MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. MEMO 524

June 1979

DIRECTIONAL SELECTIVITY AND ITS USE IN EARLY VISUAL PROCESSING

D. Marr and S. Ullman

ABSTRACT: The construction of directionally selective units, and their use in the processing of visual motion, are considered. The zero-crossings of $\nabla^2 G(x,y) * I(x,y)$ are located, as in Marr & Hildreth [1979]. In addition, the time derivative $d/dt\,[\nabla^2 G(x,y) * I(x,y)]$ is measured at the zero-crossings, and serves to constraint the local direction of motion to within 180 degrees. The direction of motion is determined in a second stage by combining the local constraints.

The second part of the paper suggests a specific model of the information processing carried out by the *X* and *Y* cells of the retina and the LGN, and certain classes of cortical simple cells. A number of psychophysical and neurophysiological predictions are derived from the theory.

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 in part by the Advanced Research Project Agency of the Department of Defence under Office of Naval Research contract N00014-75-C-0643 and in part by National Science Foundation Grant MCS77-07569

© MASSACHUSETTS INSTITUTE OF TECHNOLOGY 1979

Introduction

Motion pervades the visual world, and the human visual system uses it in several ways, to control eye movements, to separate figure from ground (Wertheimer 1923; Koffka 1935; Gibson, Gibson, Smith & Flock, 1959; Julesz 1971, chapter 4), and to recover three-dimensional structure from motion (Miles 1931, Wallach & O'Connell 1953, Ullman 1979a). To understand the differing requirements of these visual tasks, it is useful to divide them into two classes, which we shall term tasks of separation and tasks of integration. Separation tasks are those that, in principle, can rely only on instantaneous measurements of position and velocity in the image. An example of such a task is the detection of a sudden movement, which is useful for driving certain kinds of eye movement, or for helping separate figure from ground. Tasks of integration, on the other hand, are those that rely upon the accumulation of information over a period of time. For the recovery of structure and three-dimensional motion from an orthographic projection, for example, instantaneous position and velocity values are insufficient. The task requires the integration of this information over time (Ullman 1979b sections 4.2, 4.5). In the case of discrete presentation, the recovery of three-dimensional structure under orthographic projection requires three different views (Ullman, 1979a), while for tasks of separtation two frames separated by a short time interval are sufficient.

These tasks are sufficiently different that one may expect them to be carried out by separate mechanisms. Those dealing with separation tasks will be making instantaneous measurements, and will operate over short ranges and short times. Mechanisms for tasks of integration cannot be so restricted.

There is some psychophysical evidence for this dichotomy. The reversed phi phenomenon (Anstis 1970) and Braddick's (1974) short range process are both restricted to a range of 10 to 15', and ISI's below 50 msec (Anstis, 1970; Braddick, 1974; Anstis & Rogers, 1975). Apparent motion, on the other hand, can operate over much longer ranges (several degrees of visual angle) and times (400 msec, Neuhaus 1930) and some kinds of apparent motion require long ISI's to be perceived (200 msec. in Ramachandran 1973; 100-200 msec. in Julesz & Payne 1968). These may be the mechanisms involved in the correspondence process and the recovery of stucture from motion (Julesz & Payne, 1968; Ullman 1979b).

This article concentrates on tasks of separation, and it is organized into two parts. In the first, we consider the computational requirements of this kind of task, analyzing the construction of directionally selective units, and their use in the separation of moving objects from one another and from the background. In the second part, we combine this analysis with that of Marr & Hildreth (1979) to propose a specific model of the information processing carried out by the *X* and *Y* cells of the retina, the lateral geniculate nucleus, and certain classes of cortical simple cells. Finally, a number of critical psychophysical and neurophysiological predictions are derived.

I Theoretical analysis

Tasks of separation rely on the instantaneous measurement of the motions of elements in the visual field. These measurements can then be used to detect moving objects, to avoid collisions, to help carve up the visual field into objects, and so forth. There are therefore two main steps to consider, the measurement of the field of velocities over the image, and the subsequent use of these measurements. We deal with each of these in turn.

Establishing the velocity field

Establishing the velocity field means assigning velocities to elements everywhere in the image. The first question is, what are the optimal primitives whose velocity is measured? There are two general requirements to consider here. The first is that in separation tasks speed of computation is of the essence. Secondly, it is important to be sensitive to a wide range of velocities. These two requirements interact, because the fast detection of low velocities demands sensitivity to very small displacements. The human visual system, for example, can detect velocities as low as about 1'/sec (Graham 1965 p. 575; King-Smith, Riggs, Moore & Butler, 1977), and cortical simple cells in the cat can detect displacements as small as 0.87' of arc (Goodwin, Henry & Bishop, 1975).

These two requirements favour the use of early primitives. The earliest possible primitives are the raw intensity values, the next are zero-crossing segments (Marr & Poggio, 1979; Marr, Poggio & Ullman, 1979, Marr & Hidreth, 1979), and above that are edge segments. Zero-crossing here refers to the zero values in the convolution of the image I with a mask shaped like $\nabla^2 G$, where ∇^2 is the Laplacian operator, and G is a two-dimensional gaussian distribution. These zero-crossings can be thought of as the zero values in a second derivative operator applied to the filtered image. They correspond to the locations of sharp intensity changes in the image, as seen through a mask of a certain size. They are the precursors of edges. For more details, see Marr & Hildreth (1979).

There are probably several biological systems that detect relative movement directly from intensity values, for example the motion detection system of the the frog and rabbit retinae (Barlow 1953; Maturana, Lettvin, McCulloch & Pitts 1960; Maturana & Frenk 1963; Barlow & Levick 1965; Torre & Poggio 1978), of the fly (Poggio & Reichardt 1976), and possibly also retinal W-cells in higher mammalian visual systems. Such schemes are useful for saying where in the visual field a relative movement has occurred. If in addition one wishes to analyze the shape of the moving patch, it seems more sensible to try to combine the analysis of movement with the analysis of contours. The earliest stage at which this could be carried out is at the level of zero-crossing segments, and as we shall later see, the physiological data support this view.

Nature of the measurement

The use of zero-crossing segments as primitives for motion raises a substantial difficulty which we shall call the *aperture problem* (see figure 1). If the motion is to be detected by a unit that is small compared with the overall contour, the only information one can extract is the component of the motion perpendicular to the local orientation. Motion along the contour will be invisible. Hence local measurements alone fail to give either the direction or speed of movement, and can only restrict the direction to within 180^0. In other words, only the sign of the movement is given directly by the local measurement.

Therefore, using zero-crossings (or any oriented local element) as primitives divides the problem into two stages. In the first, the local sign is established, and in the second, the local signs are compared and combined. We deal now with the first stage, the construction of units that detect the sign of the movement of an oriented zero-crossing segment. We call such units ***directionally selective***.

The construction of directionally selective units

The construction of directionally selective units involves two steps; firstly, the detection of an oriented zero-crossing segment, and secondly, establishing the sign of its motion. Zero-crossing segments may be detected by the mechanism shown in figure 2 (Marr & Hildreth 1979). The basic idea is that, if the values of the convolution $\nabla^2 G * I$, which we shall write as $S(x,y,t)$ are carried by two kinds of

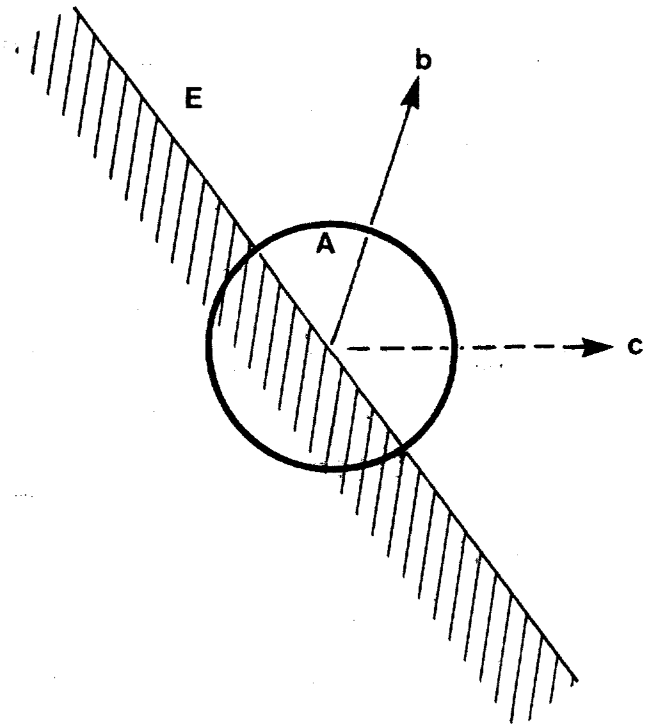
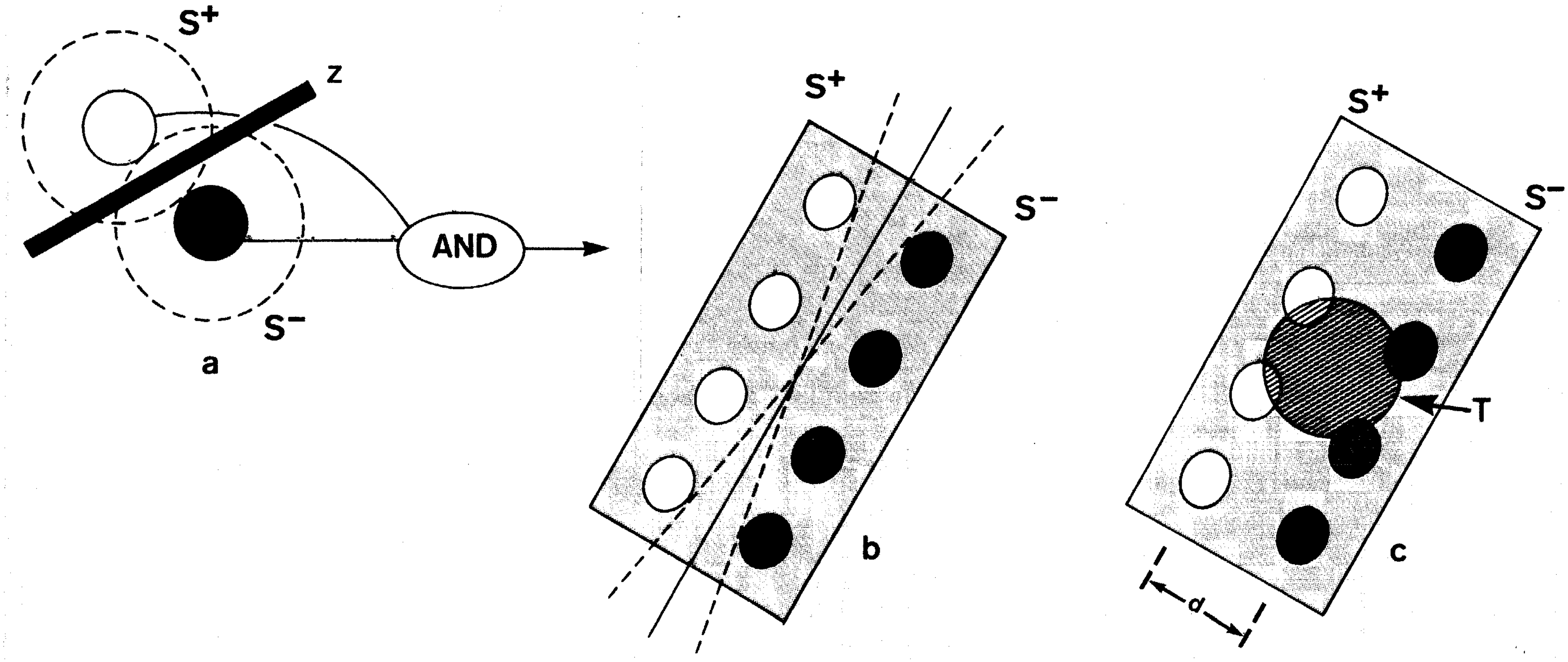


Figure 1. The Aperture problem. If the motion of an oriented element is detected by a unit that is small compared to the size of the moving element, the only information that can be extracted is the component of the motion perpendicular to the local orientation of the element. Looking at the moving edge E through a small aperture A, it is impossible to determine whether the actual motion is, e.g., in the direction of b or that of c.

