MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 675 May, 1982

Zero-crossings and Spatiotemporal Interpolation in Vision: aliasing and electrical coupling between sensors

T. Poggio, H.K. Nishihara & K.R.K. Nielsen

Abstract: We will briefly outline a computational theory of the first stages of human vision according to which (a) the retinal image is filtered by a set of centre-surround receptive fields (of about 5 different spatial sizes) which are approximately bandpass in spatial frequency and (b) zero-crossings are detected independently in the output of each of these channels. Zero-crossings in each channel are then a set of discrete symbols which may be used for later processing such as contour extraction and stereopsis. A formulation of Logan's zero-crossing results is proved for the case of Fourier polynomials and an extension of Logan's theorem to 2-dimensional functions is also proved. Within this framework, we shall describe an experimental and theoretical approach (developed by one of us with M. Fahle) to the problem of visual acuity and hyperacuity of human vision. The positional accuracy achieved, for instance, in reading a vernier is astonishingly high, corresponding to a fraction of the spacing between adjacent photoreceptors in the fovea. Stroboscopic presentation of a moving object can be interpolated by our visual system into the perception of continuous motion; and this "spatiotemporal" interpolation also can be very accurate. It is suggested that the known spatiotemporal properties of the channels envisaged by the theory of visual processing outlined above implement an interpolation scheme which can explain human vernier acuity for moving targets.

We consider, in particular, the problem of avoiding aliasing in the perifoveal visual field. It is conjectured that gap junctions (or another form of coupling) between rods and cones are needed to avoid aliasing outside the fovea. Possible implications for machine vision and imaging devices are briefly discussed.

Acknowledgement. This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology . Support for the Laboratory's artificial intelligence research is provided by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643; this work was supported in part by National Science Foundation Grant MCS-79-23110, and travel for collaboration purposes was supported under NATO grant number 122.82.

In the last seven years a new computational approach has led to promising advances in the understanding of visual perception. This approach, which may be relevant not only for the information sciences but also for the neurosciences, is mainly due to the late D. Marr and his colleagues. In this article we will briefly describe this computational theory for the very first stages of vision, since it provides an useful framework for approaching the problem of spatiotemporal acuity in human vision, which is the main topic of the paper.[1]

1.1 A Computational Approach

The central tenet of this approach is that vision is primarily a complex information processing task, with the goal of capturing and representing the various aspects of the world that are of use to us. It is a feature of such tasks, arising from the fact that the information processed in a machine is only loosely constrained by the physical properties of the machine, that they must be understood at different, though interrelated, levels. This framework, formulated by Marr & Poggio (1976), was not new: H. Simon and especially L. Harmon emphasized a similar point of view in a more general context.

In a process like vision it is useful to distinguish three levels over which one's descriptions and explanations of the process must range: a) computational theory, b) algorithm, c) implementation. These are not hard and fast divisions. The important point is that no explanation or set of explanations is complete unless it covers this range. To avoid possible misunderstandings, we wish to stress that this computational approach is not a substitute for the "traditional" methods and techniques of the neurosciences to which it is in fact complementary. It is probably fair to say that most physiologists and students of psychophysics have often approached a specific problem in visual perception with their personal "computational" prejudices about the goal of the system and why it does what it does. With few exceptions this heuristic attitude, although useful, remained at the level of prejudices; computational analysis was not a science, nor was it appreciated in the neurosciences that one was needed.

[1]Some of the material for this paper has been drawn from Poggio (1981) and Fahle and Poggio (1981).

This state of affairs is hardly surprising. The difficulties of the vision process are often not appreciated even now. Until the early 70's the field of computer science and artificial intelligence failed to realise that problems in vision are difficult. The reason, of course, is that we are extremely good at it, but in a way which cannot be subjected to careful introspection. Today we know that the problems are profound. "Ad hoc" methods and tricks have consistently failed. Marr realized what the message was. A science of visual information processing was needed to analyze a given information processing task and its basis in the physical world. Marr's work, from the breadth of the approach to its rigorous detail in the analysis of specific problems, provides a methodological lesson for this new field.

1.2 The Detection of Intensity Changes

In this section we will outline one of the very first stages in the processing of visual information, the computation of zero-crossings. The basic ideas, outlined by Marr in a paper (1976), have evolved into a scheme (Marr & Poggio, 1977) based on bandpass filtering of the image through difference of gaussians and detection of the associated zero-crossings. Marr and Hildreth (1980) have provided a number of attractive arguments for justifying this scheme from a computational point of view, although a complete formal theory is still lacking. We will outline here their main points.

The goal of the first step of vision is to detect changes in the reflectance of the physical surfaces around the viewer or in the surface orientation and distance. On various computational grounds, sharp changes in the image intensity turn out to be the best indicator of most physical changes in the surface. In natural images, intensity changes can and do occur over a wide range of spatial scales. It follows that their optimal detection requires the use of operators (that is filters) of different sizes. A sudden intensity change like an edge gives rise to a maximum or a minimum in the first derivative of image intensites or equivalently to a zero-crossing in the second derivative. Marr and Hildreth (1980) argue that the desired filter should take the second derivative of the image at a particular scale. A convenient choice for the derivative in two dimensions is the Laplacian $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$, and

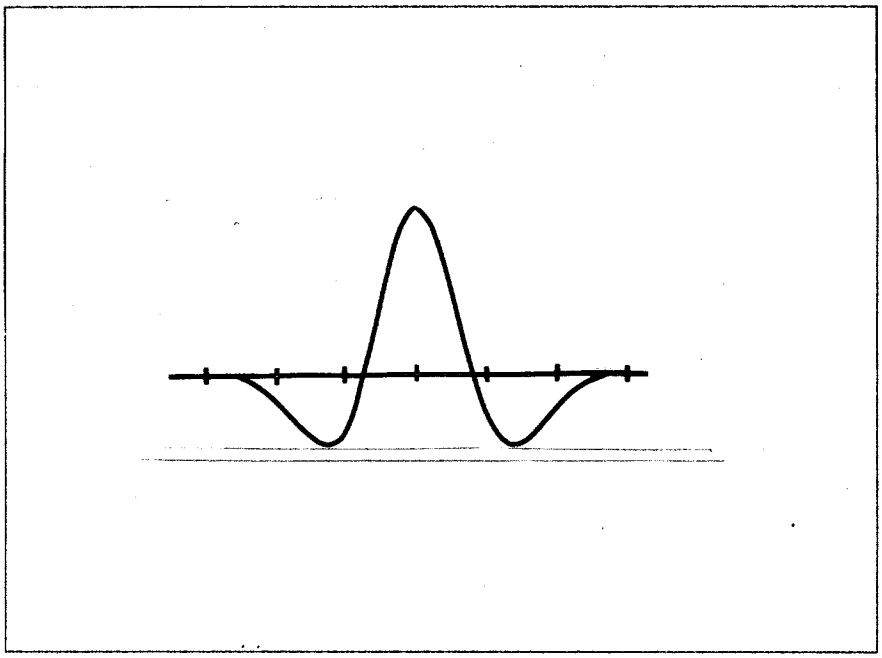

Figure 1. A cross-section of the circularly symmetric centre-surround receptive field $\nabla^2 G$.

the appropriate scale can be set by filtering the image with a 2-D Gaussian filter G, which optimally satisfies specific constraints on the real world, particularly the fact that intensity changes arising from physical objects are spatially localized at their own scale. Since the operations of taking the derivative and blurring an image are linear, the overall transformation is equivalent to convolving the image with the Laplacian of a gaussian distribution, that is with $\nabla^2 G$. As shown by fig.1, this corresponds to a centre-surround type of receptive field. Such a filter closely resembles the usual descriptions of the ganglion cell receptive field and of the psychophysical channels in human vision as the difference of two gaussians, an excitatory and an inhibitory one. Spatial filters with the centre-surround organization shown in fig. 1, are of course bandpass in spatial frequency, although their bandwidth is not very narrow.

In summary, the process of finding intensity changes at a given scale consists of filtering the image with a centre-surround type of receptive field, with a size reflecting the scale at which the changes have to be detected, and then locating the zero-crossings in the filtered image (see fig.2).

To detect changes at all scales, it is necessary only to add other channels, of different dimension,

