Visual Motion and its Role in the Stabilization of Gaze
Eds. F. A. Miles and J. Wallman

CHAPTER 1

Detecting visual motion: Theory and models

Alexander Borst and Martin Egelhaaf

Max-Planck-Institut für biologische Kybernetik, Spemannstraße 38, D-7400 Tübingen, Germany

1. Introduction

Whenever an animal is moving in its environment or when an object moves in front of the eyes, the visual system is confronted with motion. However, this motion information is not explicitly represented in the two-dimensional brightness pattern of the retinal image as sensed by the array of photoreceptors. Instead, motion has to be computed from the brightness changes in the retinal image, as one of the first and most basic processing steps in the visual system. Understanding this primary process of motion detection is important, because the performance of all subsequent processing steps depends on the information about the moving visual surround which is provided by it. In the following we shall see that it cannot be taken for granted that a biological motion detection system is able to determine correctly the direction and speed in which the different segments of the retinal image are moving.

1.1. What is motion?

In physics, the velocity of a moving object is defined as the object's spatial displacement over time. In order to detect this displacement visually, physical motion has to go along with changes in the spatial brightness distribution on the retina. If this is not the case, for example, when a homogeneous sphere is rotating under constant illumination, no motion can be seen by any mechanism whatsoever. What characterizes visual motion? Consider a sequence of video frames in which a vertical, black bar is moving to the right in front of a white background (Fig. 1A). Since the brightness does not change along the vertical axis, the representation of this stimulus can be further simplified by plotting only the horizontal brightness distribution against the temporal axis. When sketched in this space-time domain the moving bar appears as a slanted black stripe (Fig. 1B). The slope of the stripe reflects the amount of bar displacement between consecutive frames, or in other words, the velocity of object motion: The steeper the slope, the lower is the velocity. To further illustrate how motion can be distinguished from non-motion in the space-time domain, another example is shown in Fig. 2. Here, instead of a black bar, a stimulus pattern is used where the brightness is distributed statistically along one spatial dimension (Fig. 2, top). For simplicity, the brightness assumes only two values, either black or white. If the pattern is stationary, there is no change in brightness over time at any location in space. In the space-time domain this results in vertical stripes (Fig. 2A). If the pattern is moving to the right, the stripes in the space-time domain become slanted towards the right (Fig. 2B), as was the case for the moving bar. From the different appearance of the representation of static and moving patterns in the space-time domain one might tend to conclude that, in order to tell motion from non-motion, it would be sufficient to detect temporal brightness changes at a fixed spatial

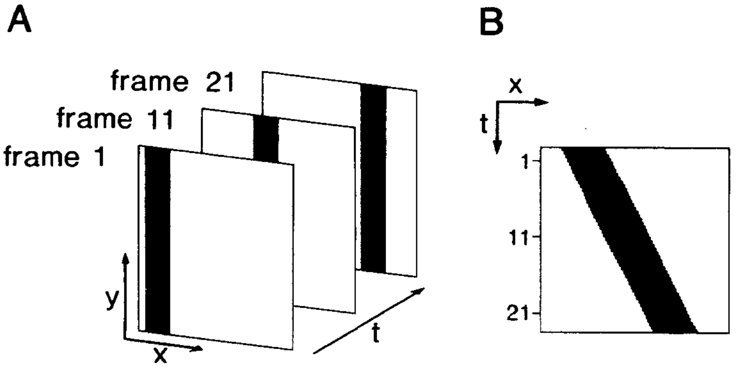

Fig. 1. **A**, A black vertical bar is moving to the right as shown by its position in three frames (frame 1, frame 11, frame 21) out of a longer sequence of consecutive frames. **B**, By omitting the vertical spatial axis this motion can be pictured in the space-time domain as a slanted black stripe. The times are indicated at which the frames shown in **A** were taken.

location. In the example shown in Fig. 2C it is illustrated that this procedure is not adequate: Random brightness fluctuations also lead to temporal changes, although there is certainly no coherent motion in this stimulus.

1.2. How to detect motion?

In all examples shown in Fig. 2 the brightness of the stimulus pattern changes in space. In addition, for non-static patterns brightness also changes