Figure 2. The detection of zero-crossings. S^- and S^+ units are combined through a logical *AND* operation (figure 2a). Such a unit would signal the presence of a zero-crossing runnig between the two sub-units. A row of similar units connected through a logical *AND* would detects the an oriented zero-crossing within the orientation bounds given roughly by the dotten lines in (b). In (c) a T unit is added to the detector in (b). If the unit is T^+, it would respond when the zero-crossing segment is moving in the direction from the S^+ to the S^-. If the unit is T^-, it would respond to motion in the oposite direction.

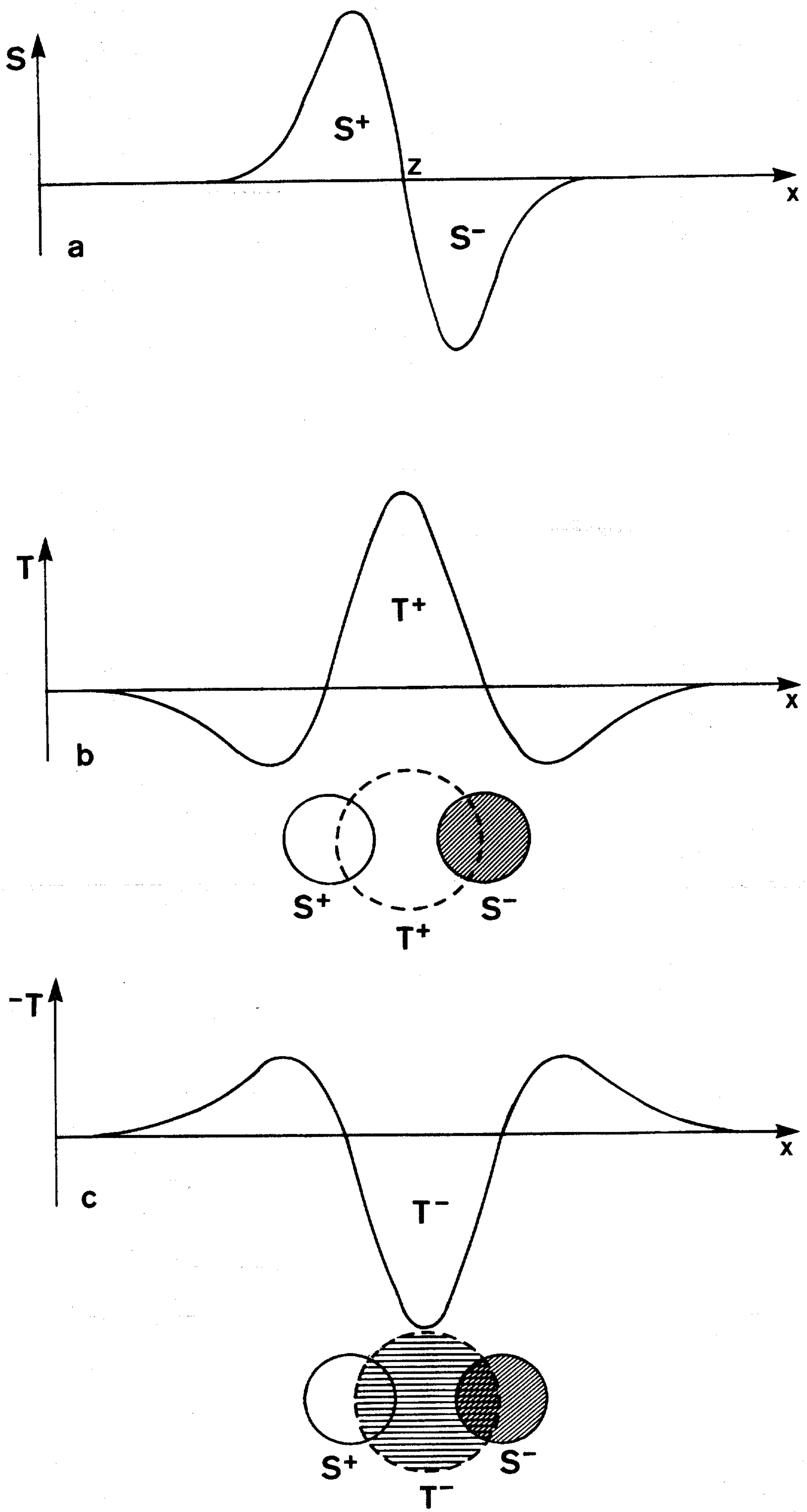

S+
z
S−
AND
a
S+
S−
b
S+
S−
T
d
c

unit, one dealing with positive values ("on-centre") and the other with negative values ("off-centre"), on-centre units will be active on one side of the zero-crossing, and off-centre units, the other side. Hence if the two sides are combined through a logical ***AND*** gate, the gate will detect the presence of a zero-crossing running between them (see figure 2a). A row of such units will detect an oriented segment of zero-crossings (figure 2b). Figure 3a illustrates the profile of the convolution values (of $\nabla^2 G * I$) in the vicinity of an isolated step change in intensity. S^+ in figure 3a indicates the position of the on-centre units, and S^-, of the off-centre units. When the zero-crossing Z lies between the two units, both are active, and the ***AND*** gate (figure 2a) performs the detection. If the two units are separated by about w, the width of the central excitatory region of the receptive field, each will be maximally stimulated by an edge midway between them. This separation thus yields the most sensitive conditions for zero-crossing detection.

It is clear from figure 3a that, if the zero-crossing is moving to the right, the value of the convolution at position Z will be increasing; and if the zero-crossing is moving to the left, the value will be decreasing. Hence by examining the sign of the time derivative of the convolution, *i.e.*, the sign of $\partial/\partial t\ (\nabla^2 G * I)$, at position Z, the direction of motion can be determined unambiguously. Figures 3b and c illustrate this. Let us write:

$$T(x,y,t) = \partial/\partial t\ (\nabla^2 G * I) = \partial/\partial t\ (S(x,y,t)).$$

Figure 3. The value of $S = \nabla^2 G * I$, and of $T = \partial/\partial t\ (\nabla^2 G * I)$ in the vicinity of an isolated intensity edge. Figure 3a shows the S signal as a function of distance. The zero-crossing in the signal corresponds to the position of the edge. Figure 3b shows the spatial distribution of the T signal when the ede is moving to the right, and (c) when it is moving to the left. Motion of the zero-crossing to the right can be detected by the simultaneous activity of S^+, T^+, S^-, in the arragement shown in (b). Motion of the zero-crossing to the left can be detected by the S^+, T^-, S^-xgp unit in (c).


S
S+
z
x
a
S−
T
T+
x
b
S+
T+
S−
−T
x
c
T−
S+
S−
T−

Then if the motion is to the right, at the instant the zero-crossing reaches Z the values of T(x,y,t) have the spatial distribution shown in figure 3b. T is strongly positive at Z, and it remains positive over a neighborhood of Z that is 2σ wide, where σ is the space-constant of the gaussian G. If the motion is to the left, the sign of T is reversed, and the situation is that shown in figure 3c.

The spatial distributions of S and T near a zero-crossing suggest a straightforward design for a robust directionally selective unit. The only measurement that we need, in addition to those for detecting a stationary zero-crossing (figure 3a), is T(x,y,t); and like the S values, we need to split T into two channels, one carrying the positive part (which we denote by T^+), and one carrying the negative part (T^-). The directionally selective unit can then be constructed from three subunits. If all of S^+, T^+, S^- are active simultaneously, and have the spatial configuration shown in figure 3b, an intensity change with higher intensities to the left (the S^+ side) is moving to the right (from S^+ to S^-). If S^+, T^- and S^- are active simultaneously (figure 3c), the same intensity change (higher intensities on the S^+ side) is moving to the left (from S^- to S^+).

Hence the oriented zero-crossing detector of figure 2b can be made directionally selective by adding an appropriate T^+ or T^- input, for example at the centre of its receptive field (as shown in figure 2c). We shall refer to a unit made directionally selective in this way as an *STS* unit. Notice that this scheme is economical in T units; the number of T-units required would be considerably less than the number of S-units.

Comments on the size and number of T channels required

There are a number of parameters that need to be chosen correctly for such a unit to function reliably. These are (i) the spatial dimensions of the S and T units; (ii) their relative positions and (iii) the temporal filter computing the time derivative in the T channel. The important questions for the performance of the device is, what is the range of angular velocities over which it performs reliably, and how does this range depend upon the spatial frequency of the stimulus?

We consider first the simplified case in which the T channel delivers the exact and undelayed temporal derivative. The sizes of the S and T units are characterized by the space constants σ_S, σ_T of their respective Gaussians. The widths w_S, w_T of the central excitatory region of these channels are given by $w_S = 2\sigma_S$, and $w_T = 2\sigma_T$ Let d denote the separation of the S^+ and S^- units (as in figure 2c).

The optimal separation of the S^+ and S^- units is w_S, since this is the distance between the positive and negative peaks in the response to a step change in intensity. The condition for proper functioning of the unit is that the T response should remain positive whenever the zero-crossing Z lies between the centres of S^+ and S^-, and Z is moving from S^+ towards S^-. For an isolated edge, if the T^+ unit is placed exactly midway between S^+ and S^-, the unit would function properly if $w_T \geq d$, and if $w_T \geq 2d$, the centre of the T^+ unit can lie anywhere between the centres of the two S units.

An ideal unit such as this will in principle be directionally selective to an infinite range of angular velocities. In practise, its response at the lower end will be determined by its sensitivity, and at the higher end will depend on the nature of the temporal filter in the T channel. Additional constraints on the size and number of T units may be introduced if the delayed derivative, rather than the derivative itself is computed. If an isolated edge moves at velocity v across a T unit that signals the time derivative delayed by τ msec, then the directionally *STS* selective unit would function properly (assuming a single T unit midway between two S units separated by a distance d) if: $v\tau + d/2 \leq \sigma_T$. Assuming again that $d/2 = \sigma_S$, we conclude that the transient channel has to be considerably larger than the stationary one. The exact size relationship would depend on the maximum velocity to which the unit is required to respond, the exact shape of the temporal filter, and the position of the T sub-units. The optimal cover of a wide range of velocities may require therefore more than a single transient channel.

Comparison with other schemes

The STS unit has several characteristics that make it well-suited to the problem of detecting directional selectivity. They are: (i) It requires only local measurements; (ii) No time delay is involved, beyond that required to compute the temporal derivative; (iii) The lower limit to the displacement that can be detected is the unit's sensitivity, and the upper limit, which depends on the temporal filter, will be high if the time constants are small. Hence a single unit can be made sensitive to a wide range of speeds. (iv) Within this range, and for a sufficiently isolated edge, the unit will be completely reliable.

Another approach to the design of a directionally selective zero-crossing unit might be to adapt the schemes proposed by Hassenstein & Reichardt (1956), Barlow & Levick (1965) and Torre & Poggio (1978). A careful analysis of this type of scheme has been given by Poggio (in preparation), in connexion with the system used by the housefly. The basic idea is essentially to detect motion by identifying the same "thing" at two different locations at two different times. The fly uses directly its detectors of intensity; for our purposes, one would use two zero-crossing detectors. The motion detecting circuitry connects one detector directly, and the other indirectly through a delay or a (temporal) low-pass filter, to an ***AND-NOT*** gate. Provided that the speed of the movement and the spatial frequency characteristics of the input are adequately restricted, the system can detect relative movement. The range which we have in mind, from about 1' per second to over 3 degrees a second, is probably too large to be accomodated by a single such system, but it could be handled by two, a small one and a larger one, operating in parallel (T. Poggio, personal communication).

The critical difference between such schemes and the one we propose is that our system does not have to wait until the stimulus has passed from the first detector to the second. It can therefore respond instantaneously, and it will be sensitive to very small displacements. In addition, unlike systems based on a pair of detectors, it does not have to effectively "guess" that whatever is exciting one detector now is the same thing that excited the other a short time ago. Guessing correctly all the time amounts to solving the correspondence problem, which is difficult (Ullman 1979b), and is furthermore unnecessary for tasks of separation.