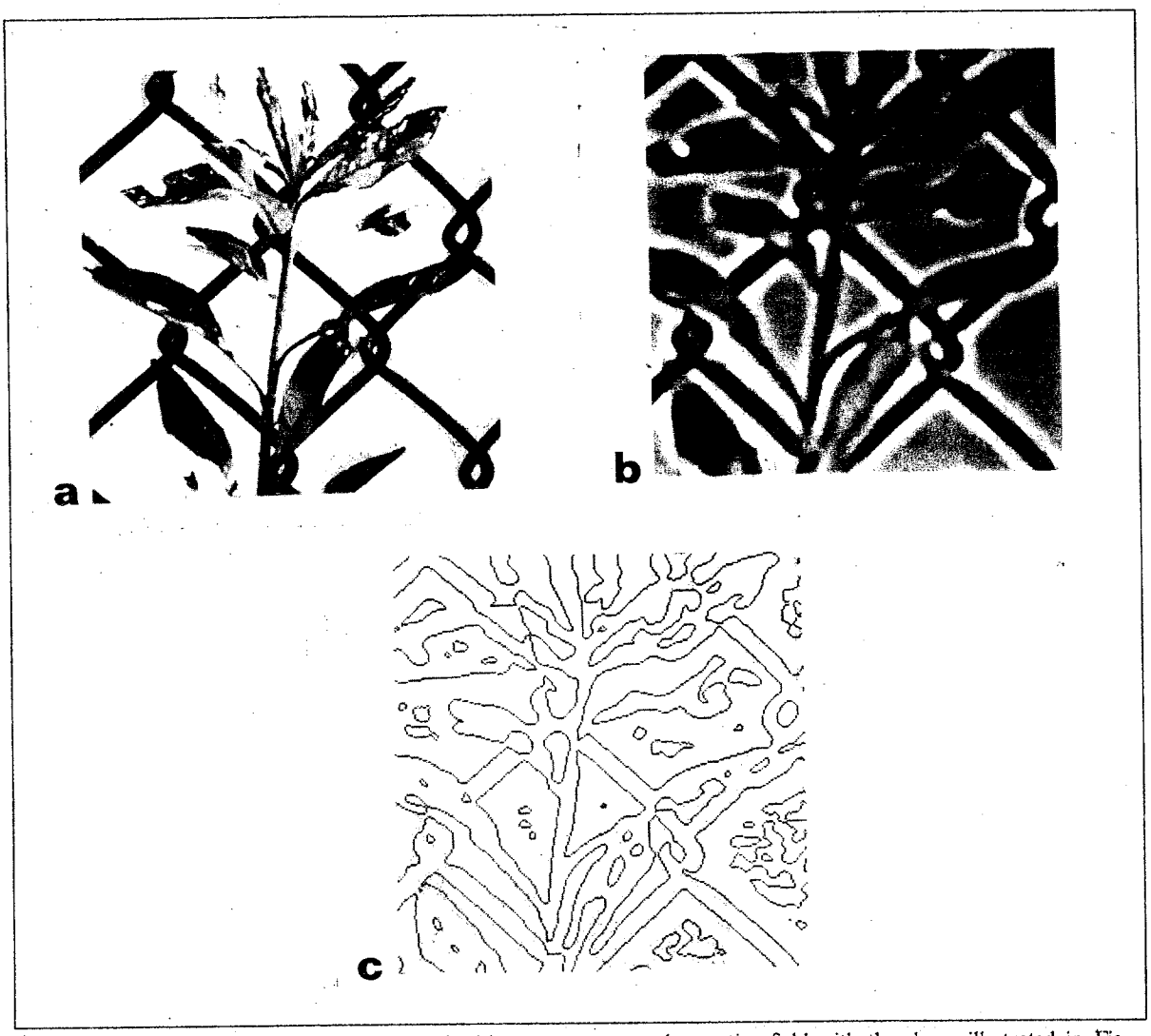

Figure 2. The image (a) has been convolved with a centre-surround receptive field with the shape illustrated in Fig. 1. (b) shows the convolved image: positive values are shown white and negative black; white (black) values would then represent the activity of the corresponding on-(off-) centre ganglion cells "looking" at the image. (c) the zero-crossings profile contains rich information about the filtered image (b) as explained in the text. Similar independent filters of smaller and larger sizes are needed to capture the whole information contained in (a). From Marr and Hildreth (1980).

and carry out the same computation for each channel independently.

Zero-crossings in each channel thus form a set of discrete symbols which are used for later processing such as stereopsis (Marr & Poggio, 1977). Marr and Hildreth, in particular, addressed the problem of how to combine zero-crossings from different channels into primitive edge elements taking advantage of physical constraints obeyed by the visual world. These and other symbolic descriptors then represent what Marr called the "raw primal sketch". Instead of describing these parts of the

theory, we shall discuss in more detail the zero-crossing detection process and the corresponding physiological and psychophysical evidence. Zero-crossings in the output of centre-surround channels represent a natural way of obtaining a discrete, symbolic representation of the image from the original "continuous" intensity values. Some recent deep results in complex analysis by B.Logan (1977) seem to support this scheme in a way which we found intriguing and fascinating from when we came across his remarkable paper. His main theorem (see Appendix 1a) states that a bandpass one dimensional signal with a bandwidth of less than 1 octave can be reconstructed completely up to a constant multiplication factor from its zero-crossings alone (if some relatively weak conditions are satisfied). From the point of view of visual information processing there is clearly no need to reconstruct the original signal. But the theorem suggests that the "discrete" symbols provided by zero-crossings are very rich in information about the original image. Unfortunately, more definite claims are as yet impossible, since an extension of the theorem to images (Appendix 1a and especially 1b; see also Marr et al., 1979) does not characterize completely the two-dimensional problem. In addition, centre-surround receptive fields are not ideal bandpass filters, as required by Logan's version of the theorem (see Appendices 1a, 1b). Clearly zero-crossings alone do not contain all the information (such as absolute intensity values), but as one of us has found in an empirical investigation, natural images filtered with $\nabla^2 G$ operators can be reconstructed to a good approximation from their zero-crossings and slopes. A successful extension of the Logan type of analysis to two-dimensional patterns may therefore represent one of the critical steps for perfecting this computational analysis of low level vision into a solid theory.

1.3 The Line Detectors/Fourier Analysis Controversy: A New Synthesis?

The previous ideas based on Logan's type of results not only lead to a satisfactory scheme for the analysis of intensity changes in an image; they also have fascinating implications for visual psychophysics and physiology, since they seem to account for basic properties of the first part of the

visual pathway. In particular these ideas explain why the image is filtered early on by approximately bandpass centre-surround receptive fields; they make more precise the notion of "edge-detectors" for extracting a symbolic description which contains full information about the image; and they state that this can be achieved only if the image was previously filtered with several independent bandpass channels — i.e. centre-surround receptive fields. As an immediate consequence these ideas also provide a solution of the long-standing controversy about edge-detectors versus frequency channels in the psychophysics and physiology of primate vision. The first stage of vision would indeed be performed to a good extent by "edge" detectors — actually zero-crossing detectors — and certainly not by Fourier analyzers; but in order for the zero-crossing detectors to extract meaningful information it is necessary that they operate on the output of independent channels, roughly bandpass in spatial frequency.

Many results from the psychophysics and physiology of early vision can be easily interpreted in this new framework. It is, for instance, not too unreasonable to propose that the $\nabla^2 G$ filtering stage is performed by ganglion cells of the retina and LGN, whereas a subclass of simple cells may represent oriented zero-crossing segments. In this context it is not important how this is implemented in detail: one of the several possibilities is that simple cells may read the zero-crossings profile from the fine grid of small cells in layer 4C of the striate cortex, where a reconstruction of the filtered image, at different scales, may be performed (via intracortical inhibition) with the goal of providing a very accurate position of the zero-crossings (see later).

Several gaps have still to be filled in the computational theory of zero-crossings. For instance, since zero-crossings do not represent the complete information about the image, it is important to characterize the other primitives that are needed. At the other levels of explanation experimental evidence in favour or against zero-crossings is of course highly desirable. Since the summer day in Tubingen where D. Marr with one of us first formulated the idea of zero-crossings in the output of independent, roughly bandpass filters, we cannot help feeling that its experimental validation — or falsification — is of critical importance for further developments of our approach to low-level vision.

2. Visual information processing: why spatiotemporal interpolation?

Any visual processor with human-level performance must be capable of analyzing time-varying imagery. The analysis starts with the spatio-temporal interpolation of the raw visual input. The spatial resolution of the photosensitive image available for processing is limited by the sampling density of the photosensitive elements in the sensor and by noise. Image motion introduces the additional problem of temporal resolution. The limiting factors are the frame rate and the integration time determined by the sensitivity of the photosensitive elements. This is of little consequence for a stationary scene, but for moving targets it poses the problem of motion smear.

The problem of high spatiotemporal resolution can be partially overcome by using better sensors with larger arrays and higher frame rate. There are, however, technological and physical limits to the spatiotemporal resolution that can be achieved in this manner, since increasing the spatial and temporal sampling rate reduces the number of photons per sensor element per cycle. Consider that since the number x of photons is Poisson distributed, $\sigma = \frac{\sqrt{x}}{2}$. The number of distinguishable levels was estimated by Barlow (1981) to be roughly $n = 2\sqrt{x}$. Thus 8 bits of resolution ($n=256$) requires about $2^{14} = 10^5$ photons. Note that the light intensity of a *bright* surface is $10^4 cd/m^2$ and this means 10^4 photons per 50 msec per sensor, assuming a sensor efficiency similar to the human cones!

Fortunately, the performance of a given sensor can be improved by appropriate spatiotemporal interpolation schemes. As we have seen, using such processes the human visual system achieves an extremely high spatiotemporal resolution compared to the sampling density of the photoreceptors and their integration time.

In summary then, temporal acuity, spatial acuity and motion smear are different facets of the same general problem posed to a visual processor by time varying imagery. We turn now to examine how the human visual processor deals with it.

2.1 Visual acuity in human vision

Since the first measurements of vernier acuity in 1892 by Wuelfing in Tubingen, the extraordinary accuracy with which the human eye can estimate the relative positions of lines or other 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. A typical example is the acuity found in reading a vernier (see inset of fig. 8a). This can be as fine as 5" of arc (Westheimer and McKee, 1975), that is 0.02mm at 1 metre 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 of no less 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 of 60 cycles/degree).

Most remarkably of all, vernier acuity is not affected by movement at constant velocity of the target in a velocity range from $0°/sec$ to at least $4°/sec$ (Westheimer & McKee, 1975). This means that a subject can detect the relative position of two lines to within a fraction of a receptor diameter (and spacing) while the whole pattern is moving across 70 receptors in 150 msec. 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 D. Burr (1979a, see also Morgan, 1980) and is shown in the top inset of fig. 8c. 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 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. (1980), the classical sampling theorem allows a correct reconstruction of the visual input from a set of discrete samples in space and time since the LGN 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 LGN 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 and Hildreth, 1980 and Marr et al., 1979).

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. As we hinted earlier, an understanding of visual interpolation may also be quite interesting from a purely information processing point of view. High resolution, smear-free real time imagery could benefit significantly from this study of human vision. 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.

