Directional Selectivity and its Use in Early Visual Processing
Author(s): D. Marr and S. Ullman
Source: *Proceedings of the Royal Society of London. Series B, Biological Sciences*, Vol. 211, No. 1183 (Mar. 6, 1981), pp. 151-180
Published by: The Royal Society
Stable URL: http://www.jstor.org/stable/35563
Accessed: 08/05/2014 05:49

Proc. R. Soc. Lond. B **211**, 151–180 (1981)
Printed in Great Britain

Directional selectivity and its use in early visual processing

By D. Marr† and S. Ullman
M.I.T. Psychology Department and Artificial Intelligence Laboratory, 79 Amherst Street, Cambridge, Massachusetts 02139, *U.S.A.*

(*Communicated by S. Brenner, F.R.S. – Received 26 February* 1980)

[Plate 1]

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 (1980). That is, the image is filtered through centre–surround receptive fields, and the zero values in the output are found. In addition, the time derivative $\partial[\nabla^2 G(x, y) * I(x, y)]/\partial t$ is measured at the zero crossings, and serves to constrain the local direction of motion to within 180°. The direction of motion can be determined in a second stage, for example by combining the local constraints. The second part of the paper suggests a specific model of the information processing by the X and Y cells of the retina and lateral geniculate nucleus, and certain classes of cortical simple cells. A number of psychophysical and neurophysiological predictions are derived from the theory.

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 *et al.* 1959; Julesz 1971, ch. 4), and to recover three-dimensional structure from motion (Miles 1931; Wallach & O'Connell 1953; Ullman 1979*a*). There is good psychophysical evidence that at least two subsystems are involved in the analysis of motion (Braddick 1980), one dealing with short-range, short-term effects, and the other operating over much larger ranges and time intervals. This is hardly surprising in view of the wide range of information processing tasks related to motion. Simple tasks, such as the separation of a moving object from its background, require much less sophisticated analysis than the recovery of three-dimensional structure from visual motion (Ullman 1979*b*, §§4.2, 4.5). For discrete presentation, for example, the recovery of three-dimensional structure under orthographic projection requires the integration of information from at least three different views (Ullman 1979*a*), while for the separation of a moving object from its background, two frames separated by a short time interval are sufficient.

In this article, we consider some of the computational problems related to the

† During the course of the publication of this paper it was learnt with regret of the death of Dr David Courtenay Marr at the early age of 35.

first of Braddick's two categories. An example of a task in this category is the detection of the coherent motion of an unchanging region against an uncorrelated background, presented in two alternating frames. For a successful performance in this task, the motion has to be restricted to a range of up to 10–15′ and an interstimulus interval (i.s.i.) less than about 80–100 ms (Braddick 1973, 1974). Performance can be disrupted if a uniform bright field is flashed during the i.s.i. (Braddick 1973). The stimuli must be delivered to the same eye (or to both eyes together (Braddick 1974)), and patterns defined by chromatic but not luminance contrast do not show this effect (Ramachandran & Gregory 1978).

Apparent motion, on the other hand, can operate over much longer ranges (several degrees of visual angle (Zeeman & Roelofs 1953)) and times (400 ms (Neuhaus 1930)) and some kinds of apparent motion require long i.s.is to be perceived (200 ms in Ramachandran *et al.* (1973); 100–200 ms in Julesz & Payne (1968)). Motion is perceived whether the i.s.i. is bright or dark; successive frames may be delivered to different eyes (Shipley *et al.* 1945); and the stimuli may be defined by chromatic contrast alone (Ramachandran *et al.* 1973; Ramachandran & Gregory 1978). These longer range processes may be the mechanisms involved in the correspondence process and the recovery of structure from motion (Julesz & Payne 1968; Ullman 1979*b*).

The particular tasks that we address here are the measurement and use of information about direction of motion, and the article is organized into two parts. In the first part, we consider the computational requirements of this kind of task, analysing 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 (1980) 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.

1. Theoretical analysis

Tasks such as the separation of a moving object from its background 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 tasks such as separation 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′/s (Graham 1965, p. 575; King-Smith *et al.* 1977), and cortical simple cells in the cat can detect displacements as small as 0.87′ (Goodwin *et al.* 1975).

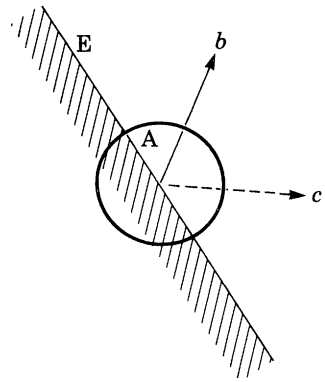


FIGURE 1. The aperture problem. If the motion of an oriented element is detected by a unit that is small compared to the moving element, the only information that can be extracted is the component of the motion perpendicular to the local orientation of the element. Looking at the moving edge E through 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*.

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 *et al.* 1979; Marr & Hildreth 1980), 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 (1980).

There are probably several biological systems that detect relative movement directly from intensity values, for example the motion detection system of the frog and rabbit retinae (Barlow 1953; Maturana *et al.* 1960; Maturana & Frenk 1963; Barlow & Levick 1965; Torre & Poggio 1978), that of the fly (Poggio & Reichardt 1976; Reichardt & Poggio 1979), and possibly also retinal W cells in higher mammalian visual systems. If one wishes to analyse the shape of the moving patch, however, 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 that one can extract is the component of the motion perpendicular to the local orientation. Motion along the contour will be invisible. Hence local

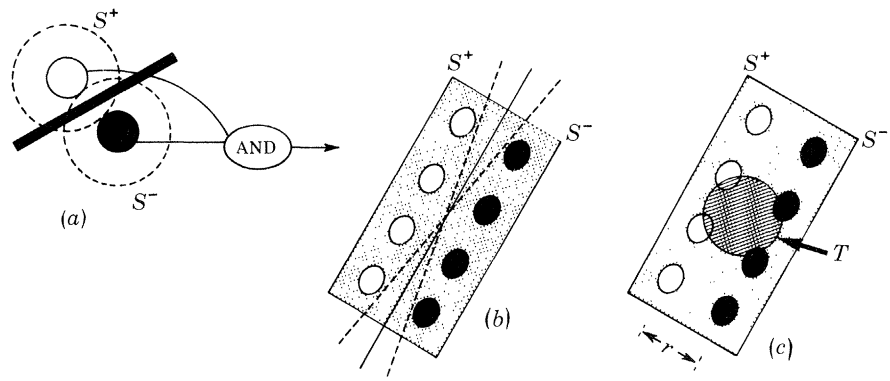


FIGURE 2. The detection of zero crossings. (*a*) S^- and S^+ units are combined through a logical AND operation. Such a unit would signal the presence of a zero crossing running between the two subunits. A row of similar units connected through a logical AND would detect an orientated zero crossing within the orientation bounds given roughly by the dotted lines in (*b*). In (*c*) a T unit is added to the detector in (*b*). If the unit is T^+, the STS unit 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 opposite direction.

measurements alone fail to give either the direction or the speed of movement, but they can restrict the direction to within 180°. In other words, the sign of the motion, relative to the zero-crossing segment, can be otained directly from the local measurement.

