets. One could argue that the first bars, containing possibly weak motion information, would carry only the spatial offset information not compensated by the temporal offset. The responses then could conceivably be biased by these first bars towards the spatial offset. An experiment was designed to test this hypothesis. The first bars were displayed simultaneously at the same position, while the last two were offset in time and space as before. Again the temporal delay failed to compensate

completely for the spatial offset. Evidently, the dominance of the spatial offset is a characteristic property of the interpolation process that cannot be simply explained by the above hypothesis.

1.6. *Summary of results*

In summary our results indicate that vernier acuity for a spatial offset is extremely good over a wide range of velocities and of separations between the stations, whereas acuity for a temporal offset strongly depends on velocity and

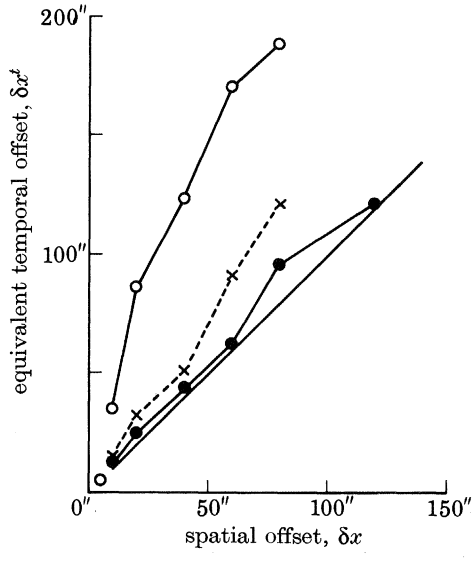

FIGURE 6. Temporal (δx^t) against spatial (δx) offset in the compensation experiment. The ordinate shows the temporal offset (in equivalent spatial units $\delta x^t = v\delta t$) needed to compensate the spatial offset shown in the abscissa: ● is for a separation between the stations $\Delta x = 2.5'$ and a velocity $v = 1.11°/s$ ($\Delta t = 37$ ms); × is for $\Delta x = 2.5'$ and $v = 5.28°/s$ ($\Delta t = 7.9$ ms); ○ is for $\Delta x = 7.5'$ and $v = 4.11°/s$ ($\Delta t = 30$ ms). Larger separations yield an even greater mismatch. The continuous diagonal indicates the loci of perfect compensation. Subject T.V.

separation, deteriorating with increased spatial separation and being optimal for temporal intervals of 10 to 40 ms. The difference between interpolation in the two situations, spatial and temporal, is directly revealed by the blur and compensation experiments. As performance is so characteristically different for spatial vs. temporal interpolation, one might think that different mechanisms are called for. In the theoretical part of this paper we shall show that this is not so. The same simple mechanisms can consistently explain all the major properties of spatiotemporal interpolation in human vision.

Section 2

2.0. *Spatiotemporal interpolation: how is it done?*

The results described in section 1 constrain the problem of hyperacuity tightly enough to justify a theoretical analysis of how spatiotemporal interpolation may be done in the visual system. The precise meaning of interpolation in terms of our visual stimuli is a well defined question, and this is the main point to discuss.

Our analysis in the next sections is mainly cast in the language of Fourier transforms and only to a lesser extent in the complementary language of receptive fields and temporal impulse responses. The reason for this is that the Fourier language provides an especially transparent and precise way of looking at the interpolation problem. Descriptions in terms of spatial and temporal frequencies are of course equivalent to descriptions in terms of real space and time, as long as the processes to be examined are linear. We consider this to be essentially true for the interpolation process revealed by vernier acuity, and that this is implemented in our visual system by standard spatiotemporal filtering operations. This point is best made with a specific example.

2.1. *A simple illustration*

Figure 7 illustrates a very simple scheme for achieving spatiotemporal interpolation of a visual pattern. The elements of this scheme could be, but do not need to be, interpreted as cells with associated receptive fields and temporal impulse responses. Visual input is sampled in space by an array of cells with a sampling density high enough to preserve the whole of the spatial information (in accordance with the sampling theorem). The input is then reconstituted in more detail on a finer grid of cells by convolving the sampled values with the function sinc x†. In effect each cell of the interpolation layer weights its input according to a centre–surround receptive field. In two spatial dimensions, a variety of filters (i.e. 'receptive fields') are capable of performing a correct interpolation (see Crick *et al.* 1981).

If the input intensity distribution is presented at discrete instants in time, temporal interpolation can be achieved by suitable temporal low pass properties of each individual pathway. If the temporal interval between presentations is small enough the effect of the filter is to reconstruct the original continuous temporal input. Spatial interpolation can then operate at each instant of time.

Figure 7*b* shows the Fourier interpretation of the spatial interpolation process (interpolation in time can be interpreted in a similar way). The effect of sampling is to replicate the original spectrum in an infinite number of side lobes. Spatial interpolation, i.e. reconstruction of the original function from its samples, is accomplished by filtering out all side lobes but the central one, which is the original spectrum.

This model is probably the simplest conceivable scheme. In it, interpolations in space and time are performed independently, since the temporal dependence of the input is not constrained in any way. We now consider the conditions under which this scheme can be effective.

† Definition: $\operatorname{sinc} x = (\sin \pi x)/\pi x$.

2.2. *More complex interpolation schemes are required*

The scheme of figure 7 can provide a correct reconstruction of a spatiotemporal input sampled at intervals $\Delta\xi$ (in space) and $\Delta\tau$ (in time) only when the input function is bandlimited in spatial frequency by f_x^c and temporal frequency by f_t^c in such a way that $\Delta\xi < 1/(2f_x^c)$ and $\Delta\tau < 1/(2f_t^c)$ (theorem 1 in appendix 2). The