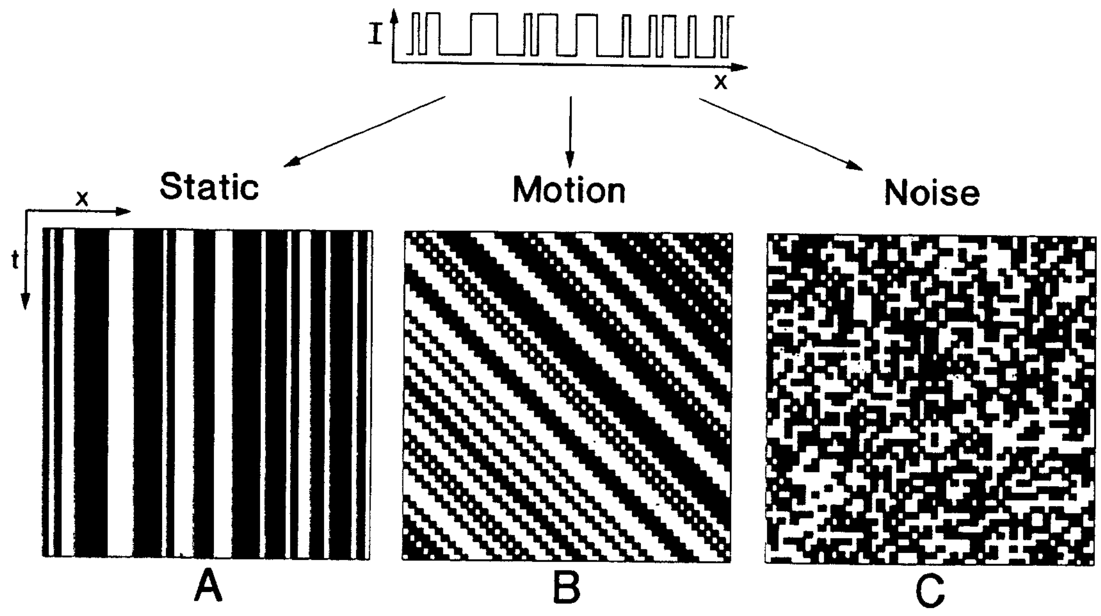

Fig. 2. A one-dimensional brightness distribution (top) is shown in the space-time domain for three different situations: **A**, When it is static, **B**, when it is moving at constant velocity to the right, and **C**, when the spatial brightness distribution fluctuates randomly in time. Motion is characterized by a slanted orientation of the resulting pattern in the space-time domain.

over time. The slanted bars in the space-time domain seem to be the critical feature which allows a moving pattern to be distinguished from both static patterns and random brightness fluctuations. How can this feature be exploited to build a mechanism detecting motion? There are basically two ways in which this can be done: i) the brightness at a particular location and at a given instant of time is compared with the brightness value at a neighboring location measured some time later (Fig. 3A); and ii) the brightness change in time observed at one location is related to the brightness change in space at the same location. For simplicity, the spatial and temporal brightness changes are approximated in Fig. 3B by the brightness difference in space and time, respectively (see below). These two strategies of motion detection form the basis of two classes of model mechanisms, the so-called *correlation-* and the *gradient-detectors*.

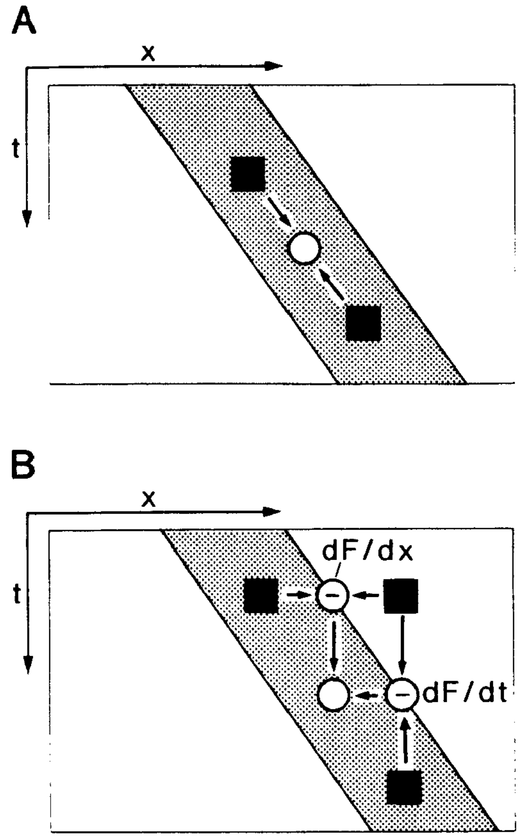

Fig. 3. Two different ways to measure visual motion. The slanted orientation of the pattern in the space-time domain can be detected by correlating the brightness values along this line (**A**), or by relating the spatial change of brightness d*F*/d*x* to its temporal change d*F*/d*t* (**B**). These are the two basic computational principles underlying *correlation detectors* (A) and *gradient detectors* (**B**), respectively. Points of measurement are indicated by black squares, operations are symbolized by open circles. In the case of the gradient detector, the spatial and temporal gradients are approximated in the most primitive way by the difference of the brightness values as measured at two spatial locations and two instants of time, respectively.

2. Models of motion detection

Various versions of correlation- and gradient-detectors have been proposed in the past to underly motion detection in biological systems (for review see Reichardt, 1961, 1987; Ullman, 1983; Buchner, 1984; Hildreth and Koch, 1987; van Santen and Sperling, 1985; Borst and Egelhaaf, 1989). Depending on the scientific tradition and the problems to be explained, some of these models were formulated as a kind of general purpose motion detection mechanism without any sophisticated elaborations, whereas others include rather specific assumptions and, consequently, assume highly specific functional properties. In the first step of our analysis, we restrict ourselves to motion in one spatial dimension such as was shown in Figs. 1–3; later we turn to motion in two spatial dimensions (Section 5).