2.2 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 under the control of analog electronics . Each bar was intensified for 0.1 msec at Δt msec intervals at n successive stations horizontally displaced by a separation Δx. 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 $10cd/m^2$. During an experimental run, a target was presented every 3 seconds. Brief displays of $n \cdot \Delta t = 150$ msec,

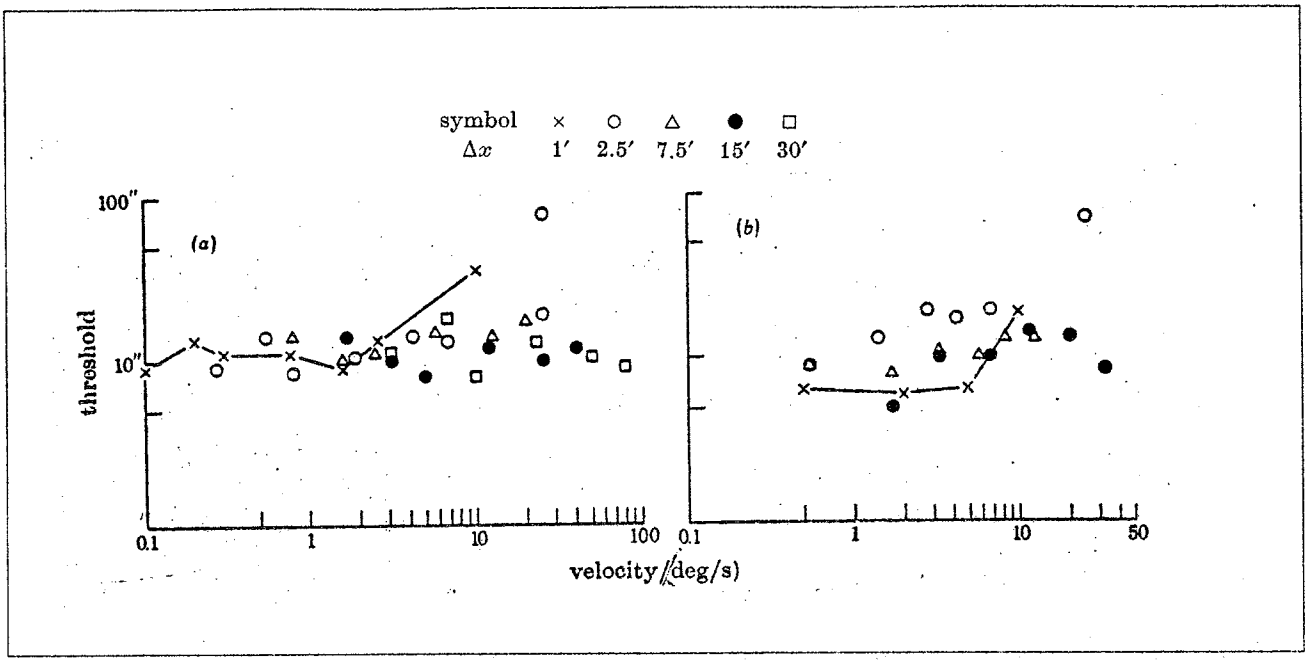

Figure 3. Vernier resolution threshold of spatial offset for different separations Δx between the stations as a function of velocity. Fig. 3a shows the data from subject AK, fig. 3b from subject TV. The standard deviation of the data is about 25%of the threshold value for fig. 1a and 20%for fig. 1b. In fig.3a the point for $\Delta x = 1'$ and $v = 10°/sec$ was measured masking the beginning and the ending of the trajectory; the same procedure did not change the threshold for the point at $v = 2.6°/sec$. Of the two points at $\Delta x = 2.5'$ and $v = 25°/sec$ in fig. 3a, the worse value has been measured under the "masking" condition way whereas the better one was measured in the standard way. In fig. 3b also the point at $\Delta x = 2.5'$ and $v = 25°/sec$ was measured with zero offset at the first and last station (from Fahle and Poggio, 1981).

with randomized direction of motion (terminating at the central fixation point)were used to prevent effective pursuit eye movements (Westheimer, 1954). 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 msec

c) the acuity for detection of mixed vernier offsets produced by a real spatial offset 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. Acuity was determined by the standard criterion of 75% correct identification. In all experiments reported here T is constant ($T = 150$ msec) and, as a consequence, the number of stations n is variable ($n = 2$ to 95). More details about the methods are given in Fahle and Poggio (1981).

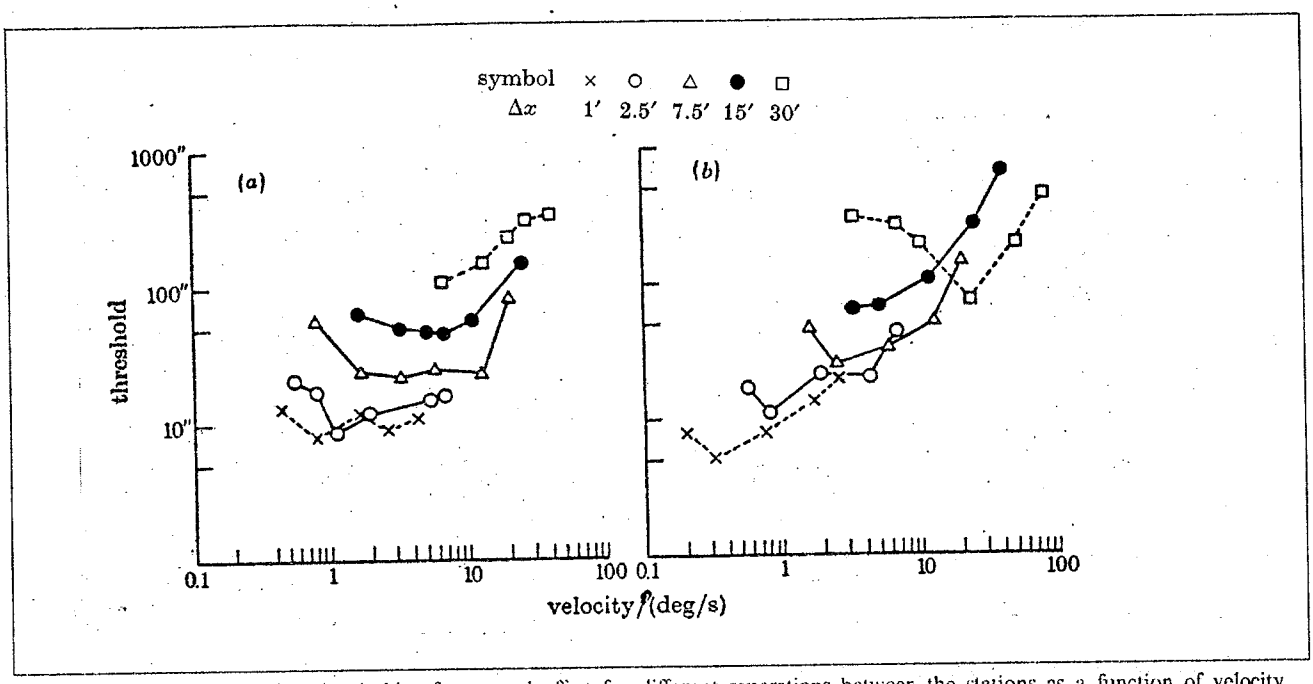

Figure 4. Vernier resolution thresholds of temporal offset for different separations between the stations as a function of velocity. Fig. 4a shows the data from subject AK, fig. 4b from subject TV. The standard deviation is about 20%of the threshold values for subject AK and 18%for subject TV (from Fahle and Poggio, 1981).

2.3 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 figs. 3a,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 Westheimer's and McKee's results (1975), which showed that vernier acuity is unaffected by rate of movement from $0°/sec$ up to $4°/sec$. Our results imply that this type of vernier acuity is relatively independent of Δt, the strobe interval.

2.4 The Temporal Type of Acuity: Dependence on v and Δx

Figs. 4a,b shows the results for temporal offsets. The accuracy of detecting the equivalent displacement is in the classical vernier acuity range (compare Burr, 1979a,b): the best value for observer AK 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

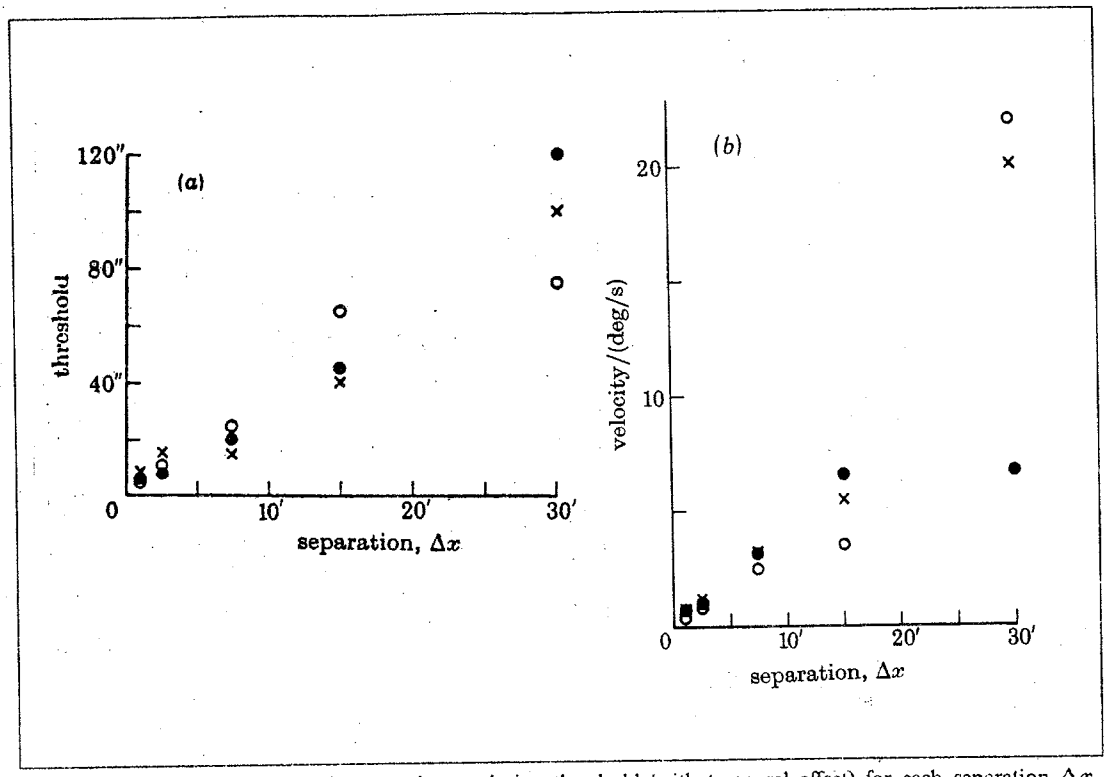

Figure 5. Fig. 5a shows the best vernier resolution threshold (with temporal offset) for each separation Δx . The data are from three subjects (partly from fig. 4a and 4b). O AK; O TV; X HW. In fig. 5b the velocity v for which optimal vernier resolution is found is plotted against the separation Δx. Same data as in fig. 5a. From Fahle and Poggio (1981).

up to half a degree, it deteriorates significantly almost in proportion to Δx(see fig. 5).