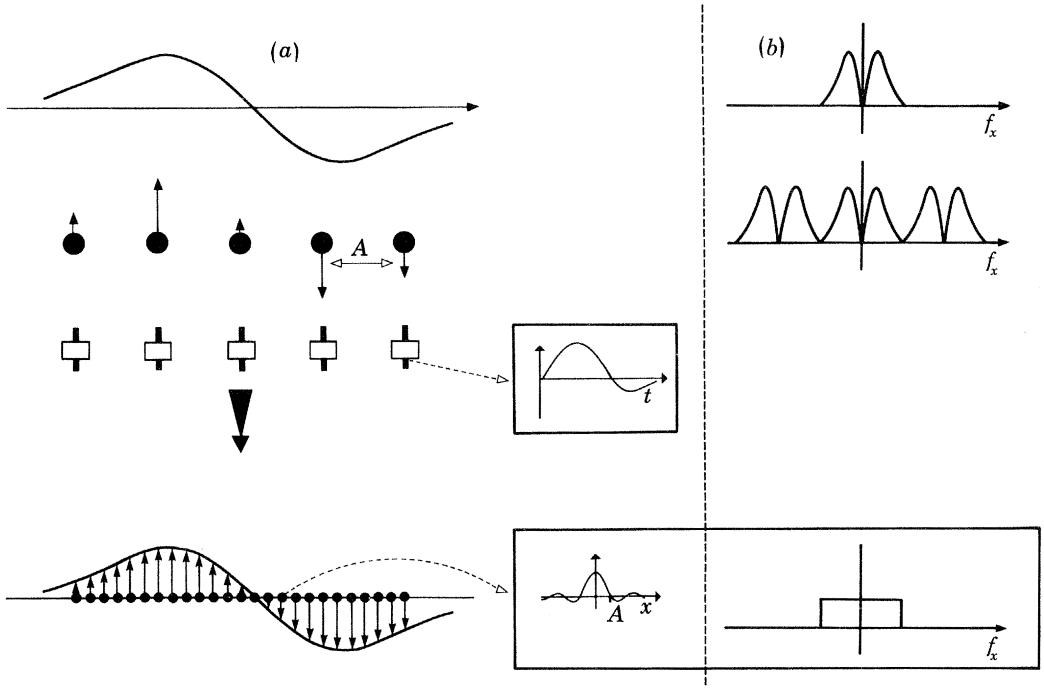

FIGURE 7. (*a*) A simple scheme for spatiotemporal interpolation. The input pattern is sampled by an array of 'cells'. Spatial interpolation is accomplished on a finer interpolation grid of cells each one weighting the sampled values with a sinc shaped receptive field (shown in the lower inset). Temporal interpolation is obtained by filtering with an appropriate low-pass or bandpass filter each of the input channels (its impulse response is shown in the upper inset). Thus a series of discrete frames of a moving pattern could be interpolated (see theorem 1 in appendix 2) into a continuous temporal function in each of the channels. The spatial input distribution outlined here represents an intensity edge as seen by centre–surround ganglion cells. (*b*) The spatial interpolation process in Fourier space. Interpolation is equivalent to filtering out the side lobes originated by the sampling process. Temporal interpolation can be interpreted in a similar way.

image that reaches the retina is indeed bandlimited in spatial frequencies to less than about 60 cycle/deg by the diffraction limited optics of the eye. Furthermore, a temporal cutoff is imposed at the level of the photoreceptors by their limited temporal resolution. The scheme of figure 7 can therefore correctly reconstruct an image sampled at intervals of less than 30″ in space (for the 2-D case see Crick *et al.* (1981)). Temporal samples of the photoreceptor activity could be interpolated under similar conditions (though regular temporal sampling in our visual system is highly implausible).

Since the spacing of the foveal photoreceptors is almost exactly matched to the eye's optics, interpolation in normal vision, when the image is a continuous function of time and space, can be accounted for by simple schemes like that of figure 7. In particular, such models could account for the vernier acuity (in the fovea) measured with real continuous motion of the retinal image. High spatial frequencies may be lost as an effect of the temporal cutoff of vision. Notice that the instantaneous position of, say, an isolated peak in a 'blurred' intensity distribution can be retrieved in principle with arbitrary precision by the process of interpolation. Although signal to noise problems limit the precision of interpolation for increasingly blurred images, acuity can remain quite high. When, however, motion of an object is simulated by presenting the image at discrete positions at separate instants, the conditions of theorem 1 are in general no longer satisfied. In our experiments we present to the eye an image that is already sampled in either time (Westheimer type of stimulus) or space (Burr type of stimulus) or both. We enforce on the system arbitrary sampling intervals Δx and Δt and this before the bandlimiting operations of the eye's optics and of the receptor kinetics come into play. Under these conditions the input function $g(x,t)$ is not ensured to be appropriately bandlimited before spatial or temporal sampling occurs. The scheme of figure 7 should for instance perform poorly when the input function is sampled in space at intervals Δx significantly coarser than the photoreceptor array. Burr's data and our own, however, show that our visual system performs surprisingly well. The simple scheme of figure 7 is also inconsistent with data on the visibility of gratings in movement (Burr 1979*b*). We are clearly forced therefore to consider other types of interpolation schemes.

2.3. *The spatiotemporal spectrum of a moving vernier*

Our analysis of alternative interpolation schemes begins with the description in frequency space of the physical stimuli corresponding to Westheimer's and Burr's experimental situations. When a spatial pattern $g(x)$ moves continuously at constant speed, the resulting spatiotemporal distribution of excitation on the retina has a simple representation in the Fourier space of temporal (f_t) and spatial (f_x) frequencies. Its Fourier transform takes values only on the diagonal line shown in figure 8*a* with a slope equal to (minus) the velocity (appendix 1). For each spatial frequency contained in the pattern, there is a corresponding unique temporal frequency. Curtailing the duration of motion (in our case to $T = 150$ ms) spreads the Fourier transform over a large area of temporal and spatial frequencies, changing the narrow line into a wider area. The spread (along the f_t axis) is the same for all our data. Thus the support of the Fourier transform shown in figure 8 (i.e. the region of the Fourier plane where the transform is different from zero) must be interpreted as being spread along f_t as a sinc function. For $T = 150$ ms the width of the spread is about 14 Hz for the central lobe of the sinc function and 28 Hz for the central lobe plus the first negative side lobe on both sides. The retinal stimulus elicited by continuous motion of a vernier at constant velocity can be described in this way (see appendix 1). The upper and the lower segment have the same line support on the f_x–f_t plane. Their Fourier transforms differ at all