Therefore, use of zero crossings (or any oriented local element) as primitives suggests a division of 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: first, detection of an oriented zero-crossing segment, and, secondly, establishment of the sign of its motion. Zero-crossing segments may be detected by the mechanism shown in figure 2 (Marr & Hildreth 1980). The basic idea is that intensity changes may be detected by seeking the zero crossings of the convolution of the second-

derivative operator $\nabla^2 G$ with the image, where ∇^2 is the Laplacian and G is a two-dimensional radially symmetric Gaussian distribution. In fact,

$$\nabla^2 G(x, y) = -(1/\pi\sigma^4)(1 - r^2/2\sigma^2)\exp(-r^2/2\sigma^2). \quad (1)$$

To detect these zero crossings, Marr & Hildreth proposed the following scheme; if the values of the convolution $\nabla^2 G * I$, which we shall write as $S(x, y, t)$, are carried by

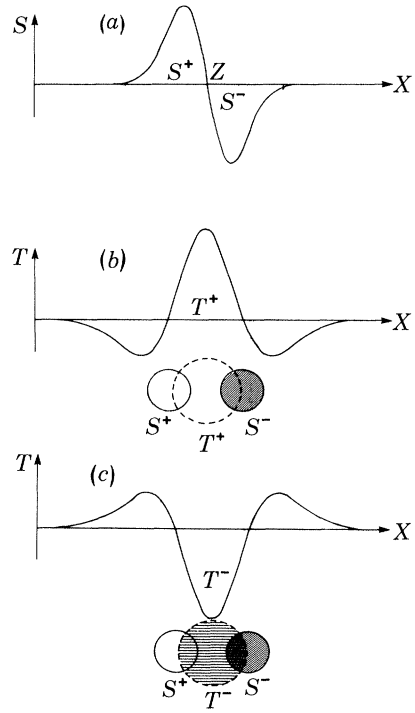


FIGURE 3. The value of $S = \nabla^2 G * I$, and of $T = \partial(\nabla^2 G * I)/\partial t$ in the vicinity of an isolated intensity edge. (*a*) S signal as a function of distance. The zero crossing in the signal corresponds to the position of the edge. Figure 3*b* shows the spatial distribution of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

two kinds of 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, on 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 2*a*). A row of such units will detect an oriented segment of zero crossings (figure 2*b*). Figure 3*a* illustrates the profile of the convolution values (of $\nabla^2 * GI$) in the vicinity of an isolated step change in intensity. S^+ in figure 3*a* 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 $\partial(\nabla^2 G * I)/\partial t$, at position Z, the direction of motion can be determined unambiguously. Figure 3b, c illustrates this. Let us define

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

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 neighbourhood 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, as for 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 orientated 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 are: 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. We see from equation (1) that the (two-dimensional) diameters d_S, d_T of the central excitatory regions of these channels are given by $d_S = 2\sqrt{2}\, s_S$, and $d_T = 2\sqrt{2}\, \sigma_T$. Let r denote the separation of the S^+ and S^- units (as in figure 2*c*).

To achieve maximal response, the optimal separation of the S^+ and S^- units should be $2\sigma_S$, which we shall write as 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 centre 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 $w_T = 2\sigma_T \geqslant r$, and, if $w_T \geqslant 2r$, 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 practice, its response at the lower end will be determined by its sensitivity, and that 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 ms, then the directionally STS-selective unit would function properly (assuming a single T unit midway between two S units separated by a distance r) if $v\tau + \frac{1}{2}r \leqslant \sigma_T$. Assuming again that $\frac{1}{2}r = \sigma_S$, 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 subunits. The optimal cover of a wide range of velocities may therefore require more than a single transient channel.

In summary then, there are various reasons why it is desirable for the T subunits to have larger receptive fields than the S subunits. Under certain conditions, however, this can introduce problems. If the zero crossings in the S channel occur too close together, the poorer spatial resolution of the T channel can lead to the wrong sign of the time derivative, essentially because the spatial average is being taken over too great an area. This can happen, for example, if a thin bar lies too close to a high contrast edge. Therefore, as in the static case (Marr & Hildreth 1980), special steps must be taken to detect configurations of nearby zero crossings, like slits and bars, and treat them appropriately.

Comparison with other schemes

The STS unit has several characteristics that make it well suited to the problem of detecting directional selectivity. (i) It requires only local measurements, occupying roughly the range of the first of Braddick's two processes. (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 course of the temporal filter is short. 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 comprehensive analysis of this type of scheme has been given by Poggio (1980), in connection with the system used by the housefly. The fly uses directly its detectors of intensity; for our purposes, one would use two zero-crossing detectors. The basic idea would then be essentially to detect motion by identifying the same zero crossing at two different locations at two different times. The motion-detecting circuitry would then connect one zero-crossing 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 distance between zero crossings in the input are adequately restricted, the system can detect movement. The range that we have in mind, from about 1′/s to over 3°/s is probably too large to be accommodated by a single such system, but it could be handled by several, operating in parallel.

The critical difference between a delay-based scheme of this kind and the one using the temporal derivative is that the latter does not have to wait until the zero crossing 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 zero-crossing detectors, it does not have to 'guess' that whatever is exciting one detector now is the same zero crossing that excited the other a short time ago. Guessing correctly all the time amounts to solving the correspondence problem, which is difficult (Ullman 1979*b*), and is furthermore unnecessary for the simpler problems posed by visual motion.

In addition, two-detector systems based on the use of a delay and an AND-NOT gate will 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. Barlow & Levick (1965) found this phenomenon in the rabbit retina, but Goodwin *et al.* (1975) looked for it 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 by means of 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 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, ch. 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 everywhere except at self-occluding boundaries*. 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. Because of the aperture problem, only the component of the motion perpendicular to the orientation of the element 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 4*a*), or, alternatively, that it is forbidden to lie on the other side (figure 4*b*). 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.

The way in which constraints can be combined is illustrated in figure 4*c*, *d*, for

the simple example 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 that they place on the direction of that motion by taking the union of their forbidden zones (figure 4*d*). 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 4*d*.