Vernier acuity of this temporal type is bad at low and high speed. As already clearly demonstrated by Burr (1979a,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 is constrained to $T = 150$ msec the stimulus consisted, at the lowest velocities, of two stations). In any case, deterioration of acuity at low velocities can be linked with a decreased sensation of motion.

A second important result is that the range of velocities for which temporal interpolation is good shifts upwards for larger separations between the stations. 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. In fact, at any separation Δx, temporal interpolation is optimal for a temporal interval Δt between 20 msec and 50 msec.

2.5 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 of the vernier pattern (see for instance Stigmar, 1971). Is temporal interpolation also degraded in the same way?

We have performed some 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 at least 15', corresponding to a cutoff frequency of around 3-4 cycle /deg. Our data show that in the experimental situation of fig. 4, blur of the pattern *improves* acuity at large separations and velocities. Fig. 6 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 case did blur of the pattern cause a reduction in temporal vernier acuity at high separations and velocities.

These data again show that temporal hyperacuity has different characteristics from spatial hyperacuity.

2.6. Spatial vs. Temporal Offset

The apparent offset δx^t produced by temporal delay δt should follow the ideal relationship $\delta x^t = v\delta t$. 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.

Fig. 7 shows that for a separation $\Delta x = 2.5'$ and a velocity $v = 1.1°/sec$ the apparent offset $\delta x^t = v\delta t$ matches rather closely the real spatial offset δx. Under these conditions spatiotemporal

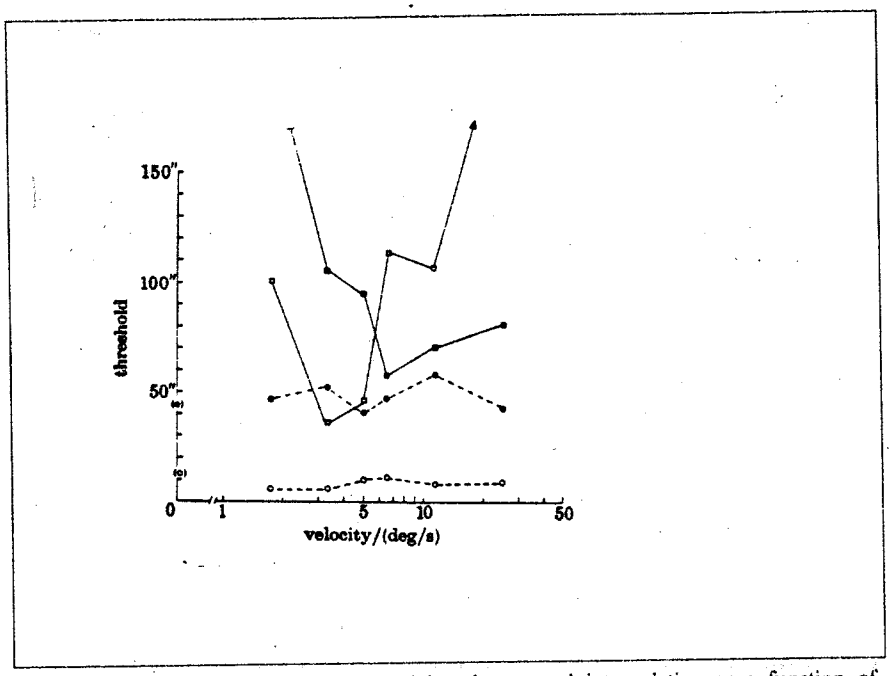

Figure 6. The effect of blur on spatial and temporal interpolation as a function of velocity for a separation between the station $\Delta x = 15'$. Vernier resolution of a spatial offset is measured with (•) and without blur (○). Vernier resolution of a temporal offset is also shown with ( ) and without ( ) blur. The screen was blurred as described in the text. Notice that the first point for spatial offset is for $v = 0°/sec$. The observer is TV. The standard deviation is about 20%of the threshold values. From Fahle and Poggio (1981).

interpolation is indeed rather precise (compare Burr and Ross, 1979). It is not so for higher velocities and/or larger separations (fig. 5). The temporal offset needed to compensate for a real spatial offset is then much larger.

3.1. Spatiotemporal Interpolation: How is it Done?

The previous results 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.

3.1.1. A Simple Illustration

Fig. 8 illustrates a very simple scheme for achieving spatiotemporal interpolation of a visual pattern.

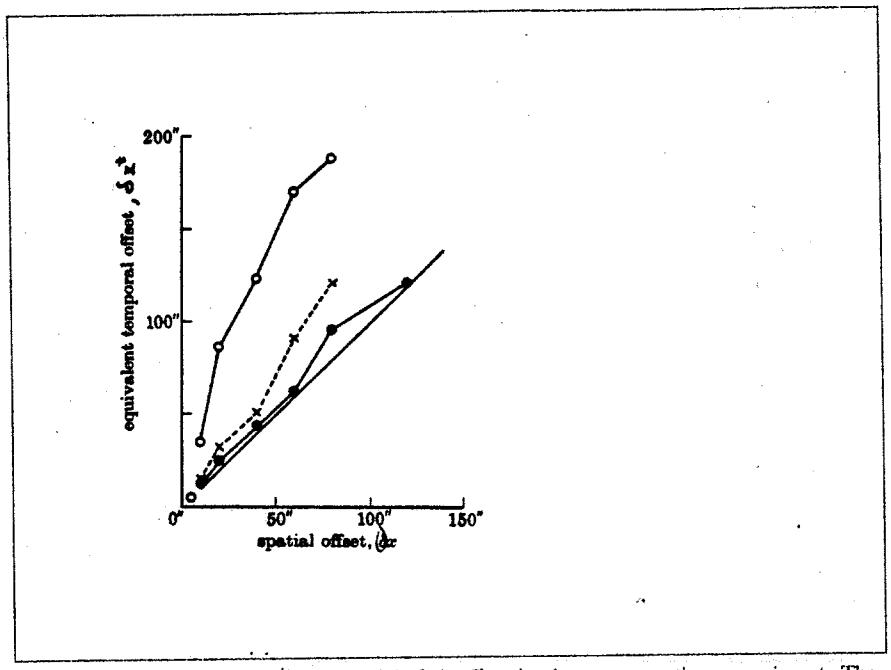

Figure 7. Temporal (δx^t) vs. spatial (δx) offset in the compensation experiment. The ordinate shows the temporal offset (in equivalent spatial units $\delta x^t = v \cdot \delta t$ needed to compensate the spatial offset shown in the abscisa. • is for a separation between the station $\Delta x = 2.5'$ and a velocity $v = 1.11°/sec(\Delta t = 37msec)$. X is for $\Delta x = 2.5'$ and $v = 5.28°/sec(\Delta t = 7.9msec)$. O is for $\Delta x = 7.5'$ and $v = 4.11°/sec(\Delta t = 30msec)$. Larger separations yield an even greater mismatch. The continuous diagonal indicates the loci of perfect compensation. Subject TV. From Fahle and Poggio (1981).

The elements of this scheme could be interpreted as cells with associated receptive fields and temporal impulse responses. Alternatively, Fig. 8 represents a computational scheme for spatiotemporal interpolation. 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 inputs according to a centre surround receptive field. A variety of filters (i.e. "receptive fields") are capable of performing a correct interpolation, especially in two spatial dimensions (see Crick et al. 1980).

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 (this scheme

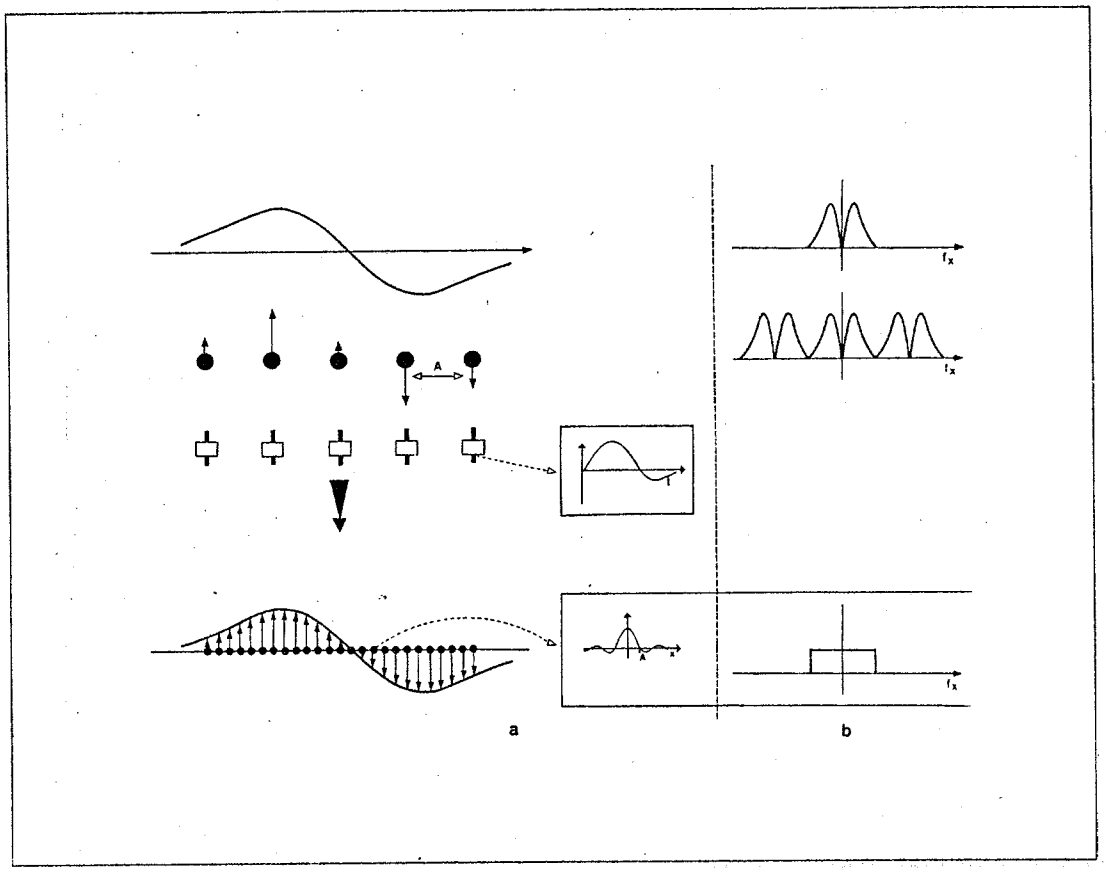

Figure 8. (a) A simple scheme for spatiotemporal interpolation. The input pattern is sampled by an array of "cells". Spatial interpolation is accomplished on a finer interpolation grid of cells each one weighting the sampled values with a sinc shaped receptive field (shown in the lower inset). Temporal interpolation is obtained by filtering with an appropriate low-pass or band-pass 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 can 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. From Fahle and Poggio (1981).

would of course operate succesfully for continuous movement of a pattern).