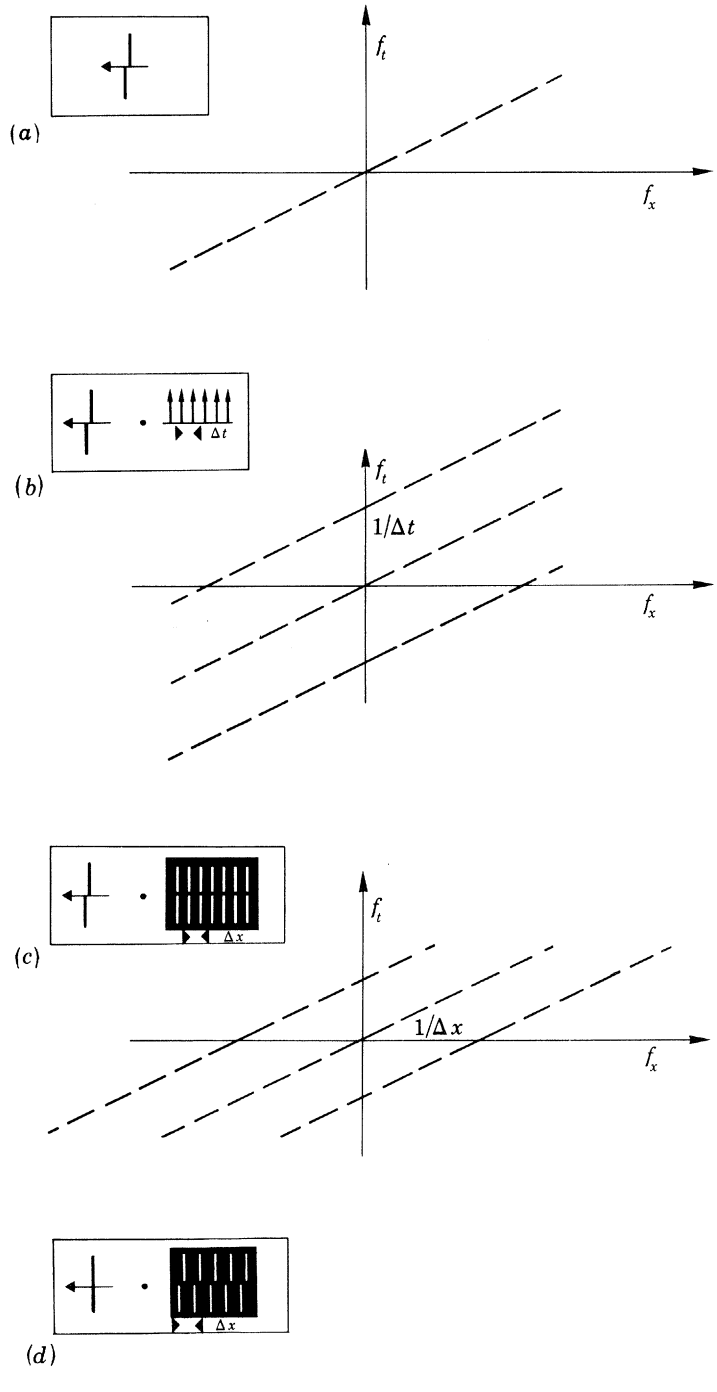

FIGURE 8. For description see opposite.

frequencies only by a phase factor that mirrors the spatial offset. The correct detection of this information underlies positional acuity.

Figure 8 summarizes the description of the two basic stimulus configurations used in this paper according to the derivation outlined in appendix 1. Westheimer's experimental situation is equivalent to looking at the continuous motion of a vernier through a series of equidistant narrow temporal slits within which the pattern is briefly visible (see figure 8*b*). Burr's experimental situation ideally corresponds to a vernier moving behind a spatial window with a series of equidistant narrow slits (see figure 8*c*). The spatial or temporal windows affect differently the spectrum of the retinal input. As shown in appendix 1 and in figure 8, in the Westheimer situation the complex spatial spectrum of the pattern, which contains amplitude and phase information, is replicated an infinite number of times along the temporal frequency axis, whereas in the Burr case the same spectrum is replicated along the spatial frequency axis. The distance between successive replicas is $1/\Delta t$ on f_t ($1/(v\Delta t) = 1/\Delta x$ on f_x) for the case of figure 8*b* and $1/\Delta x$ on f_x in figure 8*c*. An important observation is that in figure 8*b* (Westheimer stimulus) all lobes at any given f_x support exactly the same complex spectrum $\tilde{g}$. This is not so in figure 8*c* (Burr stimulus), where, instead, all lobes have the same $\tilde{g}$ at any given f_t. We re-emphasize that figure 8 describes the physical properties of the different stimuli without any reference to the human visual system.

2.4. *Computational aspects of interpolation: the constant velocity assumption*

More effective interpolation schemes are feasible if general constraints about the nature of the visual input are incorporated directly in the computation. The key observation here is that the temporal dependence of the visual input is usually due to movement of rigid objects, and that in everyday life motion has a nearly constant velocity over the times and distances that are relevant to the interpolation process

FIGURE 8. (*a*) The support on the $f_x - f_t$ plane of the Fourier spectrum associated with continuous motion of a vernier (see inset) at constant velocity $-v$. The slope of the line is v; $\tilde{g}(f_x, f_t) = \tilde{g}(f_x)$ on that line. Curtailing the duration of motion to $T = 150$ ms spreads the line into a bar-like support, corresponding to a sinc function. Under our experimental conditions all line supports shown in figure 7 are spread in this way along the f_t axis. (*b*) The support of the Fourier spectrum associated with Westheimer's type of experiment. The inset indicates that displaying the vernier stroboscopically at a sequence of times with an interval Δt is equivalent to 'looking' at the continuous motion of a vernier through a series of temporal 'slits'. This has the effect of replicating the spectrum of (*a*) along the f_t axis in an infinite number of side lobes. The distance of the lobes on f_t is $1/\Delta t$. The line encounters the f_x axis at $1/(v\Delta t) = 1/\Delta x$ (if $\Delta x = 1'$, the distance of the side lobes on f_x is 60 cycle/deg). Notice that, for any f_x, each lobe supports the same complex Fourier spectrum $\tilde{g}(f_x)$. (*c*) The support of the Fourier spectrum associated with Burr's type of experiment. Displaying the line segment of a vernier at the same position but with a slight delay is equivalent to looking at the continuous motion of a vernier through the spatial window depicted in the inset (transparent slits in otherwise opaque screen). As shown in appendix 1 this corresponds to replicating the spectrum of (*a*) along the f_x axis. The distance of the lobes is $1/\Delta x$, where Δx is the interval between successive slits in the spatial window. At a given f_x, the Fourier spectrum $\tilde{g}(f_x)$ of different lobes is in general different. (*d*) The support of the Fourier spectrum associated with the compensation experiment is the same as in (*c*). The different window corresponding to this stimulus (see inset) corresponds, however, to a different complex Fourier spectrum (see appendix 1).