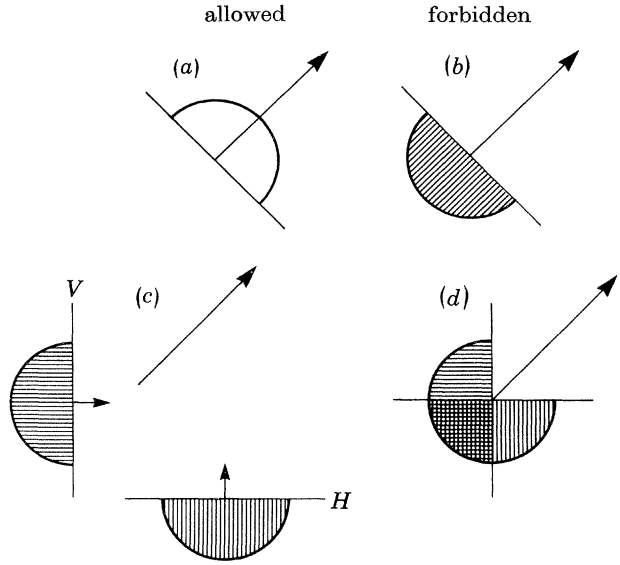


FIGURE 4. The combination of local constraints from *STS* units to determine the direction of motion. The constraint placed by a single *STS* unit is that the direction of motion must lie within a range of 180° on the allowed side of the oriented element (*a*), or, equivalently, it is forbidden to lie on the other side (*b*). (*c*) The forbidden zones for two oriented elements moving along the direction indicated by the arrow. The forbidden 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.

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, the motion of the two groups clearly cannot be compatible. Notice in this connection 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 orientated zero-crossing segment is v, and the found direction is at angle 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 to use this scheme for delineating moving objects is that objects are localized in space. If the objects are opaque and not self-occluding, their images will have an interior within which the forbidden zones in diagrams like figure 5*d* will be consistent, provided that they draw their elements from small neighbourhoods. The only exceptions to this can occur at zeros of the velocity fields, like the centre of a rotating disk. Such zeros in the velocity field can occur only at isolated points and, rarely, along lines.

Figure 5 shows an example of detection, by means of the above scheme, of a moving pattern embedded in a pair of random-dot images. A central square in figure 5*a* is displaced in figure 5*b* to the right, while the background moves in the opposite direction. Figure 5*c* depicts the zero-crossing contours of figure 5*a* filtered through $\nabla^2 G$. Figure 5*d* represents the values of the T channel, on the assumption that the two frames 5*a*, *b* were presented in rapid succession. Figure 7*a* shows the results of applying the *STS* operation to the zero crossings of figure 5*c*. The direction of movement has been colour-coded, as indicated by the star in the figure. As can be seen, red represents motion to the right, and green motion to the left. The central square is clearly delineated by discontinuities in the direction of motion. The same analysis was also applied to the natural images shown in figure 6, which are two successive frames taken from a 16 mm film of a basketball game. The results appear in figure 7*b*. Although much detail has been lost because of poor colour separation in the reproduction process, it can be seen, for example, that the left arm of player number 7 moved to the left and the player furthest to the right moved to the right. Because of the extreme sensitivity of the method, small registration errors, more or less unavoidable because of the way in which the two images were digitized, sometimes gave rise to spurious motion of the background.

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 shown in figure 4; one can use the raw output of the directionally selective units. The difficulty in this instance 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

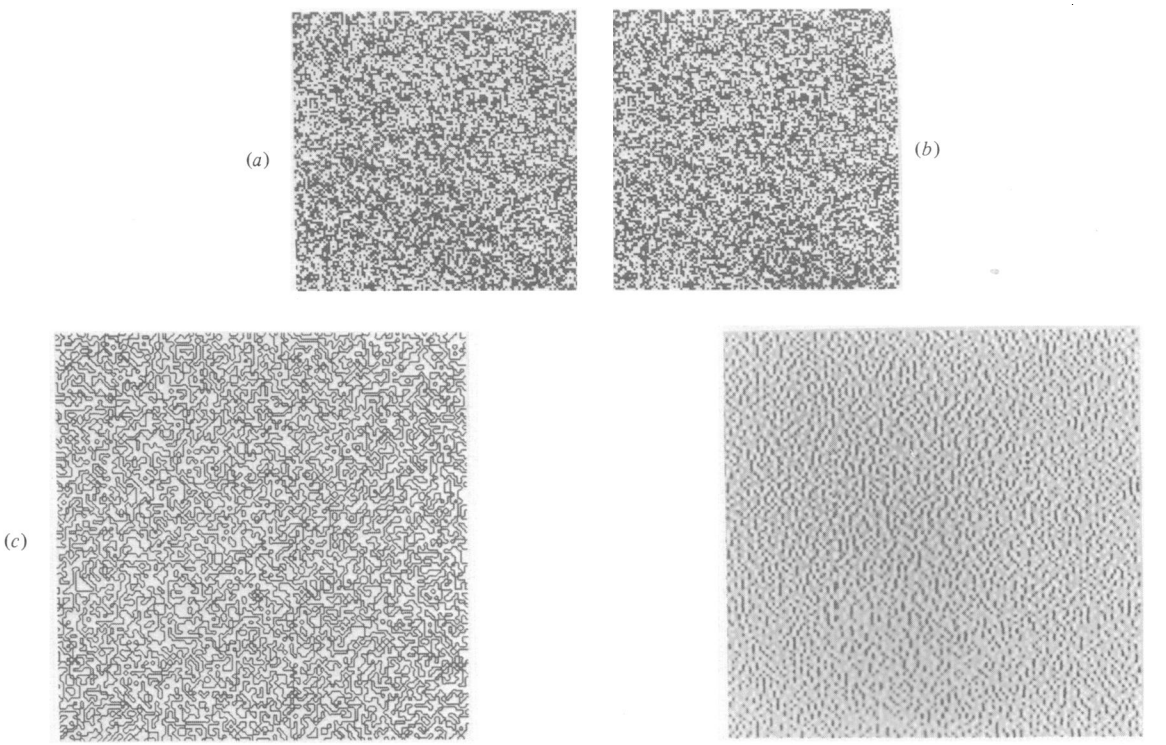


FIGURE 5. Separating a moving figure from its background by means of combinations of *STS* units. A central square in (*a*) is displaced in (*b*) to the right. The background in the two pictures moves the opposite way. Figure (*c*) shows the zero-crossing contours of (*a*) filtered through $\nabla^2 G$, and (*d*) shows the convolution of the difference between (*a*) and (*b*) with $\nabla^2 G$. If (*a*) and (*b*) are presented in rapid succession, the function shown in (*d*) approximates the value of $\partial(\nabla^2 G * I)/\partial t$. The images were 400 pixels by 400 pixels, the inner square is 200 by 200, each dot is 4 by 4, and the motions were one pixel.