Fig. 8b 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, interpolation 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.

3.1.2 Remarks on Interpolation

Before embarking on an analysis of various interpolation schemes, it is appropriate to make a few general points which arise from the discussion so far.

First, the process of computing intermediate values from samples does not depend on the existence of a finer retinotopic grid of "cells", where the results are represented. All filtering transformations indicated in Fig. 8 could be carried out at a rather symbolic level for only a few distinguished points. Thus, it is important to keep separate the problem of a process from the problem of representing its output. This paper is directly concerned only with the first issue.

Second, the goal of the interpolation process may be far more modest than a full reconstruction of the input distribution. As suggested by Crick et al. (1980), the aim of interpolating the ganglion cells' activity is to 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 very simple interpolation functions such as a normal centre-surround receptive field (Marr et al., 1980).

3.1.3 More Complex Interpolation Schemes are Required

The scheme of Fig. 8 can provide a correct reconstruction of a spatiotemporal input sampled at intervals $\Delta\varsigma$ (in space) and $\Delta\tau$ (in time) only when the input function is bandlimited in spatial (by f_x^c) and temporal (by f_t^c) frequencies in such a way that $\Delta\varsigma \leq 1/2f_x^c$ and $\Delta\tau \leq 1/2f_t^c$ (theorem 1 in Appendix 2). The image which reaches the retina is indeed bandlimited in spatial frequencies to less than about 60 cycles per degree 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 Fig. 8 can therefore correctly reconstruct an image sampled at intervals of less than 30" in space (for the 2-D case see Crick et al., 1980). 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 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 Fig. 8. In particular, such models could account for the vernier acuity measured with real continuous motion of the retinal image. 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 which is already sampled either in time (Westheimer type of stimulus) or space (Burr type of stimulus) or both. We enforce arbitrary sampling intervals Δx and Δt on the system *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 Fig. 8 should for instance perform poorly when the input function is sampled in space at intervals Δx significantly coarser than the photoreceptor array. Burr's and our data, however, show that under these conditions our visual system performs significantly better. We are clearly forced therefore to consider other types of interpolation schemes.

3.2.1 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 fig. 9a with a slope equal to the velocity (see Appendix 2). For each spatial frequency contained in the pattern, there is a unique temporal frequency corresponding to it. Curtailing the duration of motion (in our case to $T = 150msec$) 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 line supports shown in fig. 9 must be interpreted as being spread along f_t as a sinc function. For $T = 150msec$ 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 2). The upper and the lower segment have the same line support on the $f_x - f_t$ plane. Their Fourier transforms differ at all frequencies only by a phase factor which mirrors the spatial offset. The correct detection of this information underlies positional acuity.

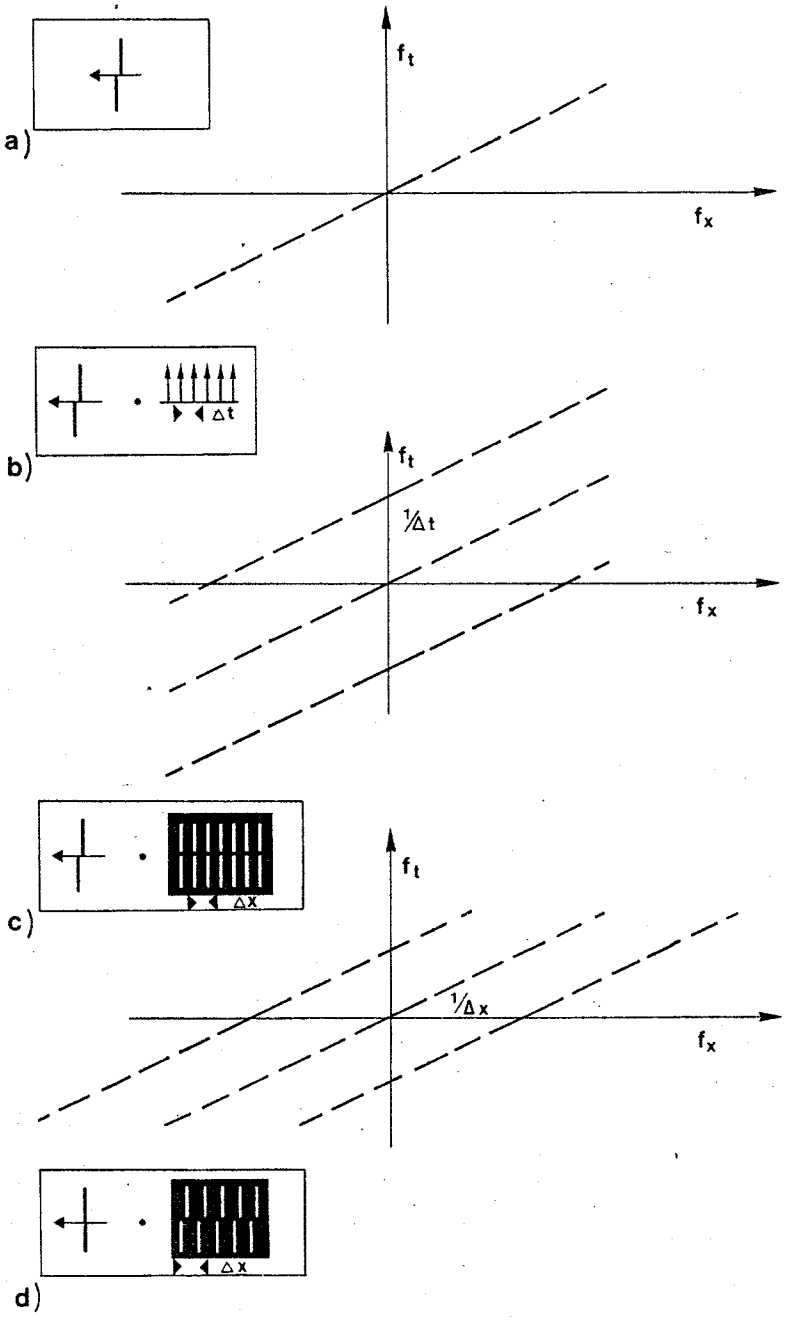

Figure 9. Legend on opposite page.

Figure 9. Legend

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. $g(f_x, f_t)$ equals $g(f_x)$ on that line. Curtailing the duration of motion to $T = 150$ msec., spreads the line into a bar-like support, corresponding to a sinc function. 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 imes 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 fig.7a 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 $1/v \cdot \Delta t = 1/\Delta x$ (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 $g(f_x)$. c) The support of the Fourier spectrum associated with Burr's type of experiment. Displaying the line segments of a vernier in 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 an otherwise opaque screen.) This corresponds to replicating the spectrum of fig.8a 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 $g(f_x)$ of different lobes is in general different. d) The support of the Fourier spectrum associated with the compensation experiment is the same as in fig.8c. The different window corresponding to this stimulus (see inset) corresponds, however, to a different complex Fourier spectrum (see Appendix 2). From Fahle and Poggio (1981).

Fig. 9 summarizes the description of the two basic stimulus configurations used in this paper according to the derivation outlined by Fahle & Poggio (1981) . 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 fig.9b). Burr's experimental situation ideally corresponds to a vernier moving behind a spatial window with a series of equidistant narrow slits (see fig.7c). The spatial or temporal windows affect differently the spectrum of the retinal input. As indicated in fig. 9, 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. An important observation is that in fig.9b (Westheimer stimulus) all lobes at any given f_x support exactly the same complex spectrum g. This is not so in fig.7c (Burr stimulus), where, instead, all lobes have the same g at any given f_t. We re-emphasize that fig. 9 describes the physical properties of the different stimuli without any reference to the human visual system.

3.2.2 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 which are relevant to the interpolation process ($T < 100msec$ and $x < 1°$). The *constant velocity assumption* leads to a more specific form of the sampling theorem, given in Appendix 2 (see also Crick et al., 1980), 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 ($x \approx vt$). From the point of view of filtering this means that spatial and temporal interpolation cannot be performed independently as in the simple scheme of Fig. 8. In the Fourier domain the constant velocity assumption constrains the spectrum of the visual input to lie on the line support shown in Fig. 9a. In the ideal case of infinitely long motion the side lobes

generated by sampling either in time (Fig. 9b) or space (Fig. 9c) 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 (Fig. 9a) 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 still apply.

3.2.3 Implementing the constant velocity scheme

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 (1979b) that these channels have somewhat overlapping supports covering the region of the $(f_x - f_t)$ Fourier plane which 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. Fig.10b 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 be effective (but not so accurate) at 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. As indicated

in Fig. 10b the stasis channels may suffer from aliasing at values of Δx for which the motion channels interpolate correctly. We assume, then, that in this scheme the wrong channels are switched off by use of velocity information.

Fig. 10c 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 any precise velocity sensitive inactivation, apart from directional selective properties.

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.