($T < 100$ ms and $x < 1°$). The constant velocity assumption leads to a specific form of the sampling theorem, given in appendix 2 (see also Crick *et al.* 1981), which states formally what is intuitively clear: the spatiotemporal sampling rate can become very low without losing information. Interpolation schemes based on the constant velocity assumption exploit the equivalence of the time and space variable. From the point of view of filtering this means that spatial and temporal interpolation cannot be performed independently as in the simple scheme of figure 7. In the Fourier domain the constant velocity assumption constrains the spectrum of the visual input to lie on the line support shown in figure 8*a*. In the ideal case of infinitely long motion the side lobes generated by sampling either in time (figure 8*b*) or space (figure 8*c*) can always be excluded by means of appropriate filters, if the precise value of v is known (e.g. by measurements). The recovery of the original spectrum (figure 8*a*) corresponds to an ideal interpolation for arbitrarily large sampling intervals (if v is known and different from zero). In the realistic case of finite duration of motion finite sampling intervals are enforced by the spread of the Fourier spectrum into a larger area, but the same basic arguments will apply.†

An interpolation scheme of this type could be implemented simply by measuring the exact velocity of movement and then reconstructing the spatiotemporal trajectory of the pattern for either temporal or spatial information. Another, more attractive possibility is suggested by the idea, supported by much psychophysical evidence, that in the human visual system there exist several channels at each eccentricity, i.e. several sets of receptive fields tuned to different spatial sizes and with different temporal properties. We imagine, following Burr (1979*b*), that these channels, with direction selective properties, have somewhat overlapping supports covering the region of the f_x–f_t Fourier plane that corresponds to the sensitive range of the visual system . 'Stasis' channels are tuned to high spatial frequencies (small receptive fields) and low temporal frequencies (sustained properties); 'motion' channels are tuned to low spatial frequencies (large receptive fields) and high temporal frequencies (transient properties). Thus, each channel is tuned to a different range of velocities, centred on the ratio between the optimal temporal and spatial frequencies characteristic for the channel: stasis channels for instance are tuned to low velocities whereas motion channels are tuned to high velocities. Thus, as revealed by detection experiments high spatial frequencies summate well in time when stationary but poorly when drifting, whereas low frequencies summate rather poorly when stationary but quite well when drifting (Burr 1979*b*; cf. Watson 1979 and Watson & Nachmias 1977). Figure 9*b* shows a set of idealized 'velocity channels' of this type. Since each channel has its own cutoff in temporal and spatial frequency, interpolation may be performed independently and with different characteristics within each channel. In the Burr type of experiment stasis channels could correctly interpolate only patterns displayed at small separations and low velocities, whereas motion channels could interpolate (but not so accurately, owing to the loss of high spatial frequencies) at

† Retinal drift motion would mediate normal vision of 'stationary' patterns (Kelly 1979). Theorem 2 in particular implies that a possible aliasing for zero velocity can be avoided by motion (at constant speed) of the retinal image. This is especially suggestive in the case of non-foveal vision.

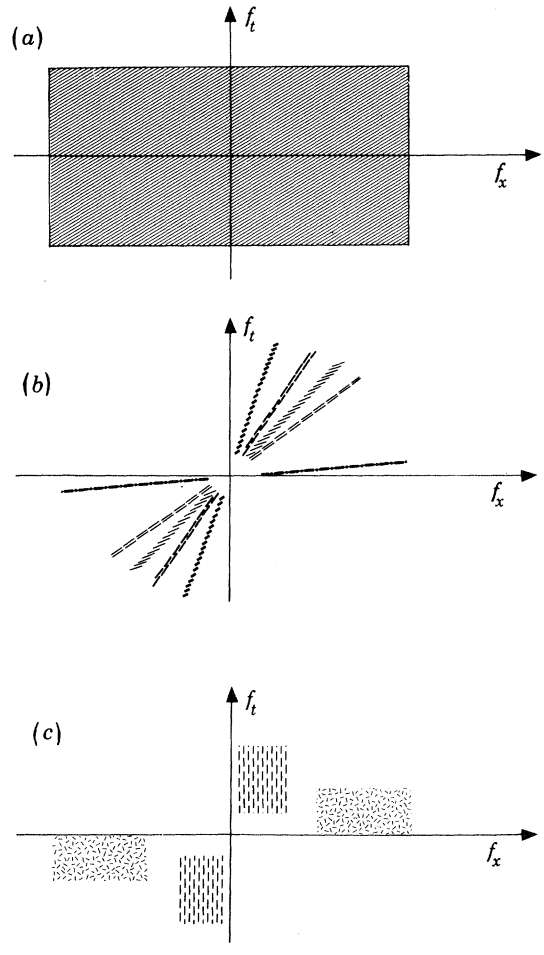

FIGURE 9. (*a*) The support on the Fourier plane of spatial and temporal frequencies of an interpolation filter corresponding to a scheme such as figure 7. (*b*) The support on the Fourier plane of a set of spatiotemporal filters ideally tuned to different velocities. A large number is needed to cover all velocities of interest. The filters are assumed to be direction selective, since they only operate in the Fourier quadrants corresponding to positive $v = f_t/f_x$ in $g(x+vt)$. A spatial pattern moving at constant velocity and sampled at spatial intervals Δx has on this plane the support shown by figure 8*c*. To avoid aliasing, the low velocity filters can be 'switched off' by information about the velocity of the motion. (*c*) A more realistic set of filters, broadly tuned to different velocities. The stasis channel is tuned to low temporal and high spatial frequencies and thus to low velocities. The motion channel is tuned to high temporal and low spatial frequencies and thus to high velocities. Intermediate channels (not shown here) may be also present. The hatched areas represent the support of such directional filters. Non-directional filters would have a symmetric support in the other two quadrants.

large separations and high velocities by filtering out the side lobes arising from the coarse spatial sampling. The complementary argument applies for coarse time sampling. The stasis channels may suffer from aliasing at values of Δx for which the motion channels interpolate correctly (cf. figure 9*b*). We assume, then, that in this scheme the wrong channels are switched off by use of velocity information.

Figure 9*c* shows a more realistic interpolation scheme of the same basic type. Instead of many channels, each one sharply tuned to velocity and inactivated when the pattern does not move at its characteristic velocity, there are a few channels coarsely tuned to velocity and without very specific velocity sensitive inactivation, apart from directional selective properties. In this specific algorithm spatial and temporal interpolation may even be performed independently within each channel (directional selective properties aside); with appropriate filter functions, then, figure 7 may be a rough description of each of the channels of the scheme of figure 9*c*. Directional selectivity, however, would require the spatio-temporal interpolation filter to change according to the direction of motion (see appendix 2).