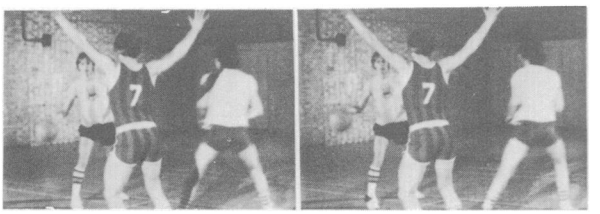
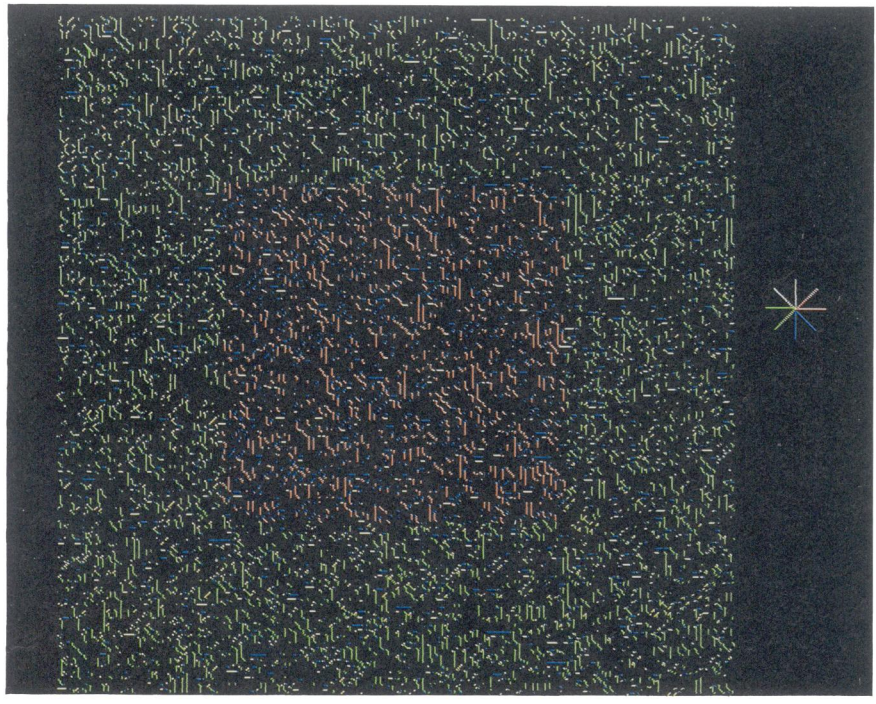


FIGURE 6. Two successive frames from a 16 mm film of a basketball game. The same analysis was applied to them as to the random dot patterns in figure 5.

DESCRIPTION OF PLATE 1

FIGURE 7. Motion assigned to zero crossings from the images of figures 5 and 6. The direction of motion was assigned according to the rules described in the text, and is represented here coded by colour, as defined by the star to the right of the images. In (*a*) the central square clearly moves right, while the surround moves left. In the zero crossings from the basketball game, the left arm of player number 7 moves to the left, while the player to the right of number 7 moves to the right.

(*a*)



(*b*)

FIGURE 7. For description see opposite.

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 *et al.* 1978) rather than truly disparity-sensitive, and the preferred motions of which are opposite in the two eyes. There is some evidence for the existence of such cells (Regan *et al.* 1979).

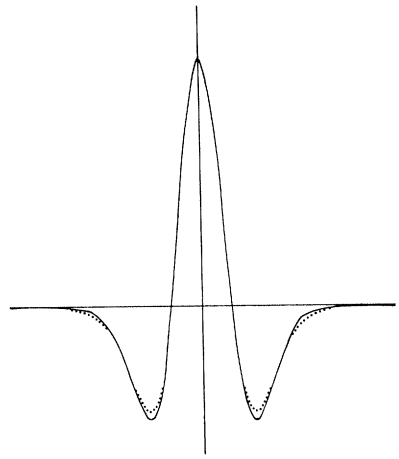


FIGURE 8. Comparing $\nabla^2 G$ to a difference of Gaussians (DOG). The dotted line is a DOG with $\sigma_i/\sigma_e = 1.6$. The solid line is an approximation of this DOG by means of $\nabla^2 G$. For more detail see Marr & Hildreth (1980, appendix B).

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 $\partial(\nabla^2 G * I)/\partial t$; 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 lateral geniculate nucleus (l.g.n.); 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 $\nabla^2 G * I$.

Spatial properties. Neurophysiology

The overall centre–surround organization of retinal ganglion cells was first discovered by Kuffler (1952, 1953). Rodieck & Stone (1965*b*) 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 Enroth-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 1980, appendix B). Figure 8 illustrates this point. The continuous curve in the figure is $\nabla^2 G$, 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 that can be easily assembled, for example, by subtracting two Gaussian 'pools' of receptors.

At the l.g.n., 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 *et al.* 1971; Hoffman *et al.* 1972; Cleland *et al.* 1973; Dreher & Sanderson 1973). Furthermore, Singer & Creutzfeldt (1970) and Levick *et al.* (1972) 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 probably little or no scatter in receptive field size (Peichl & Wassle (1979), in the cat). One possible way in which the two sizes of X and Y channels required by computational requirements (Marr & Hildreth 1980) and by psychophysical findings (Wilson & Bergen 1979) could arise, is from the limited convergence at the l.g.n. 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 *et al.* 1971; Cleland *et al.* 1973).

The overall response of retinal and l.g.n. X cells agrees closely with the predictions based on the $\nabla^2 G$ operation. Figure 9 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, fig. 6*d*, *e*) for the responses to an edge,

and from Rodieck & Stone (1965*a*, figs 1, 2), by means of traces from bars 1° and 5° wide. The predictions were calculated for bars of width $0.5w$ and $2.5w$, where $w = 2\sigma$ is the width projected onto one dimension of the central excitatory region of the receptive field. For the X-cell traces, records of on-centre cells were used for