4.1 Westheimer's Acuity: Recovery of Spatial Offset

a) 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 fig. 10a, 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 fig.10c, there are one or more channels which 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 on 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; under these conditions high frequency channels,

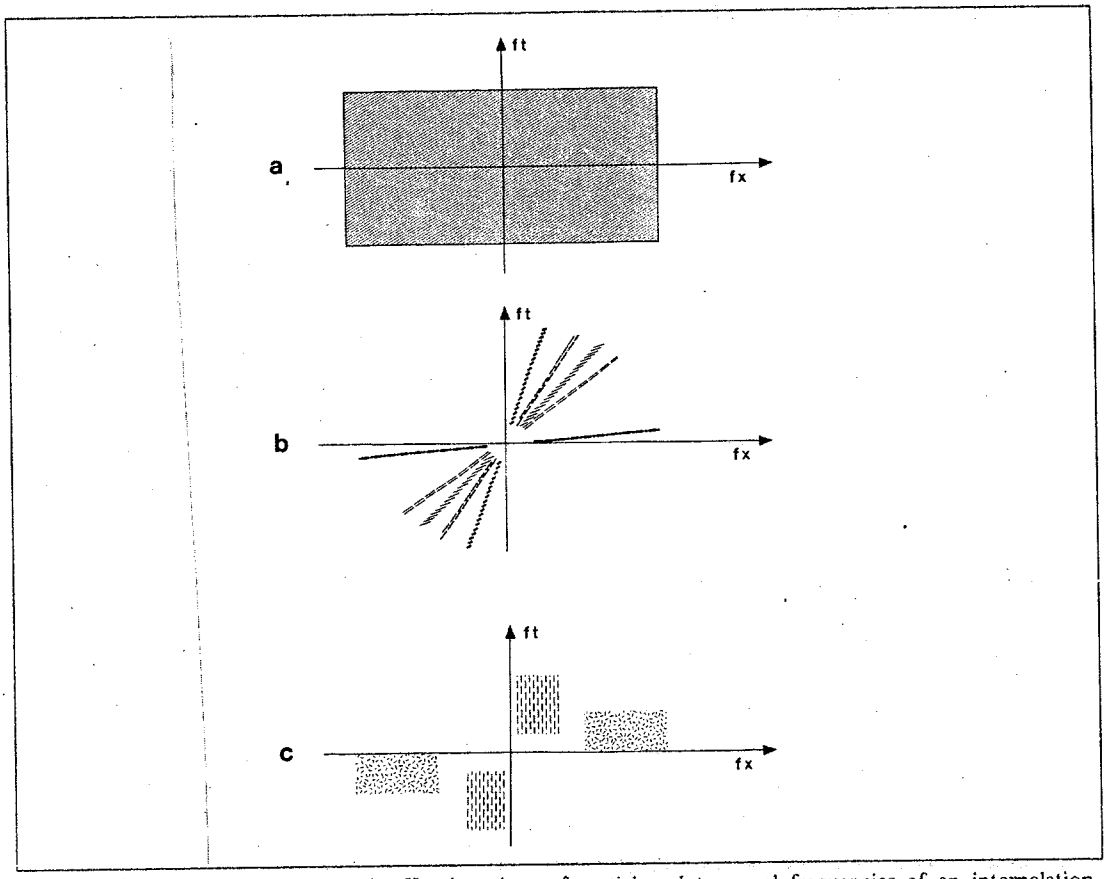

Figure 10. (a) The support on the Fourier plane of spatial and temporal frequencies of an interpolation filter corresponding to a scheme such as Fig.6. (b) The support on the Fourier plan of a set of spatiotemporal filters ideally tuned to different velocities. A large number is needed to cover all velocities 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 fig. 9c. 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 also be present. The hatched areas represent the support of such directional filters. Nondirectional filters would have also a symmetric support in the other two quadrants. From Fahle and Poggio (1981).

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. On the whole, and in the absence of a sophisticated interpolation process that always excludes all side lobes (such as the scheme of fig. 10b), one expects vernier acuity to be rather invariant for a wide range of separations and velocities. 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).

b) 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.

c) These results suggest that spatiotemporal interpolation is not performed by the "ideal" interpolation scheme of Fig. 10b. 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. The one channel scheme of Fig. 10a could explain these data on positional acuity; but as pointed out by Burr (1979b, 1980) the image should then be inevitably smeared at all but very low velocities.

4.2 Burr's Acuity: Interpolation of Temporal Offset

a) 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 (in fig.10a or fig.10c) there is a mixture of spatial frequencies, detrimental to acuity. One understands, therefore, that acuity deteriorates considerably (see fig. 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 separation Δx, an optimal velocity at which the side lobes just avoid overlap. Assuming a spread of $\approx 15Hz$ the optimal velocity (in degree/sec) should be $v = 30 \cdot \Delta x$ (Δx in degrees), which is in rough agreement with the data of fig. 5b. When the velocity approaches zero the line supports in fig. 10c 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 line spectrum of a stationary grating with no offset. Notice that we assume for the scheme of fig 10c that the vernier threshold is higher when some of the channels signal zero offset while the others still "see" the correct offset.

b) 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. The critical strobe interval which yields optimal temporal interpolation is not very different between the channels (see Fig. 5a). Though its performance may worsen at high velocities, as for the continuous motion, it should be rather invariant with respect to Δx, the separation between the stations. Fig. 5a shows that this does not happen. The opposite conclusion holds for the scheme of Fig. 10a. Its performance should deteriorate rapidly for separations Δx between the stations larger than the distance between photoreceptors, which is in conflict with Burr's and our data. An interpolation scheme of the type of Fig. 10c 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 are direction selective.

4.3 Effect of Blur

a) The interpolation scheme outlined in fig.10c 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 which 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 velocites 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), by the blur of the screen, which suppresses the high frequency grating.

b) 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.

c) 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 the wrong channels as assumed in the scheme of Fig. 10b.

4.4 Spatial vs. Temporal Compensation

a) This stimulus situation corresponds to looking at the continuous motion of a vernier through the spatial window shown in the inset of fig. 9d. The resulting Fourier support, is again as in fig. 9c: 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 fig. 10c. 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 fig. 9d) 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 fig. 7) 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.

b) 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 interpola-

tion is not fully correct.

5. Discussion

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., 1980). 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 early vision, in terms of an interpolation scheme of the type of fig.10c. The existence of independent channels tuned to different spatial and temporal frequencies seems to account for the spatiotemporal interpolation revealed by our experiments. A detailed theoretical analysis with the help of appropriate computer experiments is necessary for a quantitative evaluation of interpolation models of this type.

5.1 Explicit or implicit interpolation?

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. Clearly, hyperacuity may simply rely on a specialized routine operating on a small region of the image to answer specific questions, like the right-left choice in a vernier task. Thus 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 the array 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 automatically on the fine topographic grid of layer IVcβ by an automatic, parallel process.

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 process itself.

In the first case, the decision routine (is the upper segment to the right or to the left?) would operate on an interpolated version of the image. Thus, "reprogramming" of the vernier routine may not be expected to affect the interpolation process but only the detection criteria, contrary to the second case, in which different detection strategies may influence interpolation.

5.2 Are the Psychophysical Channels the Interpolation Filters?

Our data support interpolation schemes of the type outlined in Fig. 10c. 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 and Robson, 1968; Burr, 1979b; see also Marr et al., 1980; Wilson and Gieze, 1977, Wilson and Bergen, 1979).

These observations suggest the interesting idea that the spatial frequency tuned channels present in early human vision may be the interpolation filters *themselves.* To be completely explicit let us consider simple examples of how an interpolation scheme such as Fig. 10c might be implemented in the visual system. The first possibility is that the image is filtered before interpolation through various independent channels. Retinal or LGN 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 very similar to the corresponding LGN 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 LGN but a scatter of sizes in the

cortex (possibly in IVc). 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 may have a narrower bandwidth than retinal ones.

5.3 A prediction: interpolation must be direction selective

An explicit interpolation scheme of this type consists of a set of motion channels with direction selective properties, in the sense that the spatiotemporal interpolation filter thereby implemented must depend (in one dimension) on the sign of v (see appendix of Fahle and Poggio, 1981). As a consequence the interpolation channels should have some type of direction selective property; furthermore, cells of layer IVc -if they are involved at all - should show, despite their center-surround receptive field, some non-standard direction selective property.

6. Interpolation in the perifoveal visual field: does aliasing occur?

In the perifoveal retina, the spacing of the ganglion cells increases, as Barlow pointed out, whereas the optical cut-off remains approximately the same (for instance at 10° eccentricity; see Weale, 1976). The grid of ganglion cells is, however, matched to the spatial cut-off of the signal thereby represented: in the cat, Peichl and Wassle (1979) have shown that receptive field diameter and ganglion cell separation both increase towards the periphery so that sampling in the array of ganglion cells takes place at the interval appropriate to the cut-off frequency passed by the larger receptive fields. Thus, the grid of ganglion cells is likely to satisfy the sampling theorem (see Hughes, 1981).

A more serious, and so far unsolved, problem is whether in the perifoveal visual field the signal represented by the ganglion cells suffers from aliasing, i.e., undersampling, at the level of the photoreceptors. If only cones are involved, aliasing seems unavoidable for eccentricities larger than about 5° — 10°. The classical sampling theorem requires that the signal is lowpass filtered *before* sampling in order to avoid overlap of the sidelobes in the Fourier spectrum (i.e., aliasing). Lowpass

filtering *after* sampling cannot always avoid aliasing.

It is easy to show that ideal lowpass filtering after sampling eliminates overlap of the sidelobes only up to sampling intervals that are twice the limit set by the sampling theorem.[2] Preliminary computer experiments support these conclusions for the approximately lowpass filtering performed by a center-surround receptive field; in this case, however, effectiveness of lowpass filtering decreases more gradually with increasing sampling intervals.

This scheme is somewhat supported by Poliak's data showing that visual acuity threshold increase with eccentricity more than the separation between cones. Convergence of cones on X ganglion cells is therefore likely to increase with eccentricity.

If aliasing cannot be fully avoided, hyperacuity threshold must rise faster with eccentricity than visual resolution thresholds, a result which has been recently established by Westheimer (1982). If the reason for this were indeed aliasing, blur of the vernier pattern should improve vernier acuity in the periphery, at least in the absence of noise. Blur of the pattern corresponds to lowpass filtering of the signal *before* sampling, as required by the sampling theorem. Preliminary experiments performed to test this prediction indicate, however, that blur may improve hyperacuity only slightly, if at all (Fahle and Poggio, 1981; Westheimer, pers. comm.; Fahle, pers. comm.).