2.1. Correlation detectors

The first correlation detector was proposed on the basis of experimental studies on the optomotor behavior of insects (Hassenstein and Reichardt, 1956; Reichardt, 1961; Reichardt and Varjú, 1959; Varjú, 1959). According to common usage (van Santen and Sperling, 1984, 1985) this correlation detector will be referred to here as the *Reichardt detector*. In subsequent studies, the Reichardt model has also been applied to explain motion detection in different vertebrate species including man (e.g. Foster, 1969, 1971; Wilson, 1985; van Doorn and Koenderink, 1982a,b; van Santen and Sperling, 1984; for review see Borst and Egelhaaf, 1989). Such a detector consists of two mirror-symmetrical subunits (Fig. 4A,B). In each subunit, the signals derived from two neighboring inputs are multiplied with each other after one of them has been shifted in time with respect to the other by a delay line or some sort of temporal low-pass filter. In the case of a pure delay, such a subunit performs the operation pictured in Fig. 3A. The final detector response is given by the difference of the output signals

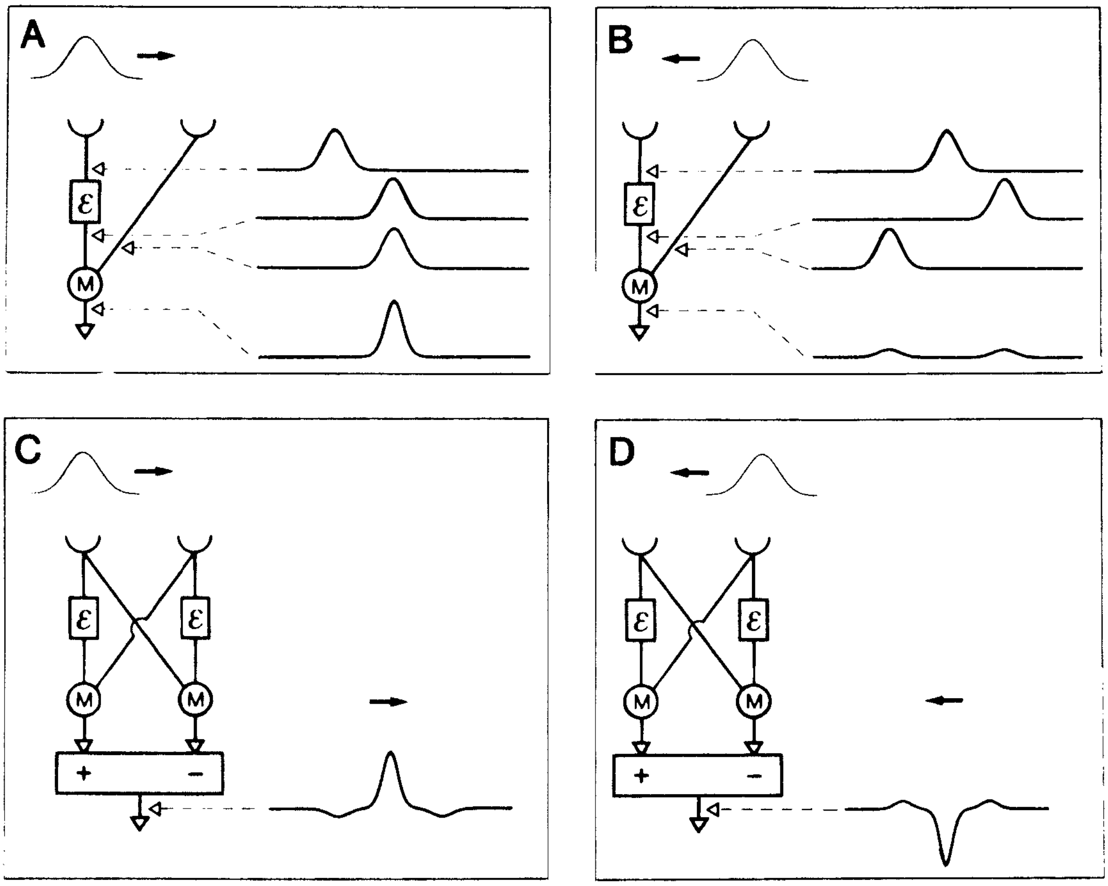

Fig. 4. Reichardt model. The detector (C,D) consists of two mirror-symmetrical subunits. In each of the subunits the detector input signals are multiplied (*M*) after one of them has been delayed (ε) with respect to the other by some sort of temporal filter. Both subunit outputs are subtracted to give the final detector output signal. To facilitate an understanding of the operations performed by such a detector, the response of a single subunit to motion in opposite directions is considered first. When the stimulus (represented by a Gaussian brightness distribution superimposed on the background brightness) passes the detector input channels they are activated in sequence. **A**, When the pattern moves in the detector's "preferred direction", the temporal separation of the signals is, more or less, compensated for by the delay in the left branch of the detector. At the appropriate velocity, both signals coincide at the multiplication stage giving rise to a comparatively large output signal. **B**, When the stimulus moves in the detector's "null direction", the temporal sequence of the signals in both channels is reversed. The delay further increases their time interval which results in two comparatively small response peaks (it should be noted that these responses arise in this example because the background brightness is represented to some extent at the movement detector input). **C**,**D**, Subtraction of the output signals of both subunits eliminates those response components which are independent of the direction of motion. The responses to motion in opposite directions show the same amplitude and time course but different signs.

of both subunits (Fig. 4C,D). The combination of a temporal delay and a multiplication is the reason why this type of detector measures the degree of coincidence of the signals in its input channels or, in other words, performs on average a spatio-temporal cross-correlation. If an object passes the detector it activates both input channels in sequence. For motion in one direction (the "preferred direction") the delay line compensates for the time-shift between the signals (Fig. 4A), for motion in the opposite direction (the "null direction"), it increases the time-shift (Fig. 4B). This leads to a large response in one but not in the other case. Although a motion detector as shown in Fig. 4A,B is directionally