In addition, all the two-detector systems known so far are based on the use of a delay and an *AND-NOT* gate (Barlow & Levick 1965; Torre & Poggio 1978). Such systems suffer from a stop-restart failure -- that is, if a stimulus moving in the null direction is halted between the two detectors for longer than the delay used by the system, when the stimulus restarts its movement, the system will give a response. A similar failure afflicts stimuli moving very slowly in the wrong direction. Goodwin, Henry, & Bishop (1975) looked for this phenomenon in directionally selective cortical simple cells, and failed to find it.

Finally, our model is clearly motivated by the physiological evidence about sustained (*X*) and transient (*Y*) cells. Given these building blocks, it is therefore natural to ask whether there are other, perhaps better ways of combining the S and T channels to yield directionally selective zero-crossing detectors. We have considered all possible logical combinations of up to three units; that is all possible combinations using the logical operations ***AND***, ***OR*** and ***NOT***, of the S and T units. One reason for considering logical combinations, as Barlow & Levick (1965) did, is that we would like our units to be robust, *i.e.* rather insensitive to the actual magnitudes of its input signals.

Of all of these possibilities, only the STS combinations and their logical equivalents yield reliable units. For example,

$(S^+$ ***AND*** T^+ ***AND*** $S^-)$ is logically equivalent to

$(S^+$ ***AND*** (***NOT*** $T^-)$ ***AND*** $S^-)$, and they are equally reliable. In a strict implementation, the second of these would respond to a stationary edge as well as to one moving in its preferred direction, whereas the first would respond only to a moving edge. Units made from logical combinations of only S cells are not directionally selective; units made only from T cells can be fooled by reversing both the contrast and the direction of movement; and a combination like $(S^+$ ***AND*** $T^-)$, while exhibiting a clear preference for motion in one direction, can give a non-zero response in the other.

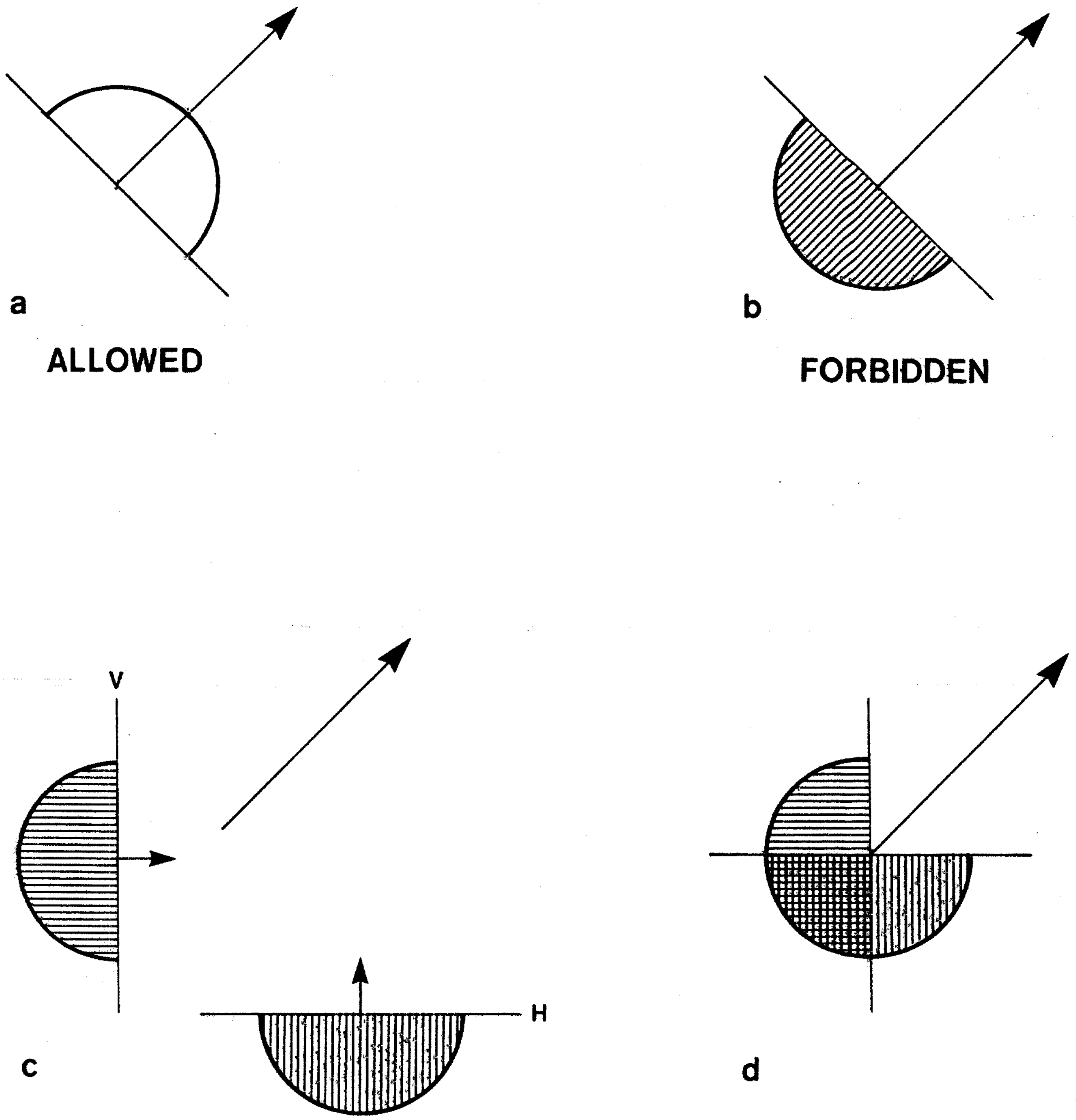
The use of directional selectivity

The movement of an object against its background can be used to delineate its boundaries, and the human visual system is efficient at exploiting this fact (Julesz 1971 chapter 4; Braddick 1974). If the complete velocity field is given (*i.e.* speed and direction at each point), object boundaries will be indicated by discontinuities in this field. This is because the motion of rigid objects is locally continuous in space and time. The continuity is preserved by the imaging process, and gives rise to what we might call the ***principle of continuous flow***, according to which ***the velocity field of motion within the image of a rigid object varies continuously almost everywhere.*** Since the motions of unconnected objects are generally unrelated, the velocity field will often be discontinuous at object boundaries. Conversely, lines of discontinuity are reliable evidence of an object boundary.

Unfortunately, the complete velocity field is not directly available from measurements made on small oriented elements. Beause of the aperture problem, only the sign of the direction of movement is available locally. This means that an additional stage is necessary for the detection of discontinuities in the velocity field. In this section, we ask how and to what extent the more limited raw information (the sign of the direction only) may be used to detect these discontinuities.

The sign of the local direction of motion determines neither the movement's speed nor its true direction, but it does place constraints on what the true direction can be (see figure 4). The constraint is that the true direction of motion must lie within the 180° range on the allowed side of the local oriented element (figure 4a), or, alternatively, it is forbidden to lie on the other side (figure 4b). The constraint thus depends on the orientation of the local element. Hence if the visible surface is textured and gives rise locally to many orientations, the true direction of movement may be rather tightly constrained.

Figure 4. The combination of local constraints from *STS* units to determine the direction of motion. The constraint placed by a single *STS* unit is that the direction of motion must lie within a range of 180° on the allowed side of the oriented element (figure 4a), or, equivalently, it is forbidden to lie on the other side (b). Figure 4c shows the forbidden zones for two orineted elements moving along the direction indicated by the arrow. The foridded zone of their common motion is the union of their individual forbidden zones, as indicated. The direction of motion is now constrained to lie within the intersection of their allowed zones, i.e. the first quadrant.


a
ALLOWED
b
FORBIDDEN
V
H
c
d

The way in which constraints can be combined is illustrated in figures 4c & 4d, for the simple case of two local elements. The true direction of motion is diagonal here. The vertically oriented directionally selective unit *V* sees motion to the right; and the horizontally oriented unit *H* sees motion upwards. If these two units share a common motion, we can combine the constraints they place on the direction of that motion by taking the union of their forbidden zones (figure 4d). The result is that the direction of motion is now constrained to lie in the first quadrant, as illustrated. The addition of further units can further constrain the true direction of motion by expanding the forbidden zone of figure 4d.

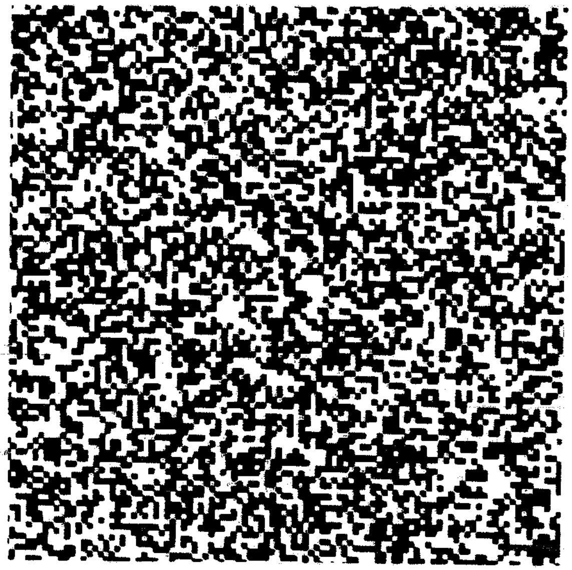
It can also be seen from the diagram how the motion of two groups of elements may be incompatible. If the allowed zone for one group of elements is completely covered by the forbidden zone of another, their motions clearly cannot be compatible. Notice in this connexion that only the direction of movement, not its speed, is used here.

Once the direction of motion has been established, for example by the method of figure 4, the true velocity field can be approximately recovered. If the measured velocity perpendicular to an oriented zero-crossing segment is v, and the found direction at θ° to the segment, then the magnitude of the true velocity is $v \arcsin(\theta)$. Such a scheme would require, however, a measurement of the speed perpendicular to the zero-crossing segment, which the basic *STS* unit does not accomplish. A system that segments a scene using *STS* like units will thus be relatively insensitive to variations in speed.

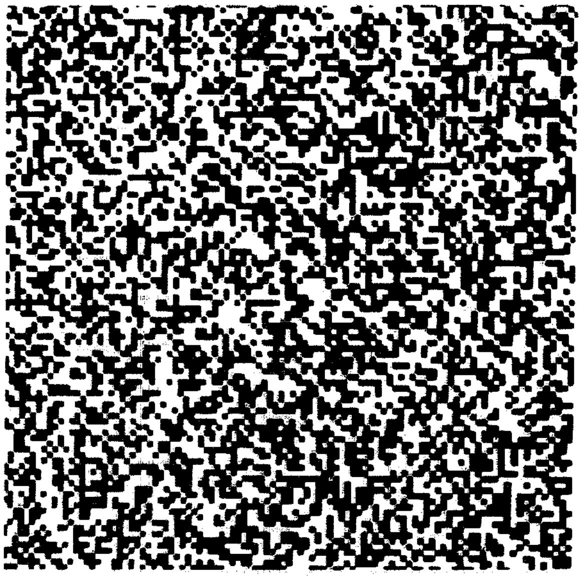
The final observation that we need in order to use this scheme for delineating moving objects is that objects are localized in space. If the objects are opaque, their images will have an interior within which the forbidden zones in diagrams like figure 5d will be consistent, provided that they draw their elements from small neighborhoods. The only exceptions to the principle of continuous flow occur at singularities in the velocity field, like the centre of a rotating disc. Such singularities can however occur only at isolated points, and there can be at most one for each rigid object; no false lines of discontinuity can be formed.

Figure 5 shows an example of detecting a moving pattern embedded in a pair of random dot images using the above scheme. A central square in figure 5a is displaced in figure 5b to the right, while the backgrounds of the two images are uncorrelated. Figure 5c depicts the zero-crossing contours of figure 5a filtered through $\nabla^2 G$. Figure 5d represents the result of applying the *STS* operation assuming that figures 5a and 5b are shown in a rapid succession. The time derivative $\partial/\partial t\ (\nabla^2 G * I)$ was computed for each position along the zero-crossing contours in figure 5c. The small light dots attached to the zero-crossing contours in 5d indicate the local direction of motion (the zero-crossing is moving towards the light dot). The central square was found to have a consistent common direction (to the right). The light dots were removed in these area, accept where errors in assigning directions occured. Since the backgrounds are uncorrelated, no consistent direction was found for this region.