In the light of this analysis we turn now to a detailed discussion of our experiments. Our main question concerns of course which type of interpolation scheme is actually used by our visual system.

2.5. *Westheimer's acuity: recovery of spatial offset*

In Fourier terms, the aim of the interpolation process is to filter out the side lobes, preserving only the central lobe, as the latter represents the Fourier spectrum of a continuously moving bar.

When both the time interval Δt between presentations and the velocity v are small, interlacing of the side lobes in the Fourier spectrum is negligible. Temporal low pass properties of the visual pathway, as in the model of figure 9*a*, suffice for eliminating the side lobes and thus achieve a correct interpolation. When Δt is large, however, interlacing is considerable in the sense that, even for the scheme of figure 9*c*, there are one or more channels that mix the main lobe with at least one of the side lobes. Because of the spread associated with the short duration of the motion sequence, actual overlap between the lobes can be significant. It turns out, however, that this does not represent a problem from the point of view of the spatial acuity measured in our experiments. At each f_x the complex Fourier spectrum of all side lobes is exactly the same. Thus, the spatial spectrum is correct irrespectively of the temporal frequency and independently of the number of side lobes contained in the support of the interpolation filters. At large Δx and high v, the presence of the side lobes turns out to be even beneficial for vernier acuity (consider figures 8 and 9*c*); under these conditions high frequency channels, which would not be stimulated by continuous motion, can obtain the correct spatial information from the side lobes, which are an artefact of the discrete time presentations (see figure 8*b*). On the whole, and in the absence of a sophisticated interpolation process that always excludes all side lobes (such as the scheme of figure 9*b*), one expects vernier acuity to be rather invariant for a wide range of separations and velocities. For very small separations ($\Delta x = 1'$) or for continuous motion, the upper velocity limit is linked to motion smear (the effect on acuity is similar to blur of the pattern, via the temporal cutoff of vision). Our data conform

well to these expectations. Notice that the presence of side lobes at high velocities and large separations corresponds to the perception not of a moving bar but of a briefly illuminated stationary grating, which carries however the correct spatial information. In this sense at large Δx and high v interpolation fails to retrieve the 'correct' spatiotemporal pattern, but still preserves spatial acuity (even at extremely high speeds). An irony of the failure of interpolation is that, as a consequence, hyperacuity of the Westheimer type is better at large Δx and high v than for successful interpolation (and continuous motion at the same speed).

The qualitative interpretation of our data in usual space–time variables is straightforward. Spatial interpolation, for instance by appropriate receptive fields, takes place correctly for each frame (i.e. for each station) even when temporal interpolation fails. Since our forced choice task measures only spatial acuity, performance is in this case independent of the interpolation of the temporal dependence of the visual input.

These results suggest that spatiotemporal interpolation is not performed by the 'ideal' interpolation scheme of figure 9*b*. For temporal aspects should then be retrieved correctly at all Δt, while acuity for high velocities should be exactly as bad as for continuous motion, irrespectively of separation.

2.6. *Burr's acuity: interpolation of temporal offset*

In Burr's experiment the situation is quite different. For any given f_x the side lobes contain different parts of the original spectrum. Thus when more side lobes lie in the support of the same channel (figure 9*a*, *c*) there is a mixture of spatial frequencies, detrimental to acuity. One understands, therefore, that acuity deteriorates considerably (see figure 2) with increasing overlap among the side lobes (large separations between the stations). At any given (large) separation, low velocities bring about a considerable overlap between the side lobes. Higher velocities reduce the degree of overlap at the expense of high spatial frequency information, which is filtered out by the temporal cutoff(s) of the visual pathway (between 20 and 50 Hz; see, for instance, Kelly 1979). Thus one expects to find for each Δx, an optimal velocity at which the side lobes just avoid overlap. If one assumes a half-spread of *ca.* 15 Hz (see §2.3) the optimal velocity (in degrees per second) should be $v = 30\Delta x$ (Δx in degrees), which is in rough agreement with the data of figure 3*b*. When the velocity approaches zero the line supports in figure 8*c* all tend to lie on the f_x axis (notice that, because of the finite presentation time T, the supports effectively overlap). In this situation information about the offset cannot be retrieved. In the limit of very high velocity the set of lobes approaches the spectrum of a stationary grating with no offset. Notice that we assume for the scheme of figure 9*c* that the vernier threshold is higher when some of the channels signal zero offset while the others still 'see' the correct offset.

When the temporal component of the filters fails to interpolate between temporal frames motion is perceived as discontinuous.† As a consequence the spatial interpolation process correctly signals zero spatial offset for each frame.

† Sensation of smooth motion is not always a good indication of the quality of interpolation. At a given velocity the perception of smooth motion might be obtained simply by increasing Δx (and thus the overlap between the side lobes of the Fourier plane). As expected the precision of the interpolation degrades (see §2.4 and figure 2). The perception of smooth motion may therefore be distinct from the interpolation process.

The scheme of figure 9*b* should clearly interpolate much better than our visual system does. Though its performance may worsen at high velocities, as for continuous motion, it should be rather invariant with respect to Δx, the separation between the stations. Figure 3*a* shows that this does not happen. The opposite conclusion holds for the scheme of figure 9*a*. Its performance should deteriorate rapidly for separations Δx between the stations larger than the distance between the photoreceptors, which is in conflict with Burr's data and our own. An interpolation scheme of the type of figure 9*c* seems consistent with these results: while small, slow 'receptive fields' would be unable to interpolate correctly at large separations Δx, large, fast receptive fields could perform a correct interpolation, if the velocity is appropriate.

The fact that spatial acuity is extremely good at separations up to 2.5′ suggests that the interpolation channels have direction selective properties. As figure 9*c* shows, this is equivalent to switching off the mechanisms responding to the wrong direction of motion. Otherwise these mechanisms may be excited by side lobes that partly lie in the wrong quadrant. For instance, at $\Delta x = 2.5'$ and low velocities the side lobes, centred at 24 Hz, invade substantially the wrong quadrants. Small non-directional receptive fields would pick up this wrong information (see figure 9*c*).