selective, it responds also to correlated input signals which are independent of the direction of motion, such as to the background brightness or its temporal changes. Since these direction-independent signals are the same in both mirror-symmetrical subunits of the complete detector, they are eliminated by subtracting the two subunit outputs from each other. The subtraction stage, therefore, increases the direction selectivity of the movement detector. If a motion detector responds to motion in opposite directions with the same amplitude but an opposite sign (Fig. 4C,D) it is called "fully opponent".

Various elaborations of the basic Reichardt model have been proposed to accommodate this motion detection scheme to the particular performance of different biological systems. The simplest schemes rely on different types of linear spatial and/or temporal frequency filters which are inserted in the input channels of the movement detector (Reichardt, 1961; van Santen and Sperling, 1984, 1985; Borst and Egelhaaf, 1989). However, Reichardt detectors can also operate on representations of the retinal image which are the result of more sophisticated pre-processing. For instance, the retinal input signals may pass nonlinearities, such as a full-wave or half-wave rectifier (Chubb and Sperling, 1988; Sperling, 1989). The movement detectors may also be fed by elements which respond only to a particular texture, spectral frequency composition or even to motion. Hence, such motion detectors are specialized to detect motion of image patches of a specific color, texture or of moving objects, such as of a spinning wheel. All these different ways of pre-processing are currently being discussed with respect to motion perception of man (Lelkens and Koenderink, 1984; Cavanagh and Mather, 1989; Zanker, 1990).

Perhaps the simplest correlation-type movement detector has been proposed by Barlow and Levick to explain their experimental findings on directionally selective ganglion cells in the rabbit retina (Barlow and Levick, 1965). The Barlow-Levick model (Fig. 5) is almost identical with respect to its layout with one subunit of the basic Reichardt model. It consists of two input lines carrying the brightness signals which are compared after one of the signals has been delayed. In contrast to the Reichardt model, this comparison is accomplished by a special logical gate, an "AND-NOT" or "veto" gate. This means that the detector's activity is suppressed when both input signals arrive simultaneously at the "AND-NOT" gate (Fig. 5A). The corresponding direction of motion is, therefore, the detector's null direction. For motion in the detector's preferred direction the veto signal arrives too late to have an effect.

Another type of correlation detector, the so-called *motion energy model* (Adelson and Bergen, 1985; Watson and Ahumada, 1985; Watson et al., 1986), was inspired by the formal equivalence of detecting the orientation of, for instance, a stationary bar in ordinary two-dimensional spatial displays and of a moving one-dimensional stimulus in the space-time domain (see Figs. 1 and 2). It is a commonplace notion that the orientation of spatial patterns can be detected by sensors with receptive fields consisting of separate elongated excitatory and inhibitory subregions; in the spatial domain such elements have been described, for instance, in the visual cortex of cats and monkeys (e.g. Hubel and Wiesel, 1962). By replacing formally one spatial dimension by the temporal dimension, motion of an object, i.e.