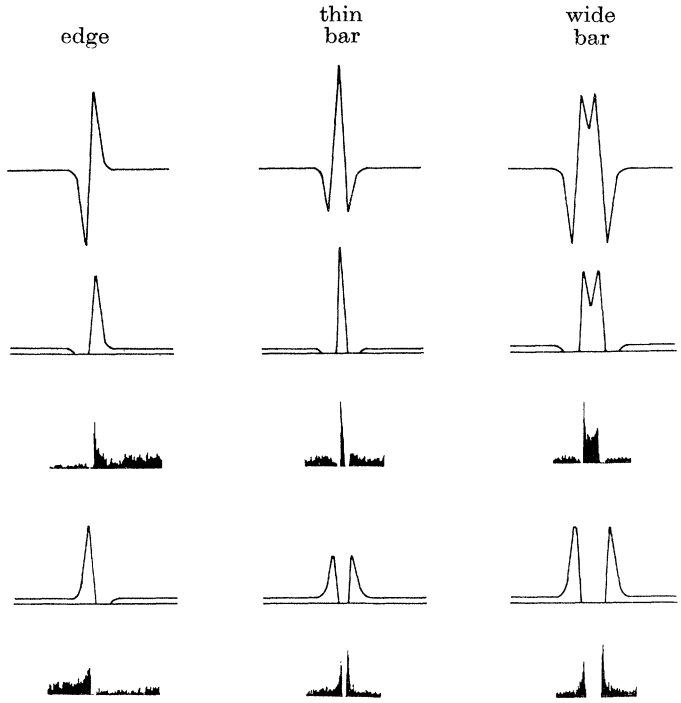


FIGURE 9. Comparison of the predicted responses of on- and off-centre X cells to electrophysiological recordings. The first row shows the response of $S = \nabla^2 G * I$ for an isolated edge, a thin bar (bar width $0.5w$, where $w = 2\sigma$, is the projected width of the central excitatory region of 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 across opposite units, or to stimuli of opposite contrast moving across the same units. The physiological responses are taken from Dreher & Sanderson (1973, fig. 6*d*, *e*) for the responses to an edge, and from Rodieck & Stone (1965*a*, figs 1, 2) by means of traces from bars 1° and 5° wide.

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 the paper by Rodieck & Stone preceded the distinction, by Enroth-Cugell & Robin (1966), between X and Y cells, and that most of the cells studied by Dreher & Sanderson (1973), including all those whose traces we have reproduced, were not classified as X or Y. Nevertheless their behaviours are quite different (compare figures 9 and 10), for example, the X-cell

response to a thin bar has the same profile as the Y-cell response to an isolated edge, and so one can 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 $\sigma_1 : \sigma_e = 1 : 1.75$, 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 circular 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 $\partial(\nabla^2 G * I)/\partial t$

We shall demonstrate that under 'reasonable' conditions, i.e. for edges and bars moving at velocities of up to a few degrees per second, Y-type retinal cells signal approximately $\partial(\nabla^2 G * I)/\partial t$. 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 convolution of a spatial receptive field sensitivity function with 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 l.g.n. levels, the Y cell's receptive field at a given eccentricity is spatially similar to that of the X cells (Rodieck & Stone 1965*a*, *b*; Rodieck 1965), only larger (Cleland *et al.* 1973).

Psychophysically, it has long been known 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 d.c. response at threshold (cf. Cowan 1977). There is some physiological evidence that the d.c. response may depend on the adaptation level (Enroth-Cugell & Shapley 1973*a*, *b*).

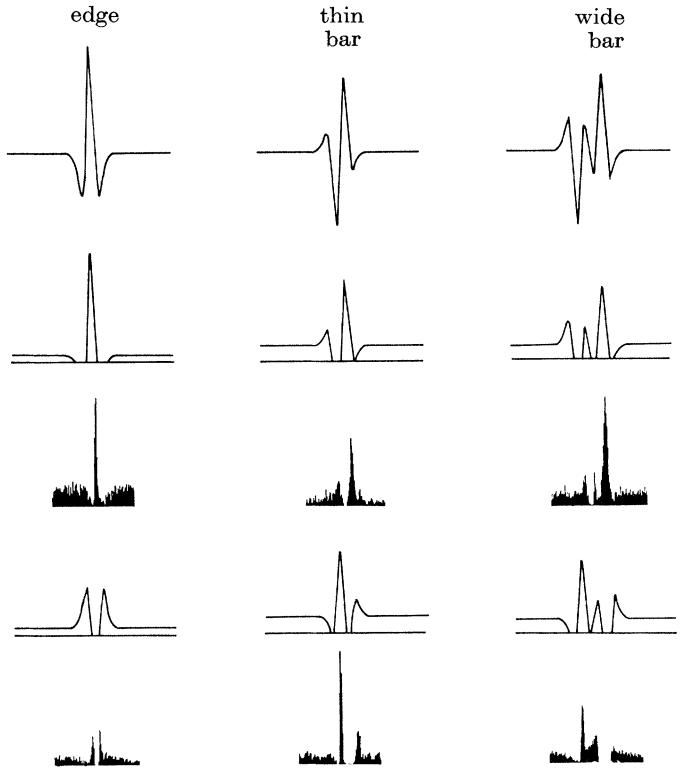


FIGURE 10. Comparison of the predicted responses of on- and off-centre Y cells to electrophysiological recordings. The first row shows the response of $T = \partial(\nabla^2 G * I)/\partial t$ for an isolated edge, a thin bar (bar width $0.5w$, where w is the projected width of the central excitatory 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 $\partial(\nabla^2 G * I)/\partial t$ 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, figs 6*b* and 8*a* for the edge responses, figs 1*d* and 2*c* for the thin bars, fig. 2*b* for the off-centre thick bar), and from Rodieck & Stone (1965, fig 5*b* for the on-centre response to a thick bar). The thin bars in these recordings were about $\frac{1}{2}°$ wide (0.4 and 0.6) and the thick bars about 5° (5.0 and 5.1). It can be seen that the observed responses are in close agreement with the predicted ones, even where both are elaborate (e.g. for wide bars).

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 the description, by Rodieck & Stone (1956*b*), of units whose response to a moving spot was 'directly correlated with the gradient of the receptive field as defined by