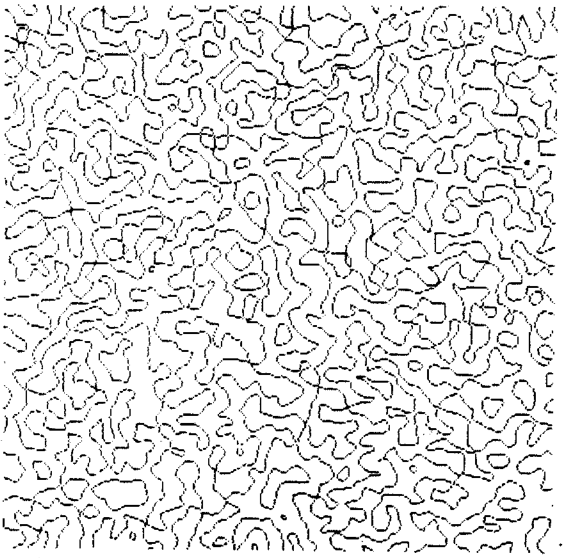
Figure 5. Separating a moving figure from its background using combinations of *STS* units. A central square in figure 5a is displaced in figure 5b a to the right. The backgroung in the two pictures is uncorrelated. Figure 5c shows the zero-crossing contours of (a) filtered through $\nabla^2 G$. The light dots in figure 5d depicts the local directions assigned to the zero-crossings by the *STS* units. The motion is in the direction of the light dots. The central area was found to have a common consistent direction, to the right. The light dots were removed from this area, except for isolated points were the direction assigned was incorrect. No consistent direction was found for the background (5e).



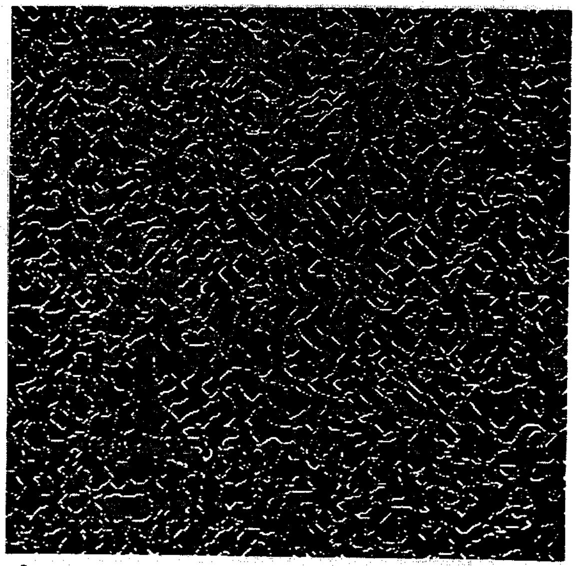
a



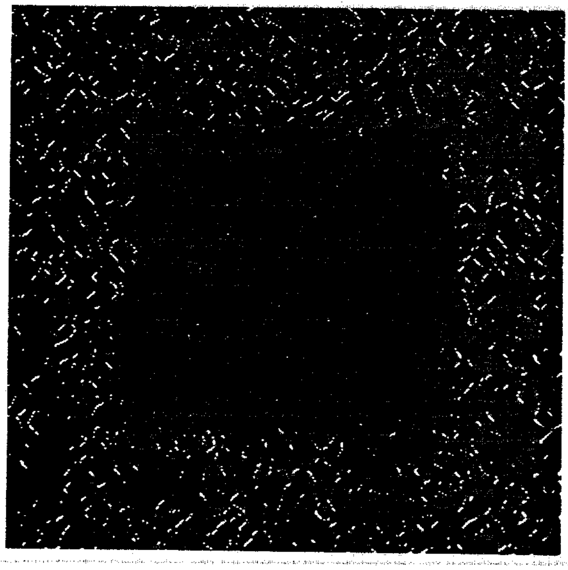
b



c



d



e

Looming

By combining directionally selective units from the two eyes, a different kind of information can be acquired. Suppose that a particular zero-crossing has been identified and assigned incompatible motions in the two images. Then the zero-crossing is moving in depth either towards (if both retinal motions have temporal components) or away from (if both have nasal components) the viewer. If motion is to the right on both retinae, the object will pass safely to the viewer's left, and *vice versa*.

For this type of analysis, one does not need to combine constraints in the manner of figure 5; one can use the raw output of the directionally selective units. The difficulty in this case lies in ensuring that both left and right detectors are looking at the same zero-crossing, and establishing this match is the essence of the stereo matching problem (Marr & Poggio 1979). If, however, one is prepared to tolerate inaccuracies from time to time, a fast looming detector can be designed that does not have to wait upon the results of stereo matching. For example, a simple looming detector can be constructed by comparing the signs of motion at corresponding retinal points. Such points will often but not always correspond to nearby points on the same moving object.

Such a scheme might rely at some point on a cell with binocular receptive fields that are incongruous (in the sense of von der Heydt, Adorjani, Hanny & Baumgartner 1978) rather than truly disparity sensitive, and whose preferred motions in the two eyes are opposite. There is some evidence for the existence of such cells (Regan, D. Beverly, K.I. & Cynader M. 1978 PRS).

Biological Implications

There are three main components to our scheme for constructing directionally selective units: (i) The computation of the convolution $\nabla^2 G*I$, (ii) the measurement of its time derivative $\partial/\partial t\ (\nabla^2 G*I)$, and (iii) their combination in the manner described by figure 3. We shall suggest that the first component corresponds to *X*-type cells in the retina and the LGN; the second to *Y*-type cells; and the third to a subclass of cortical simple cells. We consider each of the three components in turn, and for each one we shall review the available physiological and psychophysical evidence.

The Computation of $\nabla^2 G*I$

The spatial and temporal properties of retinal x-cells are appropriate for the computation of $\nabla^2 G*I$. We deal with each in turn.

Spatial properties -- Neurophysiology

The overall center-surround organization of retinal ganglion cells was first discovered by Kuffler (1952, 1953). Rodieck and Stone (1965) suggested that this organization was the result of superimposing a small central excitatory region on a larger inhibitory "dome" that extends over the entire receptive field. Rodieck (1965) and Euroth-Cugell & Robson (1966) described the two "domes" as gaussians, thus describing the receptive field as a difference of two gaussians (DOG). With appropriately chosen space constants, a DOG provides a close approximation to $\nabla^2 G$ (Marr & Hildreth 1979 appendix B). Figure 6 illustrates this point. The continuous curve in the figure is $\nabla^2 G$,

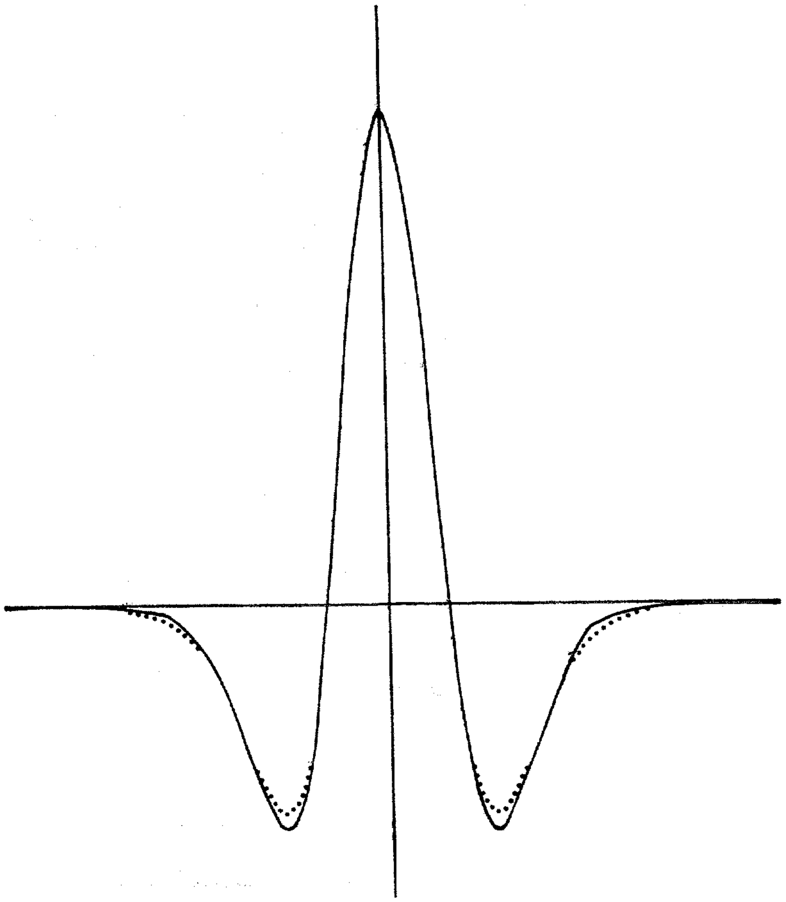


Figure 6. Comparing $\nabla^2 G$ to a difference of gaussians (DOG). The dotted line is a DOG with $\sigma_i/\sigma_e = 1.6$. The solid line is an approximation of this DOG using $\nabla^2 G$. For more detail see [Marr & Hildreth, 1979, appendix B).

and the dotted curve is its approximation by a DOG with space-constants in the ratio 1:1.6. The DOG approximation to $\nabla^2 G$ provides a physical implementation which is easily assembled by subtracting two gaussian "pools" of receptors.

At the LGN, the important properties and distinctions are preserved. The receptive fields preserve their shape (Hubel & Wiesel 1961). The *X*-*Y* and the on-off distinctions are preserved by the retino-geniculate mapping (Cleland, Dubin & Levick, 1971; Hoffman, Stone & Sherman, 1972; Cleland, Levick & Sandersen, 1973; Dreher & Sanderson, 1973). Furthermore, Singer & Creutzfeldt (1970) and Cleland, Dubin & Levick (1971a, 1971b) found that geniculate cells were for the most part driven by only one, or a very few, retinal ganglion cells.