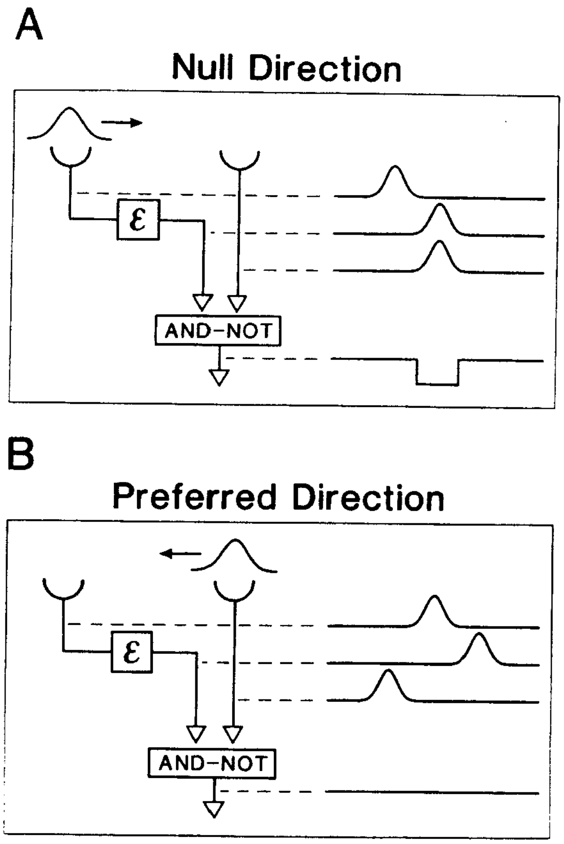

Fig. 5. Barlow-Levick model. The detector consists of two input lines, a delay line (ε) and a logical AND-NOT gate. Such an element leads always to an output, unless both of its input channels are activated simultaneously. **A**, When a visible object (drawn as a Gaussian brightness distribution) passes the detector from the left to the right, the input lines are activated in sequence. The delay line compensates for the temporal separation between the signals in such a way that, at a given velocity, they coincide at the AND-NOT gate. As a consequence, the output goes down. **B**, For motion in the opposite direction the temporal sequence of the input signals is reversed so that they are completely separated in time when arriving at the gate. Hence, the detector output does not change and remains at its normal level.

orientation in the space-time domain, can be detected by sensors with so-called oriented spatio-temporal receptive fields (Fig. 6A). As is illustrated in Fig. 6B such sensors can be constructed in two steps: (i) the retinal image is filtered in parallel by two input channels which contain different combinations of appropriate spatial and temporal linear filters. (ii) The filtered signals of the two input channels are either added or subtracted (Adelson and Bergen, 1985). However, because these sensors consist of purely linear filters they reverse the sign of their output if the contrast of the stimulus pattern is inverted. This flaw is overcome by a nonlinear operation such as squaring of the output of the linear filters (Adelson and Bergen, 1985). Although the resulting element is directionally selective, its response is temporally modulated when, for instance, a grating pattern passes its input channels (see Section 3.3). In the energy model these temporal response modulations are prevented by summing two such elements the inputs of which are filtered in a specific way. If these elements form a so-called "quadrature pair", i.e. if their responses to grating patterns with sinusoidal brightness distribution are 90° out of phase (Fig. 6C), their summed output shows a steady amplitude during motion with a constant velocity. Since the output of such a squared quadrature pair extracts a measure of the motion energy, these models are called motion energy models (Adelson and Bergen, 1985). Such a detector responds only to motion in its preferred direction and not to motion in its null direction. To obtain positive and negative responses to motion into these directions, respectively, two oppositely oriented squared quadrature pairs are subtracted from each other. Then the response is "fully opponent" (Fig. 6D).

If the Reichardt model as shown in Fig. 4 is equipped with the same spatial and temporal filters in its input channels, it assumes the same specific functional characteristics as the energy model and may even become mathematically equivalent (van Santen and Sperling, 1985; Adelson and Bergen, 1985). This identity, however, only holds for the final, fully opponent output signal of both detectors. The responses of the more peripheral processing stages may differ considerably.