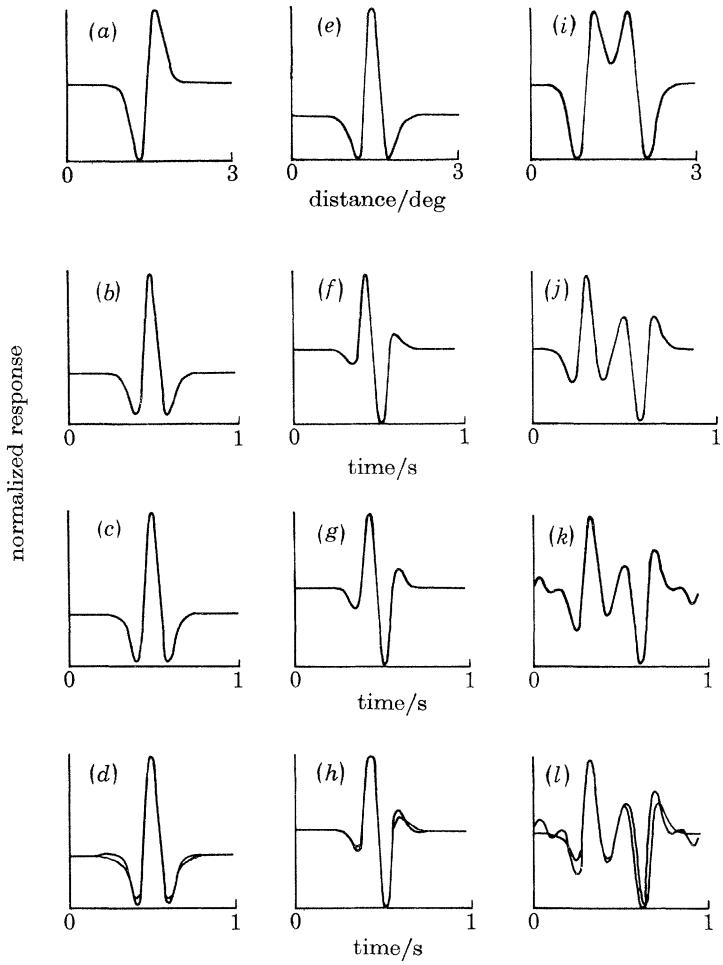


FIGURE 11. The computed response of the largest of Wilson's transient channels to a light edge ((*a*)–(*d*)), a thin bar ((*e*)–(*h*)) and a wide bar ((*i*)–(*l*)) moving at 3°/s. (*a*) The output of the spatial filter ($\nabla^2 G * I$) by means of the channel parameters from Wilson & Bergen (1979). (*b*) The theoretically predicted curve, on the assumption that the transient channel carried $\partial(\nabla^2 G * I)/\partial t$. In (*c*) is shown the output of the temporal filter by means of Wilson's contrast sensitivity curve and with the assumption that the filter is anti-symmetric; (*d*) compares (*b*) and (*c*). In (*e*)–(*h*) is shown the same analysis for a thin bar, 2′ wide, and in (*i*)–(*l*) that for a thick bar, 40′ wide. For isolated bars and edges, the psychophysical evidence is consistent with the idea that the transient channels approximate the function $\partial(\nabla^2 G * I)/\partial t$.

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 (up to a few degrees per second) are in a close agreement with the predictions based on the time-derivative operation $\partial(\nabla^2 G * I)/\partial t$. Figure 10

compares the observed responses of on- and off-centre cells, which we suppose to have been Y cells, to the predicted responses for various stimuli. All the stimuli were light (i.e. light edges, light bars); the thin bars were about $\frac{1}{2}°$ wide (0.4 and 0.6), and the thick bars, 5° (5.0 and 5.1). The traces are taken from Dreher & Sanderson (1973), figs 6*b* and 8*a* for the edge responses, figs 1*d* and 2*c* for the thin bars, fig. 2*b* for the off-centre thick bar), and from Rodieck & Stone (1965*a*, fig. 5*b* for the on-centre response to a thick bar). The predicted traces show pure values of $\partial(\nabla^2 G * I)/\partial t$, as in figure 9, the thicknesses of the thin and thick bars were respectively $0.5w$ and $2.5w$. It can be seen that the observed responses are in close agreement with the predicted ones, even where both are elaborate (e.g. for wide bars).

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 made 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 ms followed by a negative phase was explicitly suggested by Watson & Nachmias (1977), and further supported by Tolhurst (1975), Breitmeyer & Ganz (1977) and Legge (1978). The negative phase may be somewhat longer than the positive one, or may be followed by damped oscillations of small amplitude (Breitmeyer & Ganz 1977, fig. 3) without significantly affecting the results.

In the frequency domain, the temporal modulation transfer function (m.t.f.) was measured by Wilson (1980) for the largest of his two transient channels. This m.t.f. 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 ms long followed by a similar negative phase, one can approximate the phase relationships by assuming that the filter is an antisymmetric function about $t = 60$ ms. We have computed the results of applying this hypothetical filter to lines and bars moving at 3°/s. The results are shown in figure 11, and they are in a good agreement with the results of applying the operation $\partial(\nabla^2 G * I)/\partial t$.

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 nonlinear 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 transient channels, the overall time course is approximately 120 ms, 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 transient channels this delay is about 50–60 ms. Some of this delay may be compensated for by the different conduction velocities of the X and Y channels (Cleland *et al.* 1971).

Nonlinear deviations

The operator $\partial(\nabla^2 G)/\partial t$ 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 nonlinear behaviour (Enroth-Cugell & Robson 1966; Hochstein & Shapley 1976*b*). 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 1976*b*). In addition, Y cells produce a considerably stronger McIlwain periphery effect and shift effect (Cleland *et al.* 1971; Barlow *et al.* 1977; Derrington *et al.* 1979).

Although both X and Y cells exhibit nonlinearities, one might perhaps expect the Y cells to be more prone to them than are the X cells. The reason is that the measurement of $\partial(\nabla^2 G*I)/\partial t$ is quite a complicated task and requires both spatial and temporal comparisons: the centre 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. The findings of Hochstein & Shapley (1976*b*) 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 nonlinear effects are induced primarily by gratings (Enroth-Cugell & Robson 1966; Hochstein & Shapley 1976*a*, *b*). For isolated edges and bars moving at moderate velocities, however, the Y cells approximate $\partial(\nabla^2 G*I)/\partial t$, as we have seen in figure 10. 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, or oriented zero-crossing segments, and that this is carried out by combining X and Y inputs roughly in the manner illustrated by figures 2*c*, 3*b*, *c*. 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 2*c*. Its receptive field has three components, sustained on-centre X units, 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 less than about w apart (where w is the projected 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 practice, this could be implemented by a strong multiplicative interaction between the columns and the Y input, and a weaker nonlinearity 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. As we have seen earlier, more complicated receptive fields (e.g. moving bars or slits) are probably also necessary. These can be built up using XYX units as components, and may correspond for example to units like the S_2 cells of Schiller *et al.* (1976*a*, *b*, *c*).

It is hard to make quantitative predictions about the response of such units to arbitrary stimuli, because: (*a*) the actual degree of nonlinearity 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; and (*b*) there are many types of cortical cell, and probably only a minority of the measurements pertain directly to the units that we describe. The overall organization of the unit is, however, in qualitative agreement with the Hubel & Wiesel (1962, 1968) description of simple cells. The nonlinearity is supported by Schiller *et al.* (1976*b*, pp. 1324–5).

If there are two or more sizes of X units (as required by Marr & Hildreth 1980), 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 (unorientated) X-cell input to a small range of (orientated) spatial frequencies (see Campbell *et al.* 1969).

According to our scheme, directional selectivity relies upon the combination of X and Y inputs, 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 1975*a*, *b*). 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 in other tasks, such as 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 *et al.* 1968; Bishop *et al.* 1971*b*; Goodwin *et al.* 1975 (in the cat); Schiller *et al.* 1976*a* and Poggio *et al.* 1977 (in the monkey)).

If studied empirically, the directionally selective unit that we describe in figure 2*c* would be classified by Schiller *et al.* 1976*a* 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 the findings of Schiller *et al.* 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°/s, being limited at the lower end only by the sensitivity of the Y channel. The most sensible design for the Y channel is therefore to make it as sensitive as possible to small values of $\partial(\nabla^2 G*I)/\partial t$, the price of sensitivity being poor resolution in speed. 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 *et al.* (1975, tbl. 1) report velocity sensitivities down to 0.18°/s. in the cat, and psychophysical data (King-Smith *et al.* 1978) show that humans are sensitive down to about 1′/s. Both these articles support our ideas 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 *et al.* (1968) to suggest that the overall unit is assembled from smaller directionally selective subunits. This would not be necessary for the unit that 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 ms, 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.