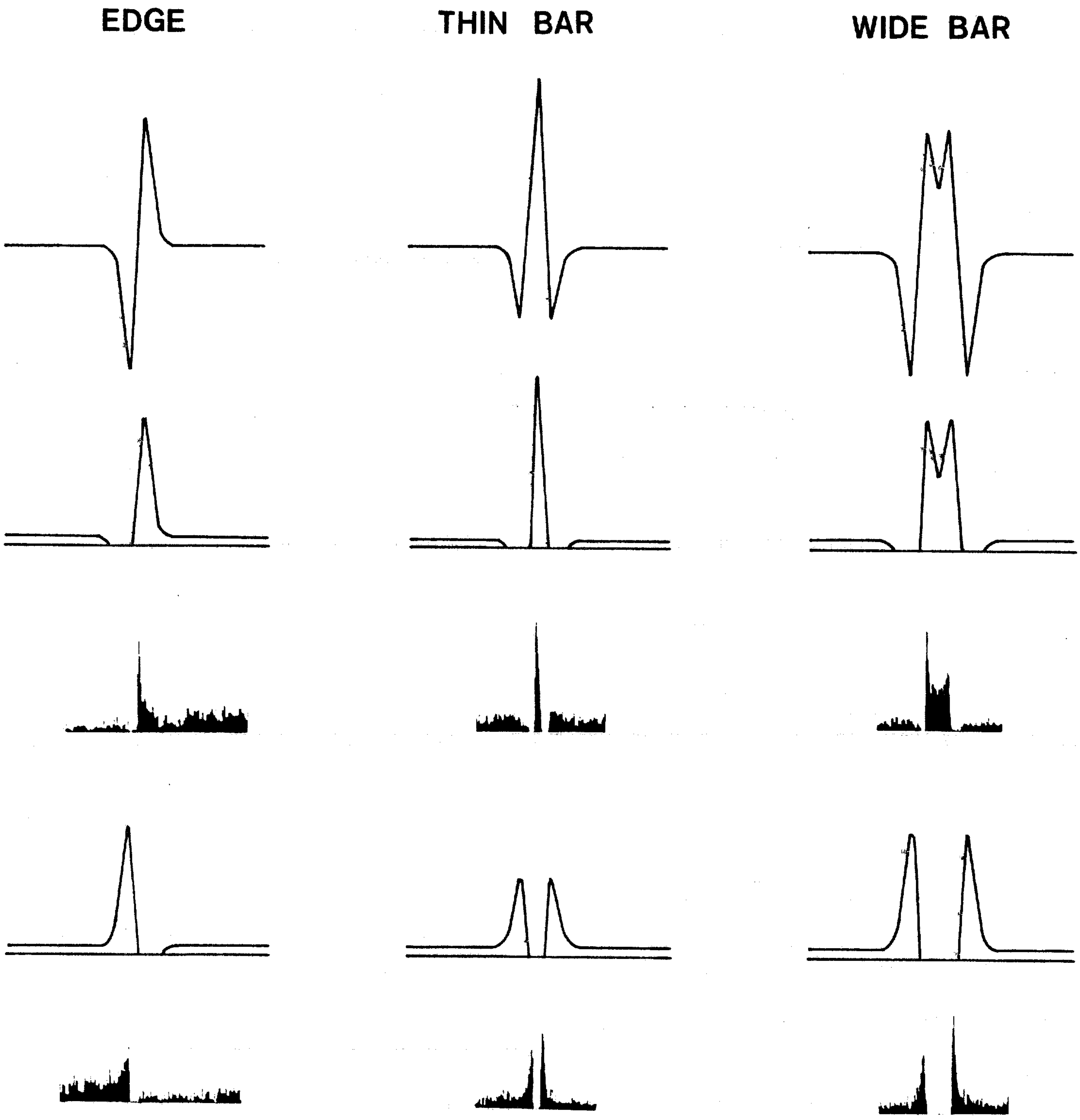
At the level of the retinal ganglion cells there is little or no scatter in receptive field size (J.G. Robson, personal communication). One possible way in which the two sizes of *X* and *Y* channels required by computational requirements (Marr & Hildreth, 1979) and by psychophysical findings (Wilson & Bergen 1979) could arise, is from the limited convergence at the LGN. Computational experiments have established that large DOGs can be constructed from the outputs of a few smaller ones. For example, five DOGs can be combined to form approximately a DOG with twice the space constant.

Temporal Properties -- Neurophysiology

Ideally, the measurement of $\nabla^2 G$ is instantaneous, *i.e.*, for an image that does not vary in time the signal should not vary in time. The ideal temporal response should therefore have no transient components. Retinal *X*-cells do exhibit a transient response but they are characterized by a strong sustained component (Cleland, Dubin & Levick, 1971; Cleland, Levick & Sanderson, 1973).

The overall response of retinal and LGN *X*-cells agrees closely with the predictions based on the $\nabla^2 G$ operation. Figure 7 compares the predicted responses of retinal or geniculate *X*-cells to their observed responses to various stimuli: a moving edge, a moving thin bar, and a moving wide bar. The predicted traces are calculated by taking either the positive or the negative part of $\nabla^2 G * I$ superimposed on a small resting or background discharge. The physiological responses are taken from Dreher & Sanderson (1973 figure 6 d & e) for the responses to an edge; and from Rodieck & Stone (1965) figures 1 and 2, using traces from bars 1 and 5 degrees wide. The predictions were calculated for bars of width w and $2.5w$, where w is the width of the central excitatory region of the receptive field. For the *X*-cell traces, records of on-centre cells were used for stimuli of opposite contrast, rather than records of off-centre cells to stimuli of the same contrast. The reason for this is that the predictions are the same for both stimuli, and there are few good published traces of the right kind for off-centre cells. Finally, it should be noted that Rodieck & Stone's paper preceeded Enroth-Cugell & Robson's (1966)

Figure 7. Comparison of the predicted responses of on- and off-centre *X*-cells to electro-physiological recordings. The first row shows the response of $S = \nabla^2 G * I$ for an isolated edge, a thin bar (bar width = w, were w is the width of the central excitatory region of the receptive field), and a wide bar (bar width = $2.5w$). The predicted traces are calculated by superimposing the positive (in the second row) or the negative (in the fourth row) parts of $\nabla^2 G * I$ on a small resting or background discharge. The positive and negative parts correspond to either the same stimulus moving in opposite directions, or stimuli of opposite contrast moving in the same direction. The physiological responses are taken from Dreher & Sanderson (1973 figure 6 d & e) for the responses to an edge; and from Rodieck & Stone (1965 figures 1 and 2), using traces from bars 1 and 5 degrees wide.



X-CHANNEL

distinction between *X*- and *Y*-cells, and that most of Dreher & Sanderson's (1973) cells, including all those whose traces we have reproduced, were not classified as *X* or *Y*. Nevertheless their behaviours are quite different (compare figures 7 and 8), and can therefore be confident of our *post hoc* classification.

Sustained Channels -- Psychophysics

The existence of channels with a sustained response, and their distinction from transient channels, has been known for a long time, and more recently their possible correspondence with the physiological *X*- and *Y*-channels has been pointed out (Tolhurst 1973; Kulikowski & Tolhurst 1973). The receptive fields of the sustained mechanisms were measured psychophysically by Wilson (1978) and by Wilson & Bergen (1979). They suggested the existence of two sizes. Both can be fitted by DOGs with $\sigma_i:\sigma_e = 1:1.75$, and with $w = 3.1'$ and $6.2'$ at the fovea. (For $\nabla^2 G$, $w = 2\sigma$, i.e., $\sigma_1 = 1.55'$, $\sigma_2 = 3.1'$). Since these measurements used elongated stimuli, they correspond to the projection of the receptive fields onto one dimension. If the receptive field were constructed from circularly symmetric DOG-shaped subfields, the measured values of w should be multiplied by $\sqrt{2}$ to obtain the values for the subfields.

Interestingly, Kulikowsky & Tolhurst (1973) found that the sustained channels are "too sustained". Unlike the physiologically measured *X*-cells, the psychophysically determined sustained channels do not exhibit a noticeable transient component.

The Computation of $\partial/\partial t(\nabla^2 G * I)$

We shall demonstrate that under "reasonable" conditions, *i.e.*, for edges and bars moving at velocities up to a few deg/sec, *Y*-type retinal cells signal approximately $\partial/\partial t(\nabla^2 G * I)$. There is both physiological (Tolhurst & Movshon 1975) and psychophysical (Wilson 1979) evidence that the spatiotemporal response of the transient channel can be described as the product of a spatial receptive field sensitivity function and a temporal impulse response function. As we did for the *X* channel, we shall examine first the spatial then the temporal response.

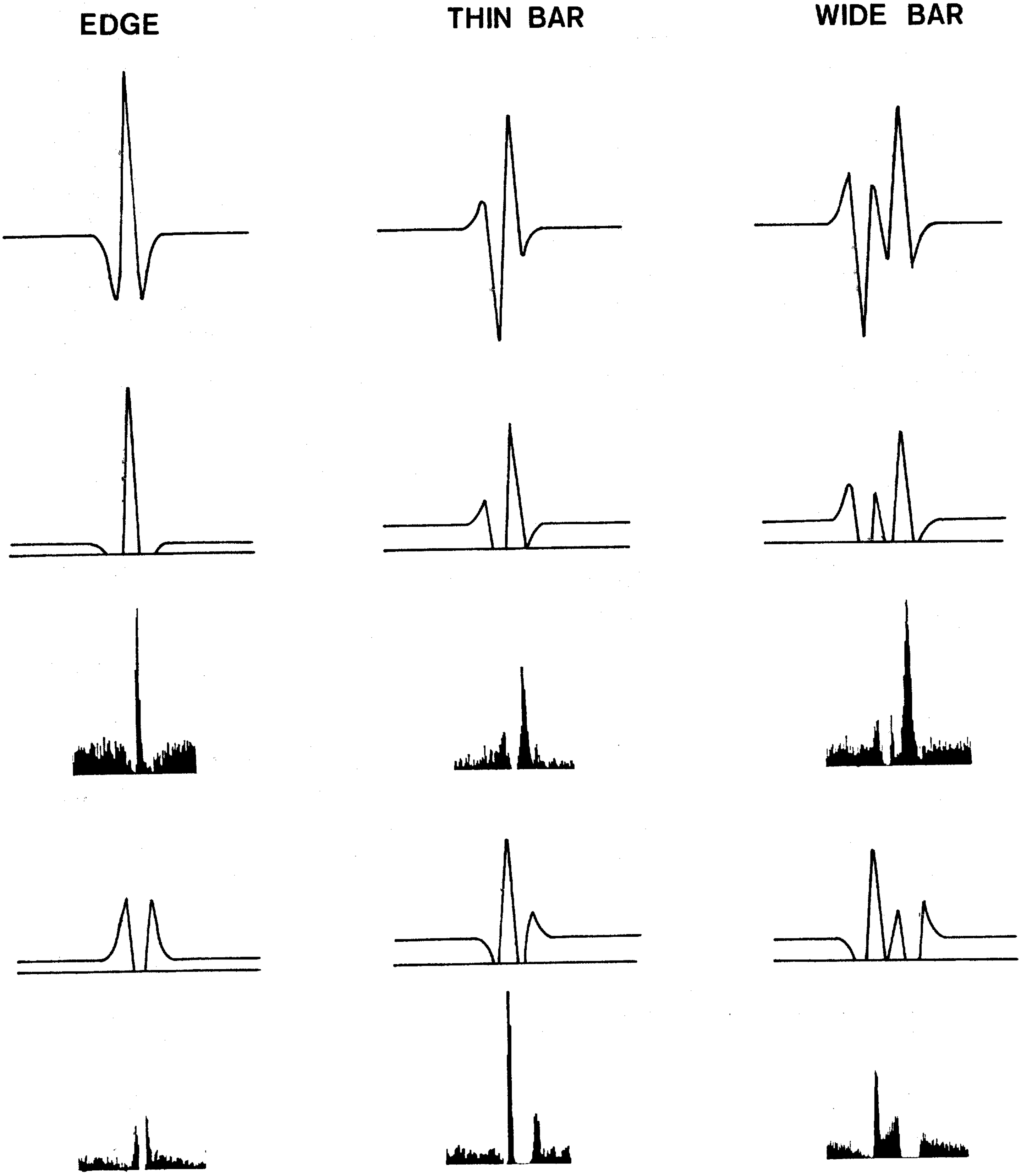
Spatial properties -- Neurophysiology and Psychophysics

Both at the retinal and the LGN levels, the *Y*-cells receptive field is spatially similar to that of the *X*-cells (Rodieck & Stone 1965a; 1965b; Rodieck 1965), only larger (Cleland, Levick & Sanderson, 1973). It has long been known psychophysically that the transient mechanisms are tuned to lower spatial frequencies, therefore having larger receptive fields than the sustained mechanisms. Recently, Wilson (1978) and Wilson & Bergen (1979) plotted the shape of the receptive fields of the transient mechanisms at threshold, and concluded that there are two distinct transient channels. The receptive fields are again DOG-shaped, and the widths of the central excitatory regions are 11.7' and 21' at the fovea (compared with 3.1' and 6.2' for the sustained channels). The ratio of the space constants is approximately 3:1, and unlike the sustained channels they seem to have a DC response at threshold (*c.f.* Cowan 1977). There is some physiological evidence that the D.C. response, as well as the size of the inhibitory region, may depend on the adaptation level (Euroth-Cugell & Shapley, 1973a & b).