2.7. *Effect of blur*

The interpolation scheme outline in figure 9*c* makes a rather strong prediction about the effect of blur. In the Westheimer case blur can only degrade vernier acuity, since it eliminates the high frequency channels. Blur of the Burr stimulus, however, should improve acuity at least at large separations and high velocities, since it eliminates side lobes that signal the absence of an offset. Our data are fully consistent with this expectation. A more perceptual but equivalent description of the effect of blur is this. At high velocities and large separations there is a strong sensation of a grating of thin, unbroken lines, corresponding to the side lobes seen by visual mechanisms tuned to low temporal and high spatial frequencies, and a weak impression of a single moving target with a clear offset, corresponding to the main lobe seen by mechanisms tuned to lower spatial and higher temporal frequencies. This ambiguity is removed, as already noticed by Burr (1979*b*), by blur of the screen, which suppresses the high frequency grating.

In other terms, blur eliminates the contribution of the small receptive fields which are unable to interpolate correctly at large separations and therefore signal zero offset. The large receptive fields, however, remain largely unaffected by blur.

The effectiveness of blur in improving vernier acuity at large Δx shows that our visual system does not normally have the intrinsic possibility of switching off, i.e. inhibiting completely, the wrong channels as assumed in the scheme of figure 9*b*. Our data, however, cannot exclude some degree of inhibitory interaction between different 'velocity channels'. As H. Barlow suggested (personal communication), mutual inhibition between the motion and the stasis channels would have several attractive properties, especially concerning the sensitivity of the visual system to low spatial and low temporal frequencies.

2.8. *Spatial versus temporal compensation*

This stimulus situation corresponds to looking at the continuous motion of a vernier through the spatial window shown in the inset of figure 8*d*. The resulting Fourier support, as shown in appendix 1, is again as in figure 8*c*: here, however, the main lobe signals no offset, corresponding to precise spatiotemporal compensation, whereas the other lobes all signal the spatial offset between the upper and lower grating of the window. In other words, exact compensation between space and time is realized only in the main, correct lobe. Thus, the spatial offset should dominate as soon as the side lobes are 'seen' by some of the channels of figure 9*c*. This is increasingly so for larger separations Δx between the stations. Correspondingly, the perception of the stationary grating carrying spatial offset information (the broken slits in the window of figure 8*d*) is expected to dominate at large separations and velocities. Again our data are consistent with these expectations. Even at relatively small separations between the stations (see figure 5) the system does not achieve a perfect interpolation, that is, removal of all side lobes. Only in this case would the temporal offset exactly cancel the spatial offset. As expected, blur improves compensation, since it helps to remove the 'wrong' side lobes, which carry information only about the spatial offset.

This experiment combines Burr and Westheimer stimuli. Since spatial interpolation always retrieves the spatial offset, this dominates for all cases in which the temporal component of interpolation is not fully correct.

2.9. *Conclusions*

To summarize, the psychophysical experiments reported here suggest that spatiotemporal interpolation in the visual system, remarkable though it is, is far from being perfect and flawless. Ideal interpolation is equivalent to filtering out the side lobes in the Fourier spectrum arising from the discrete presentations. The task is easy at small separations but requires in principle complex filters for large separations (see Crick *et al.* 1981). As our data suggest, our visual systems do not seem to use a very sophisticated spatiotemporal interpolation process. The side lobes are not effectively filtered out under all conditions. Spatiotemporal interpolation, then, can be considered as a direct consequence of the spatial and temporal properties of vision, in terms of an interpolation scheme of the type of figure 9*c*. The existence of independent, possibly direction selective channels tuned to different spatial and temporal frequencies seems sufficient to account for the spatiotemporal interpolation revealed by our experiments. A detailed theoretical analysis with the help of appropriate computer experiments will be necessary for a quantitative evaluation of specific interpolation models of this type, especially in two spatial dimensions. The goal of the interpolation process may also be more modest than a full reconstruction of the input distribution. As suggested by Crick *et al.* (1981), interpolation of the ganglion cells' activity could provide the position of the zero crossings (where activity switches from the on centre to the off centre cells) with high accuracy. This can be achieved by using simple interpolation functions such as a centre–surround receptive field (Marr *et al.* 1980).

It remains a somewhat open question whether this scheme can also explain the

psychophysical perception of smooth motion under a variety of stroboscopic presentations (at short spatial ranges; see Braddick 1980). The problem clearly lies beyond the scope of our experiments and bears upon the rules underlying the synthesis of a visual percept from the information delivered by the various channels (exemplary for stationary patterns is the discussion by Marr & Hildreth (1980)). It is, in addition, still unclear whether spatiotemporal interpolation may simplify the correspondence problem of motion (Ullman 1979; Marr 1982).

2.9.1. *Are the psychophysical channels the interpolation filters?*

Our data support interpolation schemes of the type outlined in figure 9*c*. They say, however, neither how many independent channels are needed, nor what are exactly their spatiotemporal properties. Our results seem consistent with standard characterizations of their spatial and temporal properties (Campbell & Robson 1968; Burr 1979*b*; see also Marr *et al.* 1980; Wilson & Giese 1977; Wilson & Bergen, 1979). Figure 4 suggests, for instance, that the temporal properties of the channels are quite similar; temporal sensitivity may shift towards higher temporal frequencies by no more than a factor 3 for the largest receptive field sizes (see also Burr 1979*b*). Their characteristic spatial frequency tuning, in other words the size of the associated receptive fields, probably spans a much wider range.

These observations suggest the interesting idea that the spatial frequency tuned channels present in human vision may be the interpolation filters themselves. To be completely explicit let us consider simple examples of how an interpolation scheme such as figure 9*c* might be implemented in the visual system. The first possibility is that the image is filtered before interpolation through various independent channels. Retinal or l.g.n. ganglion cells of different sizes could represent the image filtered at different resolutions. Later in the visual pathway each of these representations would be independently interpolated on a finer cortical grid of cells with a receptive field similar to the corresponding l.g.n. cells. Another possibility is that only two of the channels are present at the precortical level (e.g. X and Y) and that the measured psychophysical channels represent interpolation filters operating on their X and Y input at the cortical level. In this second case one would expect only two sizes of receptive fields, at each eccentricity, in the retina and l.g.n. but a scatter of sizes in the cortex. Thus the same retinal channel may be interpolated in two different ways, by small cortical receptive fields and by large ones, the first reconstructing the high frequency content of the retinal channel and the second emphasizing its coarser details. Notice that as a consequence cortical (interpolation) channels would have a narrower bandwidth than retinal ones.