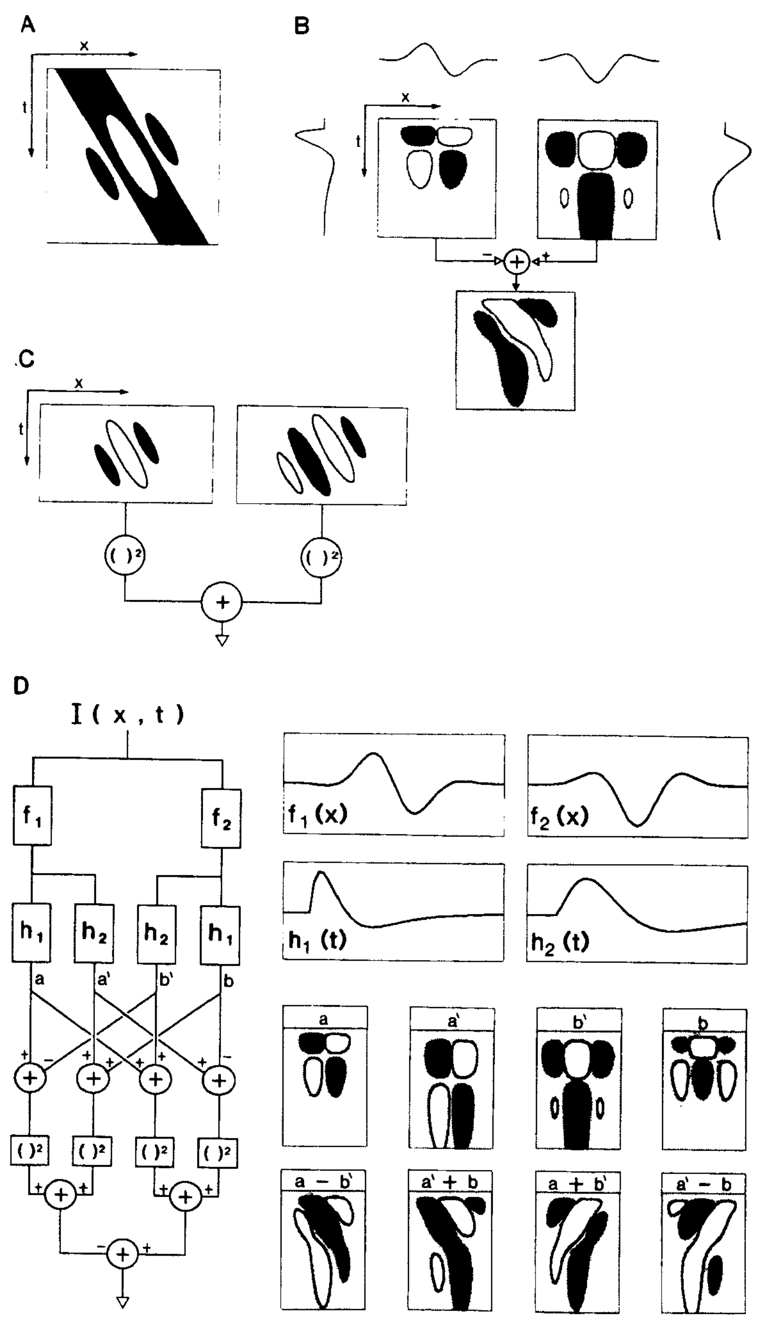

Fig. 6. Motion energy model. **A**, A dark bar moving on a bright background from the left to the right with constant velocity as pictured in the space-time domain, together with a spatio-temporal filter exactly matched to this particular motion. Shown are the excitatory and the inhibitory lobes of its receptive field. **B**, Filters with a slanted or oriented spatio-temporal receptive field can be constructed from a linear combination of two different "separable" filters which are the product of a spatial and a temporal filter function (shown by the curves on top and on the side of the receptive fields, respectively). C, Two oriented filters with the same orientation in the space-time domain but with 90° phase shift form a so-called quadrature pair. When squared and added, the resulting signal is independent of the spatial phase of the pattern. **D**, An energy model for motion detection (left) together with the receptive fields of its various processing

stages (right). It is composed of components as shown in **B** and **C**. The input signal representing the brightness as a function of space and time first passes two spatial band-pass filters f_1 and f_2 with receptive fields which are 90° out of phase ("even" and "odd"). ($f_1(x)$ and $f_2(x)$ are their spatial impulse responses.) Each of these signals is then split up and fed through the units h_1 and h_2 which have different temporal band-pass characteristics. ($h_1(t)$ and $h_2(t)$ are their temporal impulse responses.) The resulting four units (a,a',b',b) show separable spatio-temporal receptive fields. They still are not oriented in the space-time domain. Oriented or slanted receptive fields are obtained by linear combinations of these units. Two out of the resulting four units are oriented to detect motion to the left and two are oriented to detect motion to the right. The members of each quadrature pair differ along the spatial axis in that one is derived from an even and the other from an odd receptive field. Each of them is squared and added with its counterpart. The signals of both pairs are then subtracted. The result is a fully opponent unit which signals motion in opposite directions with the same time course and amplitude but an opposite sign. The optimum velocity and optimum spatial frequency range of such a detector is set by the band-pass characteristics of the input filters f_1,f_2 and h_1,h_2, respectively. Note that most of the computational expenditure of the energy model is needed to construct linear filters with oriented spatio-temporal receptive fields. (Modified from Adelson and Bergen, 1985.)

2.2. Gradient detectors

Gradient detectors calculate a motion estimate by relating the spatial change of brightness to the temporal change at a given location of the image (Fig. 3B). These changes are related physically to each other: The temporal brightness change $\partial F/\partial t$ induced by a moving object at point x is proportional to the brightness change along the spatial axis (i.e. the brightness gradient) $\partial F/\partial x$; ($F=F(x,t)$ designates the brightness of the stimulus pattern as a function of the spatial coordinate x and time t). The proportionality constant in this relation is given by the negative pattern velocity $-v=\mathrm{d}x/\mathrm{d}t$ (note: leftward and rightward motion are conventionally defined as negative and positive velocities, respectively!):

$$\partial F/\partial t = (\partial F/\partial x) \cdot (\mathrm{d}x/\mathrm{d}t) \tag{1}$$

The velocity can, thus, be formally recovered at least for motion in one spatial dimension by dividing the temporal brightness change by the spatial gradient.

$$v = -\mathrm{d}x/\mathrm{d}t = -\ (\partial F/\partial t)\ /\ (\partial F/\partial x) \tag{2}$$

The gradient scheme was first applied in computer analysis of video image sequences (Limb and Murphy, 1975; Fennema and Thompson, 1979) and only later discussed with respect to biological motion vision (e.g. Hildreth and Koch, 1987).