Temporal properties -- Neurophysiology

Our requirement for the temporal component of the Y-cell response is that it takes the time derivative of the output of the spatial filter. This is consistent with Rodieck & Stone's (1965b) description of units whose response was "directly correlated with the gradient of the receptive field as defined by flashing lights" (p. 842). Of course, no physical device can take a perfect time derivative over the entire temporal frequency range. However, the published response curves of retinal and geniculate Y-cells to bars and edges moving at moderate velocities are in a close agreement with the predictions based on the time-derivative operation $\partial/\partial t(\nabla^2 G * I)$. Figure 8 compares the predicted responses of on- and off-center cells, that we suppose to have been Y-cells, to their observed responses to various stimuli. All the stimuli were light (*i.e.* light edges, light bars), the thin bars were about half a degree wide (0.4 and 0.6), and the thick bars, 5 degrees (5.0 and 5.1). The traces are taken from Dreher & Sanderson (1973 figures 6b, 8a for the edge responses; figures 1d and 2c for the thin bars; figure 2b for the off-centre thick bar), and from Rodieck & Stone (1965 figure 5b for the on-centre response to a thick bar). The predicted traces show pure values of $\partial/\partial t(\nabla^2 G * I)$ and as in figure 7,

Figure 8. Comparison of the predicted responses of on- and off-centre Y-cells to electro-physiological recordings. The first row shows the response of $T = \partial/\partial t\ (\nabla^2 G * I)$ for an isolated edge, a thin bar (bar width = *w*, were *w* is the width of the central excitatory region of the receptive field), and a wide bar (bar width = 2.5*w*). The predicted traces are calculated by superimposing the positive (in the second row) or the negative (in the fourth row) parts of $\partial/\partial t\ (\nabla^2 G * I)$ on a small resting or background discharge. The positive and negative parts correspond to either the same stimulus moving in opposite directions, or stimuli of opposite contrast moving in the same direction. The physiological responses are taken from Dreher & Sanderson (1973 figures 6b, 8a for the edge responses; figures 1d and 2c for the thin bars; figure 2b for the off-centre thick bar), and from Rodieck & Stone (1965 figure 5b for the on-centre response to a thick bar). The thin bars in these recordings were about half a degree wide (0.4 and 0.6), and the thick bars about 5 degrees (5.0 and 5.1). It can be seen that the observed responses are in close agreement with the predicted ones, even in cases where both are elaborate, (*e.g.* the wide-bar cases).



Y-CHANNEL

the thicknesses of the thin and thick bars were respectively *w* and 2.5*w*. It can be seen that the observed responses are in close agreement with the predicted ones, even in cases where both are elaborate, (*e.g.* the wide-bar cases).

Temporal Properties--Psychophysics

Ideally, to obtain a time derivative, one subtracts from the current value of the signal its value an infinitesimal time ago. if these measurements are taken in practice, they must be taken over finite intervals of time. Hence the impulse response of the derivative-computing channel in the time domain should be composed of a positive phase followed by a phase of a similar shape but opposite sign. In the frequency domain the power spectrum should be roughly linear in frequency over the range in which the device is to operate. These expectations are supported by the psychophysical evidence.

A temporal filter composed of a positive phase of about 60 msec followed by a negative phase was explicitly suggested by Watson & Nachmias (1977), and further supported by Tolhurst (1975), Breitmeyer & Ganz (1977), Legge (1978). The negative phase may be somewhat longer than the positive one, or may be followed by damped oscillation of small amplitude (see Breitmeyer & Ganz 1977, figure 3) without significantly affecting the results.

In the frequency domain, the temporal MTF was measured by Wilson (1979) for the transient U-channel. This MTF does not characterize the temporal filter completely, since the phase information is still missing. If the overall shape of the temporal filter is indeed composed of a positive phase 60 msec. long followed by a similar negative phase, one cane approximate the phase relationaships by assuming that the filter is an antisymmetric function about $t = 60$ msec. We have computed the results of applying this hypothetical filter to lines and bars moving at 3 deg/sec. The results are shown in figure 9, and they are in a good agreement with the operation $\partial/\partial t\ (\nabla^2 G * I)$.

Deviations of the Temporal Response From a True Time Derivative

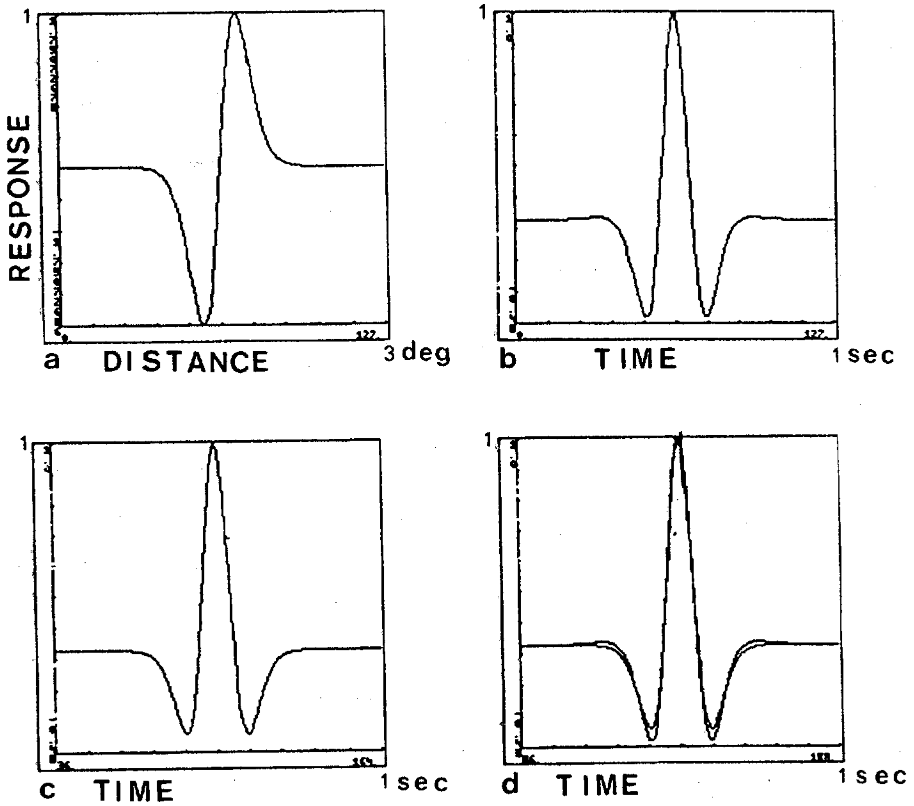
The transient channels do not take a true time-derivative. We divide the sources of aberrations into linear and non-linear types.

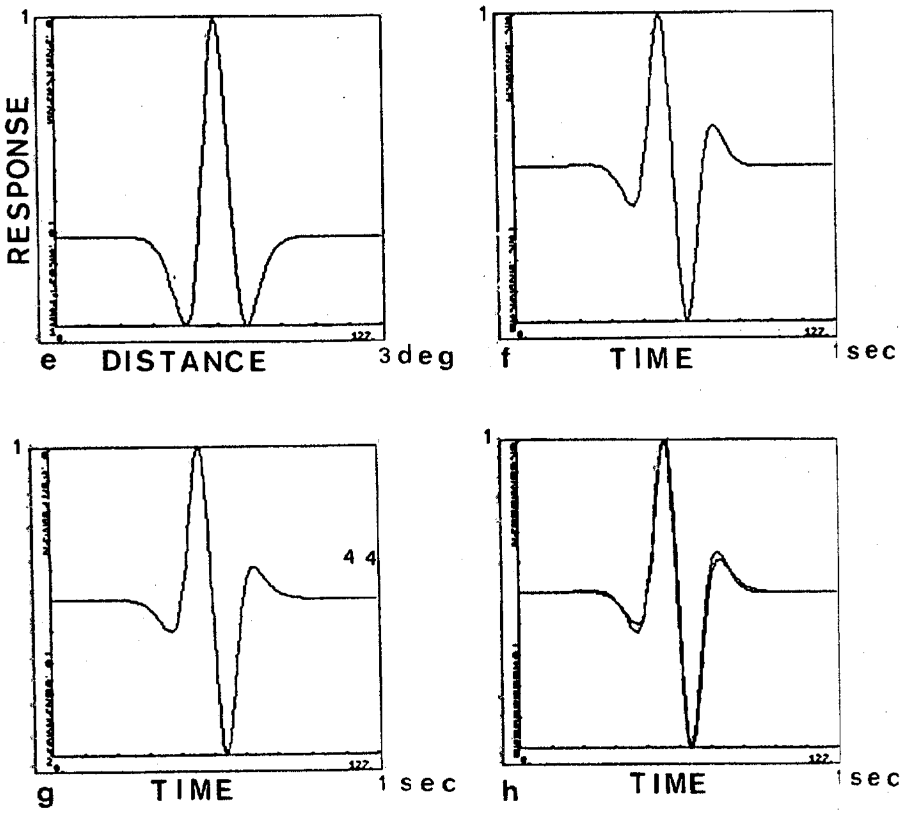
Linear Deviations

Any physical time-derivative operator will be extensive in time, not instantaneous, and this will have two consequences. (i) It will cease to function as a proper derivative for general signals whose time-constants are significantly shorter than those associated with the filter. In the frequency domain, the response of a physical device varies as $k\omega$ (where ω is the frequency) only within some range of values of ω. For the *Y*-channels, the overall time course is approximately 120 msec, and the upper limit for approximating the derivative is about 8 Hz. (ii) A delay will be introduced, because the channel signals the value of the derivative a short time ago. For the *Y*-channels this delay is about 50-60 msec. Some of this delay is

Figure 9. The computed response of the transient U-channel to a light edge (a-d) and to a thin bar (e-h) moving at 3 deg/sec. 9a: The output of the spatial filter ($\nabla^2 * I$) using the U-channel parameters from [Wilson & Bergen, 1979]. Ordinate: normalized response. Abscissa: distance (the entire range is 3 deg). 9b: The output of the temporal filter (using the contrast sensitivity curve in [Wilson, 1979] and the anti-symmetry assumption on the phase as explained in the text). Ordinate: normalized response. Abscissa: time (the entire range is 1 sec). 9c: The time derivative of 9a. 9d: Curves 9b and 9c are superimposed for comparison.

Figure 9e-f: The computed response to a 2' bar moving at 3 deg/sec. 9e: The output of the spatial filter. 9f: The output of the temporal filter. 2g: The time derivative of 9e. 9h: Curves 2b and 2c superimposed for comparison.


1
RESPONSE
a
DISTANCE
3deg
127
1
b
TIME
1sec
127
1
c
TIME
1sec
1
d
TIME
1sec


1
RESPONSE
e
DISTANCE
3deg
127
1
f
TIME
1sec
127
1
4 4
g
TIME
1 sec
127
1
h
TIME
1sec
127

compensated for by the different conduction velocities of the *X*- and *Y*-channels (Cleland, Dubin & Levick 1971).

Non-linear Deviations

The operator $\partial/\partial t(\nabla^2 G)$ is linear. As we have seen, even a linear device will inevitably deviate from a true time derivative. In addition, there are certain conditions under which *Y*-cells exhibit non-linear behavior (Euroth-Cugell & Robson, 1966; Hochstein & Shapley, 1976b). For example, experiments with gratings have revealed second-harmonic distortions, located in the surround region of the cell's receptive field, reminiscent of half-wave rectification (Hochstein & Shapley 1976b). In addition, the *Y*- but not *X*-cells exhibit the McIlwain periphery effect (Cleland, Dubin & Levick 1971).

The measurement of $\partial/\partial t(\nabla^2 G*I)$ is quite a complicated task and requires both spatial and temporal comparisons: the center must be compared with the surround, and the result "now" compared with the result a short time ago. In the retina, some of these components may be distorted, especially in view of the delay required for the comparison of values at two different times. Hochstein & Shapley's (1976b) findings suggest, for example, that the *Y*-cell surround receives a delayed contribution from the nearby units, about the size of the centres of local *X*-cell receptive fields, and that this delayed input may be a major source of the observed non-linearity. The non-linear effects are induced primarily by gratings (Euroth-Cugell & Robson 1966; Hochstein & Shapley 1976a; 1976b). For isolated edges and bars moving at moderate velocities, however, the Y-cells approximate $\partial/\partial t(\nabla^2 G * I)$, as we have seen in figure 8. Finally, it should be noted that for our scheme to function properly it is sufficient that the sign of the derivative, not its accurate value, be recovered.

The Construction of Directionally Selective Units

Our thesis is that the function of simple cells is to signal the presence, and direction of movement, of oriented zero-crossing segments; and that this is carried out by combining *X*- and *Y*-inputs roughly in the manner illustrated by figures 3b & c and 2c. There are several consequences of this thesis, and we now enumerate them, comparing them with the available neurophysiological information about simple cells.