A possible explanation for this small effect arises, if input from rods (in addition to cones) is also allowed. Aliasing in the periphery could then be largely avoided at all eccentricities by lowpass filtering the image *before* sampling, by pooling together inputs from *all* neighboring photoreceptors–rods and cones–*via* either gap junctions or synaptic coupling in second order neurons. If this prediction were correct, the decrease of vernier acuity with eccentricity would not depend on aliasing but would simply be a graded phenomenon due to the increasing spacing (in terms of visual angle) of the cortical grid and on a decreasing signal to noise ratio (because of the decreasing density of cells). The ineffectiveness of blur is consistent with this scheme. A critical test of this hypothesis may be

[2]This is achieved at the expense of a much more extensive loss of high spatial frequencies than in the case of lowpass filtering *before* sampling. Localization of an isolated feature like a zero-crossing is, however, rather unaffected by loss of high spatial frequencies, in the ideal case of small noise level.

obtained by measuring vernier acuity in the periphery under different conditions of light adaptation. An important corollary of this prediction is that the space constant of the electrical coupling should increase proportionally to cone spacing from the fovea to the periphery (the rod network may have interesting spatiotemporal properties (see Detwiler et al., 1978), possibly useful for moving patterns). Several morphological studies have demonstrated apparent connections between cones as well as between rods and cones in the vertebrate retina (see for instance Raviola and Gilula, 1975). Nelson (1977) has provided physiological evidence for the cat that cones have inputs from rods, probably mediated by the rod-cone gap junctions. The above conjecture would explain why coupling of this type is needed already at the level of the photoreceptors, whereas improvement of signal-to-noise ratio could be achieved in a simpler way with convergence of signals at a later level in the retina.

6.1 Significance for information processing and machine vision

There are various methods for reconstructing the original signal at high resolution by interpolating values measured at widely spaced intervals. The best known approach to this problem is based on the Shannon sampling theorem and on its various extensions. For static images interpolation of this type can provide a resolution much higher than the original sampling grid. Since in our framework the position of zero-crossings (and not the grey level values) is important, Hildreth and Poggio have examined the problem of interpolating the values of the $\nabla^2 G$ convolution in order to obtain precisely the location of zero-crossings. Analytical arguments, supported by computer experiments, have shown that the position of a zero-crossing can be interpolated precisely in terms of very simple interpolation functions, even by linear interpolation. For time-varying images the situation is more complicated. In the classical sampling theorem, interpolations in space and time are performed independently, since the temporal dependence of the input is not constrained in any way. Interpolation algorithms based on the constant velocity assumption discussed earlier could achieve higher spatio-temporal resolution for objects in motion, as long as the constant velocity assumption is not grossly incorrect, despite low spatial and temporal sampling rates. Positional acuity for the image features, e.g., the zero-crossings, although desirable, is not the only goal of this spatiotemporal interpolation stage. A filter

that correctly interpolates the sampled image automatically avoids any defect in the representation of the image since it reconstructs the "original" input. It avoids in particular motion smear; and it "fills in" eventual gaps either in space or time, where or when the sampled input is missing. Real time vision machines may well need such an interpolation stage and it will be interesting to see the form and the performance of a computer implementation. In particular, the "gap junction" scheme for avoiding aliasing with sparse sampling intervals may be usefully implemented in future CCD devices.

Acknowledgements. We are grateful to E. Grimson for reading the paper, to G. Weinraub for drawing the figures and to P. Rogers for her help with the manuscript.

Appendix 1a

Logan's results apply to $B_\infty(\lambda)$ functions, i.e., the restrictions to the real line of entire functions of exponential type λ whose growth (on $\mathbf{R}$)) is less than exponential. In particular, they apply to periodic functions with the exception of theorem 4 (Logan, 1977), which can be specialized to periodic functions (Logan, personal communication). If we restrict ourselves to trigonometric polynomials, it is possible to illustrate Logan's results in a simple way. It should be stressed, however, that trigonometric polynomials are a very special case and in general erroneous inferences can be made from their special properties. With this "caveat" in mind, let us consider the real band limited function

$$h(t) = \sum_{-N}^{N} C_n e^{int} \qquad C_n = \overline{C}_{-n} \tag{1}$$

which can be extended to the complex plane as

$$h(z) = \sum_{-N}^{N} C_n e^{inz}$$

$h(z)$ is for instance bandpass with one octave bandwidth if

$$C_n \simeq 0 \qquad |n| \quad \leq \frac{N}{2}$$

The complex free zeros of $h(z)$ are the complex zeros of $h(z)$ in common with its Hilbert transform $\hat{h}(z)$ where

$$\hat{h}(z) = \sum_{-N}^{N} \hat{C}_n e^{inz} \qquad \hat{C}_n = -i \quad sign(n) C_n \tag{2}$$

Let us define, given $h(z)$

$$P(z) = \sum_{A+1}^{N} C_n e^{inz}$$

$$N(z) = \sum_{-N}^{-(A+1)} C_n e^{inz} \tag{3}$$

where A is the low-frequency boundary of the spectrum of $h(z)$ (assumed in the following bandpass).

Then the free zeros of $h(z)$ are completely characterized by the following three equivalent formulations:

The free zeros of $h(z)$ are such z^*:

$$P(z^*) = 0 \qquad N(z^*) = 0 \tag{a}$$

$$h(z^*) = 0 \qquad P(z^*) = 0 \tag{b}$$

$$P(z^*) = 0 \qquad P(\overline{z^*}) = 0 \tag{c}$$

Observe that if z is a zero, $\bar{z}$ is also a zero of $h(z)$; and if z is a zero, $z + 2k\pi$ k an integer, is also a zero.

The coefficients C_n of $h(z)$ may be determined by the $2N$ roots of $h(z)$ as the solutions of the system of $2N$ equations

$$\sum_{-N}^{N} C_n e^{inz_1} = 0$$

...

$$\sum_{-N}^{N} C_n e^{inz_{2N}} = 0 \tag{4}$$

Let us now rewrite

$$h(z) = \sum_{-N}^{N} C_N e^{inz}$$

as

$$h(\varsigma) = \left(\sum_{0}^{2N} g_n \varsigma^n \varsigma^N \right) \tag{5}$$

with

$$\varsigma = e^{iz}, g_n = C_{n-N}, \mathbf{R}[z] = [0, \pi], N = :2M$$

Thus the nontrivial zeros of $h(z)$ coincide with the zeros of $\sum_0^{2N} g_n \varsigma^n$, that is, a polynomial of order $2N$. If the $2N$ roots ς would be known, it would be possible to write $2N$ equations in the $2N + 1$ real unknowns (C_n):

$$\sum_{0}^{2N} g_n \varsigma_1^n = 0$$

$$\sum_{0}^{2N} g_n \varsigma_{2N}^n = 0 \tag{6}$$

with

$$\varsigma = e^{iz}$$

Since the determinant of the roots is a Vandermonde determinant, it always has maximum rank if the roots are distinct. The question is under which conditions the real roots alone determine, apart from a multiplicative constant, the set of C_n, i.e. $h(z)$. Clearly, multiple zeros, in particular multiple real zeros, cannot be allowed. Observe that if more than $2N$ real zero-crossings would be available (in a basic period) then $h = 0$.

Under the bandpass condition ($C_n = 0$ for $n \leq A$) there are at least $2A$ real zero-crossings per period. The real unknowns are $2b$, $b = N - A$, that is the number of non-zero C_n between N and A, counted twice because they are complex numbers. A sufficient condition to ensure that there are enough zero-crossings, and thus equations, is $A = M = \frac{N}{2}$, i.e., C_n (for $n > 0$) all non-zero in $[M, 2M]$. Notice that $[M, 2M]$ i.e., one octave bandwidth would not be sufficient: in this case there would be at least 2M real roots but $2(M + 1)$ unknowns C_n. The matrix associated to the homogeneous equation in the "roots"

$$\begin{pmatrix} e^{-i2Mt_1} & e^{-i(A+1)t_1} & e^{i(A+1)t_1} & e^{i2Mt_1} \\ \cdots & \cdots & \cdots & \cdots \\ e^{-i2Mt_{2M}} & \cdots & \cdots & \cdots \end{pmatrix}$$

has rank at most $2M - 1$ (since there exists C_n such that $\sum C_n e^{inx}$ vanishes identically for $x = t_1 \ldots t_{2M}$) and this would just not suffice to specify the C_n modulus a multiplicative constant.

Although the less-than-1 octave condition is sufficient to ensure enough zero crossings, it is by no means necessary. In fact, there are classes of bandpass signals with a larger bandwidth and still enough zero-crossings.

In any case, even when there is a sufficient number of zero-crossings, the question still remains

of whether the determinant of the matrix of the "roots" $|e^{intz}|$ has maximum rank $(2M - 1)$ and therefore the C_n can be determined (modulus a multiplicative constant). If the rank is less than $2M - 1$ then the C_n are not uniquely determined and as a consequence $h(z)$ is not determined by its real roots. Logan (1977 and personal communication) has proved that

a) if a free zero exists then $h(z)$ is not uniquely determined by its real roots and

b) if there are no free zeros, $h(z)$, provided its bandwidth is appropriate, is determined, modulus a multiplicative constant, by its real zero-crossings.

In the following, we will outline Logan's main theorems for the case of trigonometric polynomials.

Theorem 1

If $h(z)$ has 1 or more free zeros, the rank r of the determinant of the roots is $r < 2M - 1$.

Proof

$h(t)$ can be written as

$$\begin{aligned} h(t) &= P(t) + N(t) \\ &= e^{-i2Mt}\{\sum_{0}^{M-1} g_n e^{int}\} + e^{i(M+1)t}\{\sum_{0}^{M-1} P_n e^{int}\} \\ &= e^{-i2Mt}\prod^{M-1}(e^{it} - e^{i\delta_J}) + e^{i(M+1)} + \prod^{M-1}(e^{it} - e^{i\delta_J}) \end{aligned} \tag{8}$$

If ϵ is a free zero of $h(t)$ then we can divide $h(t)$ by the real function

$$f(t) = (e^{it} - e^{i\epsilon})(e^{it} - e^{i\bar{\epsilon}}) = (2ie^{\frac{it+\epsilon}{2}} \sin\frac{t-\epsilon}{2})(2ie^{\frac{it+\bar{\epsilon}}{2}} \sin\frac{t-\bar{\epsilon}}{2}) = A \sin\frac{t-\epsilon}{2} \sin\frac{t-\bar{\epsilon}}{2} \tag{9}$$

with A real.