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 *et al.* 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 2*b*). 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 the S_3 cell of Schiller *et al.* 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 figures by Schiller *et al.* (e.g. 1976*a*, fig. 5) 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 of Sillito (1974, 1975*a*, *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 interneurons. 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 interneuron (cf. Toyama *et al.* 1974, fig. 15b). 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 selectively 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 in Marr & Hildreth (1980). 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 critical and not available and therefore amount to predictions (P). For experimental predictions, we make explicit their importance to the theory by a system of asterisks; three asterisks indicates a prediction that, if falsified, would disprove the theory. One asterisk indicates a prediction whose disproof remnants of the theory could survive.

Physiology

Retina and l.g.n.

(i) (A) L.g.n. X cells signal $\nabla^2 G*I$, by means of a DOG approximation (see figure 9 and Rodieck & Stone (1965*b*), Rodieck (1965), Enroth-Cugell & Robson (1966)).

(ii) (*Partly* P***.) L.g.n. Y cells signal $\partial(\nabla^2 G*I)/\partial t$. This is consistent with many published traces (see figure 10), but has not previously been formulated in this way. The three asterisks refer to obtaining reliably the sign of the derivative.

(iii) (P***) If there is no scatter in receptive field size at the retina, there must exist at least two populations of X cells at the point where zero-crossing segments are detected. One population could be formed by one-to-one connections from the retina, the other by a small convergence (approximately five-to-one).

(iv) (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 $\nabla^2 G*I$ (e.g. 30:1). Y cells on the other hand are expected to saturate at relatively low values of $\partial(\nabla^2 G*I)/\partial t$. 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 $\partial(\nabla^2 G*I)/\partial t$ is more complex and involves a delay, it might be less reliable and more prone to nonlinearities than the measurement of $\nabla^2 G*I$.

(v) (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 $\nabla^2 G*I$ in the appropriate direction.

(vi) (P**) Sizes of the channels. If the receptive field of a channel is measured psychophysically with elongated stimuli, as was done by Wilson, the width w of the central region should be $1/\sqrt{2}$ of the diameters of the underlying physiological concentric receptive fields.

Striate cortex

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

(vii) (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.

(viii) (P***) The basic directionally selective unit receives both on-centre and off-centre X inputs.

(ix) (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.

(x) (P**) 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.

(xi) (P**) For best operation, the separation of the two columns, and therefore the width of the 'sensitive' region, should be approximately equal to $w = 2\sigma$ of the X units.

(xii) (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 Y^-. More elaborate units can be constructed in a similar way. As mentioned in the section on the construction of directionally selective units, one of the Schiller *et al.* S_2 cells might be constructed from $\langle X^+\ Y^+\ X^-\ Y^-\ X^+\rangle$ subunits. If this is in fact how they are made, one would predict that S_2 cells should respond well to bars and dots moving in the preferred direction.

(xiii) (A) Directionally selective units respond well to small displacements and low velocities, and the velocity response curve is relatively flat (Goodwin *et al.* 1975; King-Smith *et al.* 1978).

(xiv) (P***) The unit should respond to any displacement that exceeds the minimum detectable and which lies within the unit's sensitive region.

(xv) (A) The basic directionally selective unit shows no start-up and no slow-motion response in the null direction (Goodwin *et al.* 1975).

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

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 but not universally adopted in the literature.

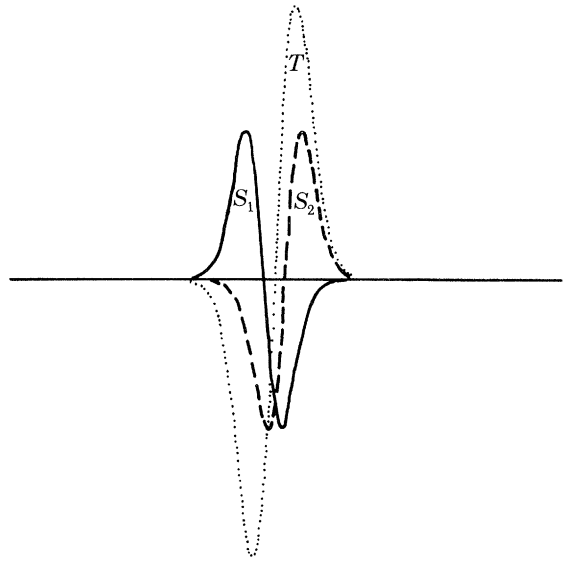


FIGURE 12. Anstis's reversed phi phenomenon is exhibited by the STS directionally selective unit. Curve S_1 shows the value of $\nabla^2 G$ for the first frame. In the second frame the edge is moved to the right and its contrast is reversed (curve S_2). Curve T shows the value of the time derivative. By applying the rules to the zero crossings in either S curve, motion to the left would be inferred.

Channel psychophysics

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

(xviii) (Almost A.) The transient channels signal $\partial(\nabla^2 G * I)/\partial t$, by means of a DOG approximation for the spatial part of the function. It appears that the time derivative is approximated by a biphasic odd function with each phase lasting about 60 ms (Watson & Nachmias 1977; Tolhurst 1975; Brietmeyer & Ganz 1977; Legge 1978; Wilson & Bergen 1979; Wilson 1980).

(xix) (A) There should be at least two sizes of sustained channel (Wilson & Giese 1977; Wilson & Bergen 1979; Marr & Hildreth 1980).

(xx) (A) If adaptation takes place at the S_1 cells and if these receive X cell

inputs of one size, then adaptation will be orientation-, direction-, and spatial-frequency-selective.

(xxi) (A) The *STS* unit should exhibit the reversed phi phenomenon described by Anstis (1970) and by Anstis & Rogers (1975). The *T* signal in the reversed phi presentation would be opposite in sign to the physical displacement, leading to motion being signalled in the direction opposite to the physical displacement as illustrated in figure 12. Since *Y* cells are not colour-specific, reversed phi should depend on the overall brightness change, regardless of colour, as observed by Anstis & Rogers.

Using directional selectivity

If a task is performed with use only of information supplied by directionally selective units of the kind we have described, then the following characteristics should be evident.

(xxii) (P***) The phenomena should occur only over short ranges (around w, or 15′ at 5° eccentricity) and short i.s.is (somewhat less than the total time course of the temporal component of the transient channel, which is about 120 ms).

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

We thank J. Batali for figures 5, 6 and 7. This work was carried out at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology, support for which is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643 and in part by National Science Foundation Grant MCS77-07569.