Spatial Organization

The basic unit is the directionally selective oriented zero-crossing detector shown in figure 2c. Its receptive field has three components, sustained on-centre *X* inputs, sustained off-centre *X* units, and a *Y* input. The *X* units need to be all the same size, and arranged in two parallel columns not closer than *w* apart (where *w* is the width of the central excitatory regions of the *X*-cell receptive fields). The transient input can in principle be satisfied by a small number of *Y*-cells whose receptive fields lie between the two columns of *X*-cells.

Our ideal scheme requires a strict logical *AND* operation between the outputs of the subunits. In practise, this could be implemented by a strong multiplicative interaction between the columns and the *Y* input, and a weaker non-linearity down the columns. Such a unit would respond optimally to a moving zero-crossing segment that extended along the entire length of the columns, but it would also respond to shorter stimuli, and even to moving spots of light. More complicated receptive fields (*e.g.*, moving bars or slits) can be built up using these units as components.

It is hard to make quantitative predictions about the response of such units to arbitrary stimuli, because (a) the actual degree of non-linearity is unknown, and this is important in determining the relations between quantities like the length and separation of the columns and the orientation sensitivity of the unit; (b) there are many types of cortical cell, and probably only a minority of the measurements pertain directly to the units we describe.

The overall organization of the unit is in qualitative agreement with Hubel & Wiesel's (1962, 1968) description of simple cells. The non-linearity is supported by Schiller, Finlay & Volman (1976" pp. 1324-5).

If there is more than one size of *X*-unit (as required by Marr & Hildreth 1979), they should innervate different simple cells, because a given simple cell should receive *X*-inputs of only one size. Hence there should be at least two populations of simple cells, each tuned as narrowly as its (unoriented) *X*-cell input to a small range of (oriented) spatial frequencies (see Campbell, Cooper & Euroth-Cugell 1969).

According to our scheme, directional selectivity relies upon the combination of *X* and *Y* inputs (Schiller 1978), and should therefore be abolished by, for example, the selective removal of the *Y* input. This view contrasts with the notion that the *X* and *Y* channels feed two separate systems, one concerned with the analysis of "form" or "pattern", and the other, with "movement" (Tolhurst 1973; Kulikowski & Tolhurst, 1973; Ikeda & Wright, 1975a & b [Exp Brain Res]). According to our view, the sustained and transient channels are more properly viewed as two components of the same analytic system. (This does not, of course, exclude the possibility that the Y channels may also be involved with the control of eye movements).

Spatio-temporal Organization

Since Hubel & Wiesel first remarked on the sensitivity of simple cells to moving stimuli, the property of directional selectivity has been the subject of many studies (Pettigrew, Nikara & Bishop 1968; Bishop, Coombs & Henry 1971a & b; Goodwin, Henry & Bishop, 1975, in the cat; Schiller, Finlay & Volman 1976^l, and Poggio, Doty & Talbot, 1977, in the monkey).

If studied empirically, the directionally selective unit we described in figure 2c would be classified by Schiller *et al* 1976^l as an S_1 cell, responding to a single contrast edge moving in one direction. The size of its sensitive region would be of the order of w for an *X*-cell, about 15' at 4° eccentricity in the monkey, which is in rough agreement with Schiller *et al*'s findings. More complex units, like their S_2 unit (a directionally selective "bar" detector), can be built up in similar ways (e.g. $X^+Y^+X^-Y^-X^+$ would detect a dark bar moving to the right).

According to our earlier calculations, our proposed unit would be reliable for velocities up to at least 3^0/sec, and at the lower end, is limited only by the sensitivity of the Y-channel. The most sensible design for the Y-channel is therfore to make it as sensitive as possible to small values of $\partial/\partial t(\nabla^2 G*I)$. Consequently, one would expect the Y-channel to saturate early (as well as earlier for higher contrasts), giving a flat response curve for a given contrast as a function of velocity.

Goodwin, Henry & Bishop (1975 table 1) report velocity sensitivities down to 0.18°/sec in the cat, and psychophysical data (King-Smith, Riggs, Moore & Butler 1978) show that humans are sensitive down to about 1'/sec. Both these articles support our predictions about the flatness of the velocity sensitivity curve.

Our proposed unit will respond not only to continuous movement but also to discrete jumps. The response of simple cells to small jumps led Pettigrew, Nikara & Bishop (1968) to suggest that the overall unit is assembled from smaller directionally selective subunits. This would not be necessary for the unit we are proposing. Because it is a single unit, and not a composite of two adjacent detectors connected for example through some kind of delay, it will respond to any jump that is small enough and fast enough. The size of the jump must be such that both the initial and final positions lie between the centres of the X^+

and X^- receptive fields; and the interval between presentations of the initial and final cells cannot much exceed 60 msec, because of the temporal characteristics of the *Y*-channel. Goodwin & Henry (1975) found in the cat that a jump of 0.87' was sufficient to elicit a response.

Unlike the ***AND-NOT*** unit proposed by Barlow & Levick (1965) for the rabbit (and see also Schiller *et al.* 1976^{IV} p. 1369), our unit will not respond in the null direction at very low velocities, nor will it exhibit a "start-up" response if movement in the null direction is halted momentarily in the centre of the receptive field. These properties were confirmed by Goodwin, Henry & Bishop (1975).

Although most simple cells prefer moving stimuli, and many respond only to moving stimuli (Hubel & Wiesel 1962; 1968), it remains an open question whether all simple cells are directionally selective (Poggio, Doty & Talbot, 1977). According to our scheme, there are two basic ways of detecting stationary zero-crossings. If in an STS unit one replaces the excitatory T^+ input by an inhibitory input from T^-, the unit would respond to a zero-crossing that was stationary or moving in its preferred direction. Alternatively, one can omit the T input altogether (cf. figure 2b). In this case the unit would have no preferred direction.

There is no direct physiological evidence for cells of this latter type. We find this surprising in view of the simplicity and usefulness of such a unit. A possible candidate is Schiller *et al.*'s S_3 cell, which appears not to be directionally selective, responding equally to an edge of fixed spatial contrast moving in either direction. On closer examination, however, S_3 cells are somewhat enigmatic. If they were straightforward $\langle X^+ X^- \rangle$ units, the "sensitive" regions of such cells for edges moving in the two directions should coincide, yet in Schiller *et al.*'s figures, they are about 15' apart. It would therefore be interesting to know how certain it is that the separation is 15', and whether it is the same for all S_3 cells.

Intracortical structure

The recent studies by Sillito (1974, 1975a & b, 1977) suggest that both directional selectivity and orientation sensitivity involve inhibitory interactions. Directionality is abolished, and orientation sensitivity is impaired by bicuculline, which is thought to act antagonistically to GABA, thought to be a cortical inhibitory transmitter.

In our scheme, directionality depends wholly, and orientation sensitivity depends partly, on *AND*-like interactions between specific visual afferents. It is possible that the neural implementation of such interactions depends on the use of inhibitory interneurones. Although there are certainly many possible neural schemes, it is perhaps interesting to consider one in detail.

The basic *AND*-like operation can be implemented by a multiplication. Simple synaptic mechanisms of the type proposed by Torre & Poggio (1978) can achieve a multiplication, but also introduce a linear term that is unwanted here. It would be possible to eliminate this term *via* a linear inhibitory interneurone (*c.f.* Toyama, Matsumari, Ohno & Tokashiki, 1974 Figure 14B). If such inhibition were blocked, the linear term would reappear, destroying the *AND*-like nature of the interaction. This would abolish directionality but its disruption of orientation selectivity would be only partial, since the basic consequences of the geometry of the receptive field would remain.

The analysis of these effects will of course depend critically on the precise logical structure that is used for an STS unit -- whether for example one uses T^+ or (*NOT* T^-).

Experiments

In this section, we summarize the experiments that are important for the theory as set out here and by Marr & Hildreth (1979). We separate psychophysical experiments from neurophysiological ones, and divide the experiments themselves into two categories according to whether their results are critical and are already available (A), or are critical and not available and therefore amount to predictions (P). In the case of experimental predictions, we make explicit their importance to the theory by a system of stars; three stars indicates a prediction which, if falsified, would disprove the theory. One star indicates a prediction whose disproof remnants of the theory could survive.

Physiology

Retina and LGN

1 (A) LGN *X*-cells signal $\nabla^2 G*I$, using a DOG approximation (see figure 8 and Rodieck & Stone, 1965; Rodieck 1965; Enroth-Cugell & Robson 1966).

2 (Partly P***) LGN *Y*-cells signal $\partial/\partial t(\nabla^2 G*I)$. This is consistent with many published traces (see figure 8), but has not previously been formulated in this way. The three stars refer to obtaining reliably the sign of the derivative.

3 (P***) If there is no scatter in receptive field size at the retina, there must exist at least two populations of *X*-cells in the LGN. One population is formed by one-to-one connexions from the retina, the other by a small convergence (approximately five-to-one).

4 (P**) Response characteristics of *X*- and *Y*-cells. The response of *X*-cells should increase monotonically without saturating over a wide range of values of D^2G*I (e.g. 30:1). *Y*-cells on the other hand are expected to saturate at relatively low values of $\partial/\partial t(\nabla^2G*I)$. That is, the response curve of *Y*-cells as a function of velocity should be flat. Saturation should occur at higher velocities for lower contrasts. In addition, since the measurement of $\partial/\partial t(\nabla^2G*I)$ is more complex and involves a delay, it might be less reliable and more prone to non-linearities than the measurement of ∇^2G*I.

5 (P**) *Y*-cells should be sensitive to small displacements (of the order of 1'), and should respond to any jump that changes the value of ∇^2G*I in the appropriate direction.

6 (P**) Sizes of the channels. The values of w at the geniculate should be $\sqrt{2}$ times their sizes as measured psychophysically with elongated stimuli.

Striate Cortex

We now list the predicted properties of the basic directionally selective unit. Taking current neurophysiological data into account, it seems that the S_1 cells described by Schiller *et al.* (1976[1]) are the most likely candidates for such units.

7 (P^{***}) The basic directionally selective unit receives both *X* and *Y* inputs. Directional selectivity depends on the *Y* input and would be abolished by its complete removal. The output should be abolished or diminished, unless an S(*NOT* T)S unit is used.

8 (P^{***}) The basic directionally selective unit receives both on-centre and off-centre *X* inputs.

9 (partially P^{***}) The basic geometry of the unit should be as in figure 2, a column of on-centre *X*-units lying adjacent to a column of off-centre *X*-units. The centres of the *Y*-units (of which there must be at least one) should coincide roughly with the central axis of the unit.

10 (P^{**}) All of the *X* subunits should be of the same size. The *Y* subunits need not be the same size as the *X* subunits. For proper operation, *w* for the *Y* subunits should not be smaller than the separation of the two columns of *X* subunits.

11 (P**) For best operation, the separation of the two columns, and therefore the width of the "sensitive" region, should be approximately equal to w of the X units.

12 (P***) The preferred direction of a unit that receives X^+, X^-, and excitatory Y^+ input is from the X^+ to the X^-. If the unit receives excitatory Y^- input, the preferred direction is from the X^- to the X^+. If the Y input is inhibitory, the preferred directions are reversed, and the units would also respond to stationary stimuli.

Comments: This describes the geometry of the basic STS unit, a directionally selective edge (zero-crossing segment) detector realized physiologically by units like X^+, Y^+ and X^-. More elaborate units can be constructed in a similar way. As mentioned in the section on the construction of directionally selective units, one of Schiller ***et el.***'s S_2 cells might be constructed from $\langle X^+\ Y^+\ X^-\ Y^-\ X^+\rangle$ subunits. If this is in fact how they are made, S_2 cells should respond well to bars and dots moving in the preferred direction.

13 (A) Directionally selective units respond well to small displacements and low velocities, and the velocity response curve is relatively flat (Goodwin, Henry & Bishop, 1975; King-Smith, Riggs, Moore & Butler, 1978).

14 (P***) The unit should respond to any displacement that exceeds the

minimum detectable and which lies within the unit's sensitive region.