The resulting $\frac{h(t)}{f(t)}$ is still a periodic bandpass function of the form

$$\frac{h(t)}{f(t)} = \sum_{-2M}^{-M} S_n e^{int} + \sum_{M}^{2M} S_n e^{int} \tag{10}$$

and actually of reduced bandwidth. Multiplication of $\frac{h(t)}{f(t)}$ by any arbitrary $[a - \cos(t - \sigma)], a > 1$ which can be always written as $C \sin \frac{t-\gamma}{2} \sin \frac{t-\bar{\gamma}}{2}$, provides a periodic bandpass function with the same bandwidth as the original $h(t)$ but different from it despite the same real zeros. Notice that if ϵ is not a free zero, $\frac{h(t)}{f(t)}$ will no longer be a periodic bandpass function. This means that the determinant associated with the homogeneous equation 7 has at most rank $r = 2M - 2$.

Theorem 2

If $h(t)$ has no multiple and no free zeros the rank of the determinant of the real "roots" is $r = 2M - 1$.

Proof

Clearly r cannot be $r > 2M - 1$. If h_1 and h_2 have the same bandwidth and the same real zeros, then

$$h_1 h_2 + \hat{h}_1 \hat{h}_2 = \sum_{0}^{2M-1} g_n e^{int} \tag{11}$$

$$h_1 h_2 - \hat{h}_1 \hat{h}_2 = \sum_{0}^{2M-1} P_n e^{int} \tag{12}$$

as it is easy to check by substitution of equation (2). If the real zeros are $2M$ in number and distinct, the Vandermonde determinant associated to the real roots of equation 12 is different from zero; thus, the unknowns g_n are identically zero. The same argument implies that all P_n are also identically zero. Thus, $\frac{h_1}{\hat{h}_1} - \frac{h_2}{\hat{h}_2} = M(t)$.

Now $M(t)$ is any function with the same zeros (real and complex) of h_1. But h_1 is a bandlimited function $h_1(t) = \sum_{-2M}^{2M} C_n e^{int}$ which is uniquely determined (apart from a multiplicative constant) by its 4M real and complex zeros. Thus h_1 and h_2 must coincide identically and the theorem follows. The theorem can be generalized allowing for real zeros.

Finally, a short remark about the multiple and free zero condition. It is rather intuitive that multiple and free zeros are not generic; assume, for instance, that the polynomial $\sum_{-N}^{N} C_n e^{int}$ has a free zero. It is enough to perturb one of the coefficients C_n to annihilate the free zero. Similarly, if the trigonometric polynomial is a sample function of a random process, the coefficients C_n would be random numbers, as well as the zeros of the associated polynomial $\prod^{2N}(\zeta - \zeta_i)$. The probability that a zero is free (i.e. with $\zeta_i = \rho \quad e^{i\theta}$, ζ_i is free *iff* $\frac{1}{\rho} e^{i\theta}$ is also a zero) is usually very low.

Appendix 1b

Logan's result can be extended to the case of a two-dimensional entire function $f(x, y)$ if it is bandpass in x with a band-width strictly less than an octave and band-limited in y . In this case, the restriction of f to a one-dimensional line l_x in the x, y plane parallel to the x axis will be bandpass with less than an octave band-width. Provided the free-zero condition is met, Logan's theorem tells us that the zeros of f along l_x determine f there up to a multiplicative constant. To determine f everywhere up to a multiplicative constant, these parallel slices must be tied together.

The following lemma shows that Logan's theorem can be invoked for f restricted to a line l_θ which is not parallel to the X axis. l_θ will intersect all slices l_x parallel to the x axis, so determining f up to a multiplicative constant on l_θ determines f up to the same constant along each of the slices l_x.

Lemma

If $f(x, y)$ is ideally bandpass with band-width strictly less than an octave in x and band-limited in y then there is an $\epsilon > 0$ such that f along all slices, l_θ which make an angle $\theta < \epsilon$ with the X axis, will be bandpass with band-width less than an octave.

Proof

The support of the Fourier transform of f is confined in ω_x to the intervals $I_1 = (-2a + \delta, -a - \delta)$ and $I_2 = (a + \delta, 2a - \delta)$ and in ω_y to the interval $J = (-b, b)$ for some positive δ, a, and b. Observe that the support of the Fourier transform of a slice l through f is confined to the projection of the support of the Fourier transform of f onto the ω_l axis. The rectangles $I_1 \times J$ and $I_2 \times J$ will project into the intervals $(-2a, a)$ and $(a, 2a)$ on l_ω provided that l makes a sufficiently small angle with the x axis.

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 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 $\Delta\varsigma$ and $\Delta\tau$ are such that $\Delta\varsigma \leq 1/2f_x^c$ and $\Delta\tau \leq 1/2f_\tau^c$, where f_x^c and f_τ^c are the spatial and temporal bandwidths.

Theorem 2 (Crick et al., 1981; Fahle & Poggio, 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 (temporal) 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 Fig. 9; 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 $f(x)$ 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 filter is obtained by adding to the spatiotemporal impulse response its Hilbert transform with a sign controlled by the sign of v (in the case of Fig.8 the Hilbert transform of the spatial point spread function is an odd function).

References

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

Barlow, H.B., "Critical limiting factors in the design of the eye and visual cortex," *Proc. Roy. Soc. Lond. B* **212** (1981), 1-34.

Burr, D.C., and Ross, J., "How does binocular delay give information about depth?," *Vision Research* **19** (1979), 523-532.

Burr, D.C., "Acuity for apparent Vernier offset," *Vision Research* (1979a), 835-837.

Burr, D.C., "On the visibility and appearance of objects in motion," *Ph.d. Thesis* **University of Cambridge** (1979b).

Burr, D., "Motion smear," *Nature* **284** (1980), 164-165.

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

Crick, F.H.C., Marr, D.C., Poggio, T., "An Information-Processing Approach to Understanding the Visual Cortex.," *In: The Organization of the Cerebral Cortex, Ed. F. Schmitt* **M.I.T. Press** (1980).

Detwiler, P.B., Hodgkin A.L., McNaughton P. A., "A surprising property of electrical spread in the network of rods in the turtle's retina," *Nature,Lond.* **274** (1978), 562-565.

Fahle, M., Poggio, T., "Visual hyperacuity: spatialtemporal interpolation in human vision," *Proc. R. Soc. Lond. B* **213** (1981), 451-477.

Hughes, A., "Cat retina and the sampling theorem," *Exp. Brain Res.* **42** (1981), 196-202.

Kelly, D.H., "Motion and vision:II. Stabilized spatio-temporal threshold surface," *J.Opt.Soc.Am.* **69**

(1979), 1340-1349.

Logan,B.F., "Information in the zero-crossings of band pass signals," *Bell Syst. Tech. J.* **56, 487** (1977), 510.

Marr, D., "Early Processing of Visual Information," *Phil. Trans. R. Soc. Lond.B.* 275 (1976), 483-524.

Marr,D. and Hildreth,E., "Theory of edge detection," *Proc.R.Soc.Lon.B.* 207 (1980), 187-217.

Marr, D., Poggio, T., Hildreth, E., "Smallest channel in early human vision," *J. Opt. Soc. Am.* **70** (1980), 868-870.

Marr, D.C., Poggio, T., "From Understanding Computation to Understanding Neural Circuitry. In: Neuronal Mechanisms in Visual Perception," *Neurosciences Res. Prog. Bull., Eds. E. Poppel, R. Held, J.E. Dowling* **15, No. 3** (1976), 470-488.

Marr, D., Poggio, T., "A Computational Theory of Human Stereo Vision," *M.I.T. A.I. Memo 451* (1977).

Marr, D., Ullman, S., Poggio, T., "Bandpass Channels, Zero-crossings, and Early Visual Information Processing," *J. Opt. Soc. Am.* **69, No. 6** (1979), 914-916.

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

Nelson, R., "Cat cones have rod input:a comparison of response properties of cones and horizontal cell bodies in the retina of the cat," *J. Comp. Neurol.* **172** (1977), 109-136.

Peichl, L., Wassle, H., "Size, scatter and coverage of ganglion cell receptive field centers in the cat retina," *J. Physiol. Lond.* **291** (1979), 117-141.

Poggio, T., "Trigger Features of Fourier Analysis in Early Vision: A New Point of View. In: "The role

of feature detectors"," **ed.P.B.Gough and S.Peters Springer** (1981), in press.

Raviola, E., Gilula N.B., "Intramembrane organization of specialized contacts in the outer plexiform layer of the retina," *J. Cell Biol.* **65** (1975), 192-222.

Stigmar, G., "Blurred visual stimuli. II. The effect of blurred visual stimuli on vernier and stereoacuity," *Acta Ophthalm.* **18** (1971), 364-379.

Wacle, R.A., "Ocular optics and evolution," *J. Opt. Soc. Am.* **66, 10** (1976), 1053-1054.

Westheimer, G., "Eye movement responses to a horizontally moving visual stimulus," *Arcs. Ophthal.* **52** (1954), 932-941.

Westheimer, G., "Diffraction theory and visual hyperacuity.," *Am. J. Optometry & Physiological Optics* **53, No. 7** (1976), 362-364.

Westheimer, G., "The spatial grain of the perifoveal visual field.," *Vision Res.* **22** (1982), 157-162.

Westheimer, G., and McKee, S.P., "Visual acuity in the presence of retinal-image motion," *J. Opt. Soc. Am.* **65, No.** 7 (1975), 847-850.

Wilson, H. and Giese,S.C., "Threshold visibility of frequency gradient patterns," *Vis. Res.* **17** (1977), 1177-1190.

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

Wuelfing, E.A., "Ueber den kleinsten Gesichtswinkel.," *Z. Biol.* **29** (1892), 199-202.