2.9.2. *Neural mechanisms*

Interpolation can be regarded as a spatiotemporal filtering of the input transmitted from the retina. This is the point of view taken in this paper. We cannot advance any hypothesis as to where this filtering stage may be localized in the brain on the basis of our psychophysical data alone. Throughout this paper we have used the term 'interpolation' without necessarily implying a direct reconstruction of the pattern of visual activity, say its zero-crossing profile in the various channels,

somewhere in the visual pathway. The interpolation scheme suggested by our data may be implemented as an 'implicit interpolation', that is, as a computational process involving manipulation of symbolic quantities, or it may depend on an 'explicit reconstruction' of a coded version of photoreceptor activity on a fine retinotopic grid of neurons. These extreme possibilities, and all in between, can be implemented in a variety of ways.

For instance, activity may be reconstructed on the fine grid of layer IVc by spatiotemporal properties of intracortical connections: all or part of the channels may thus be independently realized by such connections. To achieve the required spatiotemporal filtering, connectivity in this layer should implement an appropriate spatiotemporal receptive field. For instance, a 'motion' channel that is bandpass in f_t and lowpass in f_x requires a broad receptive field and a temporal impulse response with a positive peak followed by a negative phase (cf. Burr 1979*b*; see Barlow 1979; van Seelen 1973). Although directional selectivity is likely to be an important property of the interpolation filters, neurons in an interpolation grid do not need to be direction selective themselves, since they may be controlled by other direction selective cells. Interestingly, data about vernier acuity for strongly blurred patterns (see for instance figure 5 and Stigmar (1971)) would imply that the coarse 'motion' channel still requires a rather fine interpolation grid, only about six times sparser on a linear scale than the 'stasis' channel, consistently with the high accuracy which can be theoretically achieved in the interpolation of blurred isolated features. Several of the properties of classical vernier acuity as characterized mainly by Westheimer and coworkers can be explained by neural schemes of this type (Westheimer & Hauske, 1975; Westheimer & McKee 1977*a*, *b*, 1979).

On the other hand, a specific, more symbolic process could read the output of retinal ganglion cells and perform the correct interpolation for any desired position and time. In this case interpolation would be implicit and mixed with the decision procedure itself.

2.9.3. *Interpolation: why?*

The heart of any theory of a visual process is the answer to the question, what is it for? In §2.4 we have assumed that one of the goals of our visual system is spatiotemporal interpolation of the visual input. Under the assumption of constant velocity motion, we have derived an interpolation scheme that seems consistent with our data on vernier acuity. Although stereopsis certainly needs precise position information, we find it difficult to believe that high positional accuracy is the only reason why our visual system has evolved spatiotemporal interpolation. But why then interpolation?

All our analysis has stressed that the interpolation needed for high positional acuity can be regarded as a specific filtering of the visual image. A filter that correctly interpolates the visual input automatically avoids any defect in the neural representation of the visual image, since it reconstructs the original input. It avoids in particular motion smear (see Burr 1980); also it 'fills in' the eventual gaps, in either space or time, where or when the retinal input is missing. Quite interestingly uniform velocity, assumed in the filtering scheme that we favour, is

a necessary requirement for avoidance of motion smear (Burr 1979*b*). It is thus possible, extending a suggestion of Barlow (1979), that the goal of interpolation is broader than simple precision and involves some 'correction' in the visual representation of the visual input by enforcing what we may call 'continuity of the visual field' in space and time. Whether such ideas may be developed in detail remains of course to be seen.

Appendix 1

We give here the complex Fourier spectrum corresponding to the various experimental stimuli (see figure 8).

(*a*) If a pattern $g(x)$ moves at constant velocity $-v$ along the x axis, the resulting spatiotemporal function is $g(x+vt)$ with the Fourier transform

$$\tilde{g}(f_x, f_t) = \tilde{g}(f_x)\ \delta(f_t - vf_x),$$

where $\tilde{g}(f_x)$ is the Fourier transform of $g(x)$ and $\delta(\cdot)$ is the usual delta distribution. Figure 8*a* shows the support of the function $\tilde{g}(f_x, f_t)$ on the f_x, f_t plane.

(*b*) Westheimer's experiment (see inset of figure 8*b*) corresponds to sampling on a regular array of time points the spatiotemporal input stimulus $g(x, t)$ comb $(t/\Delta t)$, where comb $(t) = \sum_n \delta(t-n)$ is the sampling function consisting of an array of delta functions spaced at intervals of width 1. Burr's experimental situation (see inset of figure 8*c*) corresponds to sampling on a regular array of space points at intervals Δt the spatiotemporal function $g(x+vt)$. Thus the image is $g(x, t)$ comb $(x/\Delta x)$. The corresponding complex Fourier spectra $[\tilde{g}(f_x)\,\delta(f_t - vf_x)] *$ comb $(f_t \Delta t)$ and $[\tilde{g}(f_x)\,\delta(f_t - vf_x)] *$ comb $(f_x \Delta x)$ are represented in figure 8*b* and figure 8*c*.

(*c*) The stimulus of figure 6 corresponds to

$$[\tilde{g}(f_x)\,\delta(f_t - vf_x)] * \text{comb}\,(f_x \Delta x) = \sum_n \tilde{g}(f_x - n/\Delta x)\,\delta[f_t - v(f_x - n/\Delta x)]$$

for the upper segment of the vernier pattern and

$$[\tilde{g}(f_x)\,\delta(f_t - vf_x)] * [\text{comb}\,(f_x \Delta x) \exp(-\mathrm{i}2\pi\,\delta x f_x)]$$
$$= \sum_n \tilde{g}(f_x - n/\Delta x)\,\delta[f_t - v(f_x - n/\Delta x)] \exp(-\mathrm{i}2\pi\,\delta x\, n/\Delta x)$$

for the lower segment (δx is the spatial offset between the upper and the lower segment in the vernier). For a comparison Burr's stimulus corresponds to the same spectrum for the upper segment and

$$[\tilde{g}(f_x)\,\delta(f_t - vf_x) \exp(-\mathrm{i}2\pi\,\delta x f_x)] * \text{comb}\,(f_x \Delta x)$$
$$= \sum_n \tilde{g}(f_x - n/\Delta x)\,\delta[f_t - v(f_x - n/\Delta x)] \exp[-\mathrm{i}2\pi\,\delta x\,(f_x - n/\Delta x)]$$

for the lower segment. In the latter case all terms, for every n, contain information about the offset and especially for $n = 0$ (for $n > 0$ the offset information that a channel can 'see' is poor, because the corresponding low frequency phase factor can be badly detected in the frequency modulated side lobes). In the previous case only the term with $n = 0$ shows a precise compensation of spatial and temporal

offset; all other terms carry information on the offset between the upper and lower gratings of the 'window' of figure 8*d*. Notice that Westheimer stimulus gives

$$\tilde{g}(f_x, f_t) = \tilde{g}(f_x) \sum_n \delta[f_t - v(f_x - n/\Delta x)]$$

for the upper segment and

$$\tilde{g}(f_x, f_t) = \tilde{g}(f_x) \exp(-i2\pi\, \delta x f_x) \sum_n [f_t - v(f_x - n/\Delta x)]$$

for the lower segment where all terms, for each n, contain the same information about the spatial offset.