However, this formalism does not represent a model in the same sense as the different correlation models which were discussed above: It is just a mathematical formulation of the relation of spatial and temporal brightness changes induced by a moving object. It does not specify how the temporal and spatial change of the image brightness should be measured. There are various ways in which this can be done. The spatial pattern gradient may be approximated in the simplest way by the difference between the brightness at two adjacent image points and the temporal derivative by the temporally high-pass filtered brightness signal (Fig. 7) (Buchner, 1984). The response properties of this simple type of hardware-implemented gradient detector differ considerably from the ones of the "ideal" mathematical scheme (see below). However, more sophisticated approaches are available in computer vision to approximate spatial and temporal brightness changes (e.g. Ballard and Brown, 1982; Horn, 1986). The performance of gradient models based on such approximations may then come much closer to the properties of the formalism given by Eqn. 2 and, thus, to an "ideal" gradient scheme. Whereas the gradient detectors discussed so far operate directly on the brightness values of the retinal image, modified models have also been proposed which are fed by spatially band-pass filtered versions of the retinal image (Wang et al., 1989; Srinivasan, 1990).

As a consequence of the divisive nonlinearity which is characteristic of gradient detectors, the velocity estimate may become indefinite. This can happen when the denominator of the expression relating the approximation of the temporal

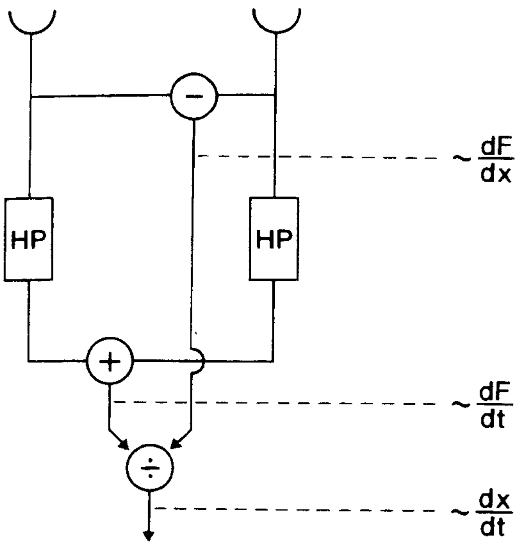

Fig. 7. Gradient model. This detector computes the ratio of the temporal and spatial change of the pattern brightness. In the model shown here the temporal gradient dF/dt is approximated by a temporal high-pass filter (HP) and the spatial gradient dF/dx is approximated by the difference between the two input signals. Here, the mean of both high-pass filtered input signals is used in the numerator of the division operation. Note, that the performance of this simple hardware implemented gradient scheme differs substantially from the ideal mathematical one.

and spatial brightness gradient, respectively, becomes zero while the numerator assumes a finite value. For instance, this situation may be encountered if the detector looks at a spatially homogeneous pattern which just changes its brightness as a consequence of changes in illumination. Appropriate means have been proposed to overcome this problem; amongst others, these include the addition of a small constant in the denominator of the ratio between the approximations of the