15 (A) The basic directionally selective unit shows no start-up and no slow-motion response in the null direction (Goodwin, Henry & Bishop, 1975).

16 (partly A, P*) Directional selectivity should be completely abolished, and orientation sensitivity impaired, by eliminating inhibitory interneurones that are driven by the specific visual afferents and which synapse to the directionally selective units (Sillito 1975b; 1977).

17 (P**) There should exist cells concerned with computing the local direction of motion. These cells should receive input from directionally selective units within a local neighbourhood. Their output should correspond to the allowed sector illustrated in figure 5.

Psychophysics

The psychophysical predictions are less critical than the physiological ones, because most of what the theory would predict for the input channels is already known, and the accessible characteristics of the later stages depend too much on quirks of the particular implementation that is used. Our predictions for the channels follow directly from the assumption that the sustained channels correspond to the *X*-cells, and the transient channels to the *Y*-cells, a view first suggested by Tolhurst (1973) and widely held in the literature.

Channel psychophysics

18 (A) The sustained channels signal (a DOG approximation to) $\nabla^2 G * I$ (Wilson & Giese 1977; Wilson & Bergen 1979).

19 (Almost A) The transient channels signal $\partial/\partial t(\nabla^2 G * I)$, using a DOG approximation for the spatial part of the function. It appears the time derivative is approximated by a biphasic odd function with time constants of about 60 msec (Watson & Nachmias 1977; Tolhurst 1977; Breitmeyer & Ganz 1977; Legge 1978; Wilson & Bergen 1979; Wilson 1979).

20 (A) There should be at least two sizes of sustained channel (Wilson & Giese 1978; Wilson & Bergen 1979; Marr & Hildreth 1979).

21 (A) If adaptation takes place at the S_1 cells, and these receive *X*-cell inputs of one size, then adaptation will be orientation, direction, and spatial-frequency selective.

22 (A) The *STS* unit should exhibit the reversed phi phenomnon described by Anstis [1970] and Anstis & Rogers [1975]. The T signal in the reversed phi presentation would be opposite in sign to the physical displacement, leading to signal of motion in the direction opposite to the physical displacement. Since *Y* cells are not color-specific,

reversed phi should depend on the overall brightness change, regardless of color, as observed by Anstis & Rogers.

Using directional selectivity

If tasks of separation are carried out using only information supplied by directionally selective units of the kind we have described, then they will exhibit the following characteristics:

23 (P***) The phenomena should occur only over short ranges (around w, or 15' at 5 degrees eccentricity) and short ISI's (not more than the total time course of the temporal component of the transient channel, about 120 msec).

24 (P**) If speed (and not direction) is the only available discriminant, separation should be difficult.

25 (P***) The amount of information that can be obtained from directional selectivity depends on the direction of movement and on the orientation of the moved elements (*cf.* figure 5). The same velocity field may be seen as coherent or incoherent depending on the orientations of the moved elements. The reason is that two nearby velocity vectors will produce the same directional sign on an element oriented roughly perpendicular to them, but different signs on an element whose orientation bisects them.

26 (P*) If the formation of coherent groups proceeds roughly in the manner of figure 5, one might expect to see clusters of locally coherent motions in even purely random display sequences.

Acknowledgement: we thank J. Batali for figure 5.

References

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

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

Barlow, H. B. 1953 Summation and inhibition in the frog's retina. ***J. Physiol. (Lond.), 119***, 69-88.

Barlow, H. B., & Levick, W. R. 1965 The mechanism of directional selective units in rabbit's retina. ***J. Physiol. (Lond.) 178***, 477-504.

Bishop, P. O., Coombs, J. S. & Henry, G. H. 1971 Responses to visual contours: Spatio-temporal aspects of excitation in the receptive fields of simple striate neurons. ***J. Physiol. (Lond.) 219***, 625-657.

Bishop, P. O., Coombs, J. S. & Henry, G. H. 1971 Interaction effects of visual contours on the discharge frequency of simple striate neurons. ***J. Physiol. (Lond.) 219***, 659-687.

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

Breitmeyer, B. & Ganz, L. 1977 Temporal studies with flashing gratings: Inferences about human transient and sustained channels. ***Vision Res. 17***, 861-865.

Campbell, F. W., Cleland, B. G., Cooper, G. F. & Enroth-Cugell, C. 1968 The angular selectivity of visual cortical cells to moving gratings. ***J. Physiol. (Lond.) 198***, 237-250

Cleland, G., Dubin, M. W. & Levick, W. R. 1971 Sustained and transient neurons in the cat's retina and LGN. ***J. Physiol. (Lond.) 217***, 473-496.

Cleland, B. G., Levick, W. R. & Sanderson, K. J. 1973 Properties of sustained and transient ganglion cells in the cat retina. ***J. Physiol. (lond.) 228***, 649-680.

Dreher, B. & Sanderson, K. J. 1973. Receptive field analysis: Responses to moving visual contours by single lateral geniculate neurons in the cat. ***J. Physiol. (Lond.) 234***, 95-118

Enroth-Cugell, C. & Robson, J. D. 1966 The contrast sensitivity of retinal ganglion cells of the cat. ***J. Physiol. (Lond.) 187*** 517-522.

Goodwin, A. W., Henry, G. H. & Bishop, P. O. 1975 Direction selectivity of simple striate cells: properties and mechanism. ***J.***

Neurophysiol. *38*, 1500-1523.

Goodwin, A. W. & Henry, G. H. 1975 Direction selectivity of complex cells in a comparison with simple cells. *J. Neurophysiol*. *38*, 1524-1540.

Graham, C. H. 1965 Perception of Movement. In: C. H. Graham (ed.) 1965 *Vision and visual perception*. New York: Wiley.

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

Hochstein, S. & Shapley, R. M. 1976a Quantitative analysis of retinal ganglion cell classification. *J. Physiol*. *262*, 237-264.

Hochstein, S. & Shapley, R. M. 1976b Linear and non linear spatial subunits in Y cat retinal ganglion cells. *J. Physiol. (Lond.) 262*, 265-284.

Ikeda, H. & Wright, M. J. 1972 Receptive field organization of 'sustained' and 'transient' retinal ganglion cells which subserve different functional roles. *J. Physiol. (Lond.) 227*, 769-800.

Ikeda, H. and Wright, M. J. 1975. Spatial and temporal properties of "sustained" and "transient" neurons in area 17 of the cat's visual cortex. *Exp. Brain Res*., *22*, 363-383.

Julesz, B. 1971 *Foundations of cyclopean perception*. University of Chicago Press.

Julesz, B. & Payne 1968

King-smith, P. E., Riggs A., Moore, R. K. & Butler, T. W. 1977 Temporal properties of the human visual nervous system. *Vision Res*. *17*, 1101-1106.

King-Smith, P. E. 1978 Analysis of detection of a moving line. *Perception 7*, 449-458.

Koffka, K. 1935 *Principles of Gestalt Psychology*. New York: Harcourt, Bracr & World.

Kulikowski, J. J. & Tolhurst, D. J. 1973 Psychophysical evidence for sustained and transient detectors in human vision. *J. Physiol*., *232*, 149-162.

Maturana, H. R., Lettvin, J. Y., McCulloch, W. S. & Pitts, W. H. 1960 Anatomy and physiology of vision in the frog (*rana pipiens*). *J. gen. Physiol*. *43*, 129-176.

Maturana, H. R. & Frenk, S. 1963. Directional movement and horizontal edge detectors in peigon retina. *Science*, ***142***, 977-979.

Marr, D. & Poggio, T. 1979 A computational theory of human stereo vision. *Oroc. R. Soc. Lond. B*, (in the press).

Marr, D., Poggio, T. & Ullman, S. 1979. Bandpass channels, zero-crossings, and early visual information processing. *J. Opt. Soc. Am.*, (in the press).

Marr, D. & Hildreth, E. 1979 Theory of edge detection. *Proc. R. Soc. Lond. B*, submitted for publication.

Miles, W. R. 1931 Movement interpretations of the silhouette of a revolving fan. *Amer. J. Psychol. 43*, 392-405.

Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. 1978 Spatial and temporal conrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex. *J. Physiol. (Lond.) 238*, 101-120.

Neuhaus, W. 1930. Experimentelle Untersuchung der Scheinbewegung. *Arch. ges. Psychol.*, *75*, 315-348.

Pettigrew, J. D., Nikara, T. & Bishop, P. O. 1968 Response to moving slit by single unit in cat striate cortex. *Exp. Brain Res. 6*, 373-390.

Poggio, T. & Reichardt, W. 1976. Visual control of orientation behavior in the fly. II. *Q. Rev. Biophys. 9(3)*, 377-438.

Poggio, G. F., Doty, R. W. Jr., & Talbot, W. H. 1977 Foveal striate cortex of behaving monkey: Single-neuron response to square-wave grating during fixation of gaze. *J. Neurophysiol. 40*, 1369-1391.

Poggio, T. & Reichardt, W. 1976 Visual control of orientation behaviour in the fly. Part II. Towards the underlying neural interactions. *Quarterly Rev. Biophys. 9*, 377-438.

Ramachandran, V. S., Madhusudhan, V. R. & Vidyasagar, T. R. 1973 Apparent movement with subjective contours. *Vision Res. 13*, 1399-1401.

Rashbass, C. 1976 Unification of two contrasting models of the visual incremental threshold. *Vision Res. 16*, 1281-1283.

Rodieck, R. W. & Stone, J. 1965a Response of cat retinal ganglion cells to moving visual patterns. *J. Neurophysiol. 28*, 819-832.

Rodieck, R. W. & Stone, J. 1965b Analysis of receptive fields of cat retinal ganglion cells. *J. Neurophysiol. 28*, 833-849.

Schiller, P. H., Finlay, B. L. & Volman, S. F. 1976a Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *J. Neurophysiol.* ***39***, 1288-1319.

Schiller, P. H., Finlay, B. L. & Volman, S. F. 1976b Quantitative studies of single-cell properties in monkey striate cortex. II. Orientation specificity and ocular dominance. *J. Neurophysiol.* ***39***, 1320-1333.

Schiller, P. H., Finlay, B. L. & Volman, S. F. 1976c Quantitative studies of single-cell properties in monkey striate cortex. III. Spatial frequency. *J. Neurophysiol.* ***39***, 1334-1351.

Sillito, A. M. 1975 The effectiveness of bicuculline as an antagonist of GABA and visually evoked inhibition in the cat's striate cortex. *J. Physiol. (Lond.)* ***250***, 287-304.

Sillito, A. M. 1975 The contribution of inhibitory mechanisms to the receptive field properties of neurons in the striate cortex of the cat. *J. Physiol. (Lond.)* ***250***, 305-329.

Sillito, A. M. 1977 Inhibitory processes underlying the directional specificity of simple, complex and hypercomplex cells in the cat's visual cortex. *J. Physiol. (Lond.)* ***271***, 699-720.

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

Tolhurst, D. J. 1975 Sustained and transient channels in human vision. *Vision Res.* ***15***, 1151-1555.

Tolhurst, D. J. & Movshon, J. A. 1975. Spatial and temporal contrast sensitivity of striate cortical neurons. *Nature*, ***257***, 674-675.

Torre, V. & Poggio, T. A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond. B* ***202***, 409-416.

Ullman, S. 1979a The interpretation of structure from motion. *Proc. R. Soc. Lond. B* ***203***, 405-426.

Ullman, S. 1979b *The interpretation of visual motion*. Cambridge, Mass. M.I.T. Press.

Wallach, H. & O'Connell, D. N. 1953 The kinetic depth effect. *J. Exp. Psychol* ***45***, 205-217.

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

Wertheimer, M. 1923 Untersuchungen zur Lehre von der Gestalt, II. *Psychol. Forsch. 4*, 301-350.

Wilson, H. R. 1978 Quantitative characterization of two types of line-spread function near the fovea. *Vision Res.*, *18*, 971-981.

Wilson, H. R. & Bergen, J. R. 1979 A four mechanism model for threshold spatial vision. *Vision Res.*, In press.

Wilson, H. R. 1979 Spatiotemporal characterization of a transient mechanism in the human visual system. (Submitted for publication).