Appendix 2

We consider a one dimensional pattern $g(x)$. Arbitrary, non-rigid movement of this pattern produces a spatiotemporal image $g(x, t)$. Rigid movement of the same pattern at constant speed $+v$ gives an image $g(x, t) = g(x - vt)$. We state here the classical sampling theorem for the first case and an appropriate modification of it for the second case.

Theorem 1 (classical sampling theorem). If a signal $g(x, t)$ is bandlimited in spatial and temporal frequencies it can be recovered exactly by independent interpolation in space and time of its sampled values, provided that the sampling separations $\Delta\xi$ and $\Delta\tau$ are such that $\Delta\xi < 1/(2f_x^c)$, $\Delta\tau < 1/(2f_t^c)$, where f_x^c and f_t^c are the spatial and temporal cutoff frequencies.

Theorem 2 (Crick *et al.* 1981). Assume that the spatiotemporal signal $g(x, t) = g(x - vt)$. The function g can then be reconstructed at the desired resolution from its spatial samples. The required sampling density can be decreased arbitrarily by knowledge of the velocity v. If only the sign of the velocity is available the maximum sampling distance can be twice the classical limit for stationary patterns.

Comments

(*a*) The proof of these results can be easily obtained from diagrams in the f_x–f_t Fourier plane (see figure 7; Crick *et al.* 1981).

(*b*) Theorem 1 requires the function $g(x, t)$ to be bandlimited before sampling takes place, since overlap of the frequency lobes as an effect of sampling usually leads to an irretrievable loss of information. This condition is not needed in theorem 2. Overlap never occurs (for infinitely long motion) even when the pattern $g(x)$ is not bandlimited in spatial frequency. Any desired part of the original spectrum can be recovered exactly (without aliasing) by an appropriate interpolation filter.

(*c*) The spatiotemporal filter implementing the interpolation depends on v. Assume, for instance, to endow an interpolation scheme with direction selective properties (i.e. to use information about the sign of v): it can be shown that the new spatiotemporal impulse response is obtained by adding to the original

spatiotemporal impulse response its Hilbert transform with a sign controlled by the sign of v (as in the case of figure 7, the Hilbert transform of a symmetric spatial point spread function is an odd function).

We thank H. Barlow, D. Burr, F. Crick, G. Palm, W. Reichardt, W. Richards, S. Ullman, G. Westheimer and especially B. Rosser for many useful comments and for reading various versions of the manuscript. We also wish to thank our subjects A. Koch, T. Voigt, and H. Wallis. H. Wenking and M. Herre built most of the electronic equipment. We are grateful to G. Weinraub and L. Heimburger for drawing the figures.

References

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

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

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

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

Burr, D. C. 1979*b* On the visibility and appearance of objects in motion. D.Phil. thesis, University of Cambridge.

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

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

Crick, F. H. C., Marr, D. C. & Poggio, T. 1981 An information-processing approach to understanding the visual cortex. In *The organization of the cerebral cortex* (ed. F. O. Schmitt), pp. 505–533. Cambridge, Massachusetts: M.I.T. Press. (Also available as *M.I.T. A.I. Memo* no. 557 (1980).)

Kelly, D. H. 1979 Motion and vision. II. Stabilized spatio-temporal threshold surface. *J. opt. Soc. Am.* **69**, 1340–1349.

Marr, D. 1982 *Vision.* San Francisco: Freeman. (In the press.)

Marr, D. & Hildreth, E. 1980 Theory of edge detection. *Proc. R. Soc. Lond.* B **207**, 187–217.

Marr, D., Poggio, T. & Hildreth, E. 1980 Smallest channel in early human vision. *J. opt. Soc. Am.* **70**, 868–870.

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

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

Morgan, M. J. 1979 Perception of continuity in stroboscopic motion: a temporal frequency analysis. *Vision Res.* **19**, 491–500.

Morgan, M. J. 1980*a* Spatiotemporal filtering and the interpolation effect in apparent motion. *Perception* **9**, 161–174.

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

von Seelen, W. 1973 On the interpretation of optical illusions. *Kybernetik* **12**, 111–115.

Stigmar, G. 1971 Blurred visual stimuli. II. The effect of blurred visual stimuli on vernier and stereo acuity. *Acta ophthal.* **49**, 364–379.

Ullman, S. 1979 The interpretation of visual motion. Cambridge, Massachusetts: M.I.T. Press.

Watson, A. B. 1979 Probability summation over time. *Vision Res.* **19**, 515–522.

Watson, A. B. & Nachmias, J. 1977 Patterns of temporal interaction in the detection of gratings. *Vision Res.* **17**, 893–902.

Westheimer, G. 1954 Eye movement responses to a horizontally moving visual stimulus. *Archs Ophthal., N.Y.* **52**, 932–941.

Westheimer, G. 1976 Diffraction theory and visual hyperacuity. *Am. J. Optom. physiol. Opt.* **53**, 362–364.

Westheimer, G. & Hauske, G. 1975 Temporal and spatial interference with vernier acuity. *Vision Res.* **15**, 1137–1141.

Westheimer, G. & McKee, S. P. 1975 Visual acuity in the presence of retinal-image motion. *J. opt. Soc. Am.* **65**, 847–850.

Westheimer, G. & McKee, S. P. 1977*a* Integration regions for visual hyperacuity. *Vision Res.* **17**, 89–93.

Westheimer, G. & McKee, S. P. 1977*b* Spatial configurations for visual hyperacuity. *Vision Res.* **17**, 941–947.

Westheimer, G. & McKee, S. P. 1979 What prior uniocular processing is necessary for stereopsis? *Invest. Ophthal. vis. Sci.* **18**, 614–621.

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

Wilson, H. R. & Giese, S. C. 1977 Threshold visibility of frequency gradient patterns. *Vision Res.* **17**, 1177–1190.

Wülfing, E. A. 1892 Ueber den kleinsten Gesichtswinkel. *Z. Biol.* **29**, 199–202.