References

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

Anstis, S. M. & 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., Derrington, A. M., Harris, L. R. & Lennie, P. 1977 The effects of remote retinal stimulation on the responses of cat retinal ganglion cells. *J. Physiol., Lond.* **269**, 177–194.

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*a* 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*b* Interaction effects of visual contours on the discharge frequency of simple striate neurons. *J. Physiol., Lond.* **219**, 659–687.

Braddick, O. J. 1973 The masking of apparent motion in random-dot patterns. *Vision Res.* **13**, 355–369.

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

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

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. & Enroch-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.

Cowan, J. D. 1977 Some remarks on channel bandwidths for visual contrast detection. *Neurosci. Res. Progr. Bull.* **15**, 492–517.

Derrington, A. M., Lennie, P. & Wright, M. J. 1979 The mechanism of peripherally evoked responses in retinal ganglion cells. *J. Physiol., Lond.* **289**, 299–310.

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.

Enroth-Cugell, C. & Shapley, R. M. 1973*a* Adaptation and dynamics of cat retinal ganglion cells. *J. Physiol., Lond.* **233**, 271–309.

Enroth-Cugell, C. & Shapley, R. M. 1973*b* Flux, not retinal illumination, is what cat ganglion cells really care about. *J. Physiol., Lond.* **233**, 311–326.

Gibson, E. J., Gibson, J. J., Smith, O. W. & Flock, H. 1959 Motion parallax as a determinant of perceived depth. *J. exp. Psychol.* **8**(1), 40–51.

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

Goodwin, A. W., Henry, G. H. & Bishop, P. O. 1975 Direction selectivity of simple striate cells: properties and mechanism. *J. Neurophysiol.* **38**, 1500–1523.

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

Hassenstein, B. & Reichardt, W. 1956 Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption der Rüsselkäfers. *Chlorophanus. Z. Naturf.* **116**, 513–524.

von der Heydt, R., Adorjani, C., Hanny, P. & Baumgartner, G. 1978 Disparity sensitive and receptive field incongruity of units in the cat striate cortex. *Expl Brain Res.* **31**, 523–545.

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

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

Hoffmann, K.-P., Stone, J. & Sherman, S. M. 1972 Relay of receptive-field properties in dorsal lateral geniculate nucleus of the cat. *J. Neurophysiol.* **35**, 518–531.

Hubel, D. H. & Wiesel, T. N. 1961 Interactive action in the cat's lateral geniculate body. *J. Physiol., Lond.* **155**, 385–398.

Hubel, D. H. & Wiesel, T. N. 1962 Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol., Lond.* **160**, 106–154.

Hubel, D. H. & Wiesel, T. N. 1968 Receptive fields and functional architecture of monkey striate cortex. *J. Physiol., Lond.* **195**, 215–243.

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

Ikeda, H. & Wright, M. J. 1975*b* Retinotopic distribution, visual latency and orientation tuning of 'sustained' and 'transient' cortical neurons in area 17 of the cat. *Expl Brain Res.* **22**, 384–397.

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

Julesz, B. & Payne 1968 Differences between monocular and binocular stroboscopic movement perception. *Vision Res.* **8**, 433–444.

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.

Koffka, K. 1935 *Principles of Gestalt psychology.* New York: Harcourt, Brace & World.

Kuffler, S. W. 1952 Neurons in the retina: organization, inhibition and excitation problems. *Cold Spring Har. Symp. quant. Biol.* **17**, 281–292.

Kuffler, S. W. 1953 Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* **16**, 37–68.

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

Legge, G. 1978 Sustained and transient mechanisms in human vision: temporal and spatial properties. *Vision Res.* **18**, 69–81.

Levick, W. R., Cleland, B. G. & Dublin, M. W. 1972 Lateral geniculate neurons of cat: retinal inputs and physiology. *Invest. Ophthalmol.* **11**(5) 302–311.

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

Marr, D. & Poggio, T. 1979 A computational theory of human stereo vision. *Proc. R. Soc. Lond.* B **204**, 301–328.

Marr, D., Poggio, T. & Ullman, S. 1979 Bandpass channels, zero-crossings, and early visual information processing. *J. opt. Soc. Am.* **69**(6) 914–916.

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 pigeon retina. *Science, N.Y.* **142**, 977–979.

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

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

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

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

Poggio 1980 (In preparation.)

Poggio, G. F., Doty, Jr, R. W. & 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. II. Towards the underlying neural interactions. *Q. Rev. Biophys.* **9**, 377–438.

Ramachandran, V. S. & Gregory, R. L. 1978 Does colour provide an input to human motion perception? *Nature, Lond.* **275**, 55–56.

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

Regan, D., Beverley, K. I. & Cynader, M. 1979 Stereoscopic subsystem for position in depth and for motion in depth. *Proc. R. Soc. Lond.* B **204**, 485–501.

Reichardt, W. E. & Poggio, T. 1980 Visual control of flight in the fly. In Recent theoretical developments in neurobiology (ed. W. E. Reichardt, V. B. Mountcastle & T. Poggio). *Neurosci. Res. Prog. Bull.*, pp. 135–150.

Rodieck, R. W. 1965 Quantitative analysis of cat retinal ganglion cell responses to visual stimuli. *Vision Res.* **5**, 583–601.

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

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

Schiller, P. H., Finlay, B. L. & Volman, S. F. 1976*a* 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. 1976*b* 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. 1976*c* Quantitative studies of single-cell

properties in monkey striate cortex. III. Spatial frequency. *J. Neurophysiol.* **39**, 1334–1351.

Shipley, W. G., Kenny, F. A. & King, M. E. 1945 Beta-apparent movement under binocular, monocular, and interocular stimulation. *Am. J. Psychol.* **58**, 545–549.

Sillito, A. M. 1974 Effects of iontophoretic application of bicuculline on the receptive field properties of simple cells in the visual cortex of the cat. *J. Physiol., Lond.* **242**, 127–128P.

Sillito, A. M. 1975*a* 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*b* 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.

Singer, W. & Creutzfeldt, O. D. 1970 Reciprocal lateral inhibition of on- and off-center neurons in the lateral geniculate body of the cat. *Expl Brain Res.* **10**, 311–330.

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, Lond.* **257**, 674–675.

Torre, V. & Poggio, T. 1978 A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond.* B **202**, 409–416.

Toyama, K., Matsunami, K. Ohno, T. & Tokashiki, S. 1974 An intracellular study of neuronal organization in the visual cortex. *Expl Brain Res.* **21**, 45–66.

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

Ullman, S. 1979*b* *The interpretation of visual motion*. Cambridge, Massachusetts: 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 on 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. 1980 Spatiotemporal characterization of a transient mechanism in the human visual system. *Vision Res.* **20** (5), 443–452.

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

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

Zeeman, W. P. C. & Roelofs, C. O. 1953 Some aspects of apparent motion. *Acta psychol.* **9**, 159–181.