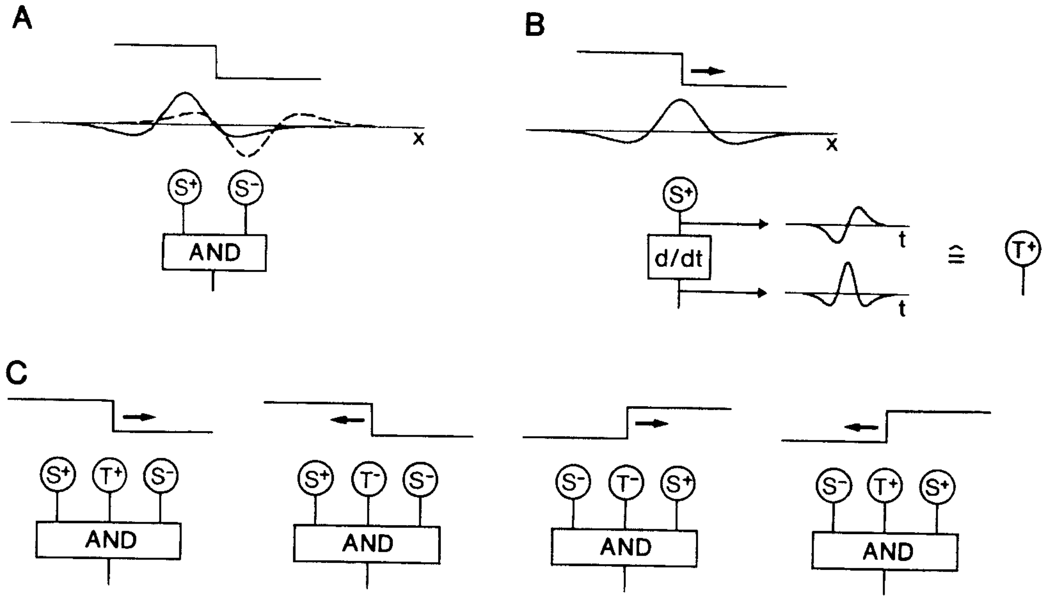

Fig. 8. Marr-Ullman model. **A**, An on-center or S^+ unit and an off-center or S^- unit having overlapping receptive fields are maximally excited by an edge of appropriate contrast polarity. By feeding both signals through a logical AND gate the edge can be detected. Edges of opposite contrast polarity are detected by a unit where the S^+ and S^- unit are interchanged. **B**, A T^+ unit with an on-center receptive field responding to the temporal change of its input. The time course of the responses to a bright-dark edge moving to the right and its temporal derivative are shown. Note that the same signals are obtained for a dark-bright edge moving to the left. **C**, Various combinations of STS-triplets detect the direction of motion of edges of both contrast polarities. In each detector the output signals of all three subunits are fed through a logical AND gate which leads only to an output when all three input signals exceed a certain threshold value, i.e. when the edge has the right polarity and is moving in the proper direction.

temporal and spatial brightness gradients (Wang et al., 1989).

The *Marr-Ullman model* (Marr and Ullman, 1981) (Fig. 8) which is also frequently classified as a variant of a gradient detector circumvents this problem by using an "AND" gate rather than a division to relate an estimate of the temporal and spatial brightness changes. Hence, in this respect it is similar to the Barlow-Levick model and, thus, to a correlation detector. The Marr-Ullman model is based on two processing steps. In the first step it determines the location of edges, i.e. the steepest local brightness changes, in the retinal image. By relating their polarity to the corresponding temporal brightness change, in a second processing step, the direction of pattern motion can be determined. The location and polarity of the edges are detected by a conjunction of a pair of adjacent units with antagonistic center-surround organization, one responding to positive brightness values ("on-center", S^+), the other to negative brightness values ("off-center", S^-) (Fig. 8A) (Marr and Hildreth, 1980; Marr and Ullman, 1981). Combining such edge detectors of either polarity with another unit which responds to either temporal brightness increments or decrements (T^+- and T^--units, Fig. 8B) motion detectors can be constructed which signal the direction of moving edges (Fig. 8C).

2.3. *Physiological implementation*

The motion detection models described so far belong to the group of so-called *algorithmic models*: They describe the different processing steps by formal operations such as convolutions, subtractions, multiplications etc. without referring to the specific properties of the biological substrate by which these computations are realized in the nervous system. In contrast, *physiological models* start off from the physiological properties of neurons, synapses etc. and try to build from these a neuronal machinery for motion vision. Of course, physiological models still are models and, for that reason, have a formal character: When describing the action, for instance, of a synapse they make use of electrical circuits as an equivalent for the neuronal membrane. Hence, "physiological models" are formal models the same way as algorithmic models are: They are settled just one step closer to the biological substrate.

Many components of the algorithmic motion detection models can be implemented quite easily in neuronal terms. Some of them have even been inspired by the properties of different types of visual interneurons. This is true, most notably, for the different types of spatial and temporal filters that were proposed in the input channels of various versions of correlation detectors or the Marr-Ullman model. For instance, spatial band-pass filters are well approximated by the antagonistic center-surround organization of the receptive fields of ganglion cells in the vertebrate retina (e.g. Rodieck and Stone, 1965; Enroth-Cugel and Robson, 1966). It is more demanding to account for the nonlinear interaction between the input signals of the movement detector in neuronal terms. Mainly two types of proposals have been made of how this processing step might be realized in biological movement detectors.

The *shunting inhibition model* (Fig. 9) has been considered as a physiological means to account for the AND-NOT gate of the Barlow-Levick model (Thorson, 1965; Torre and Poggio, 1978). In this model signals of the two input lines of the movement detector are thought to alter in a postsynaptic neuron the conductances of different ions, an excitatory and an inhibitory one (Fig. 9A). During motion in the preferred direction signals are transmitted by the excitatory channel (Fig. 9B), whereas during motion in the null direction the action of the excitatory synapse is shunted by the simultaneous activation of the inhibitory synapse (Fig. 9C). A multiplication, the nonlinear interaction of the Reichardt model, is approximated by the mechanism of shunting inhibition, if two conditions are met: (i) the reversal potential of the inhibitory ion is close or equal to the resting potential of the postsynaptic cell; (ii) the input signals are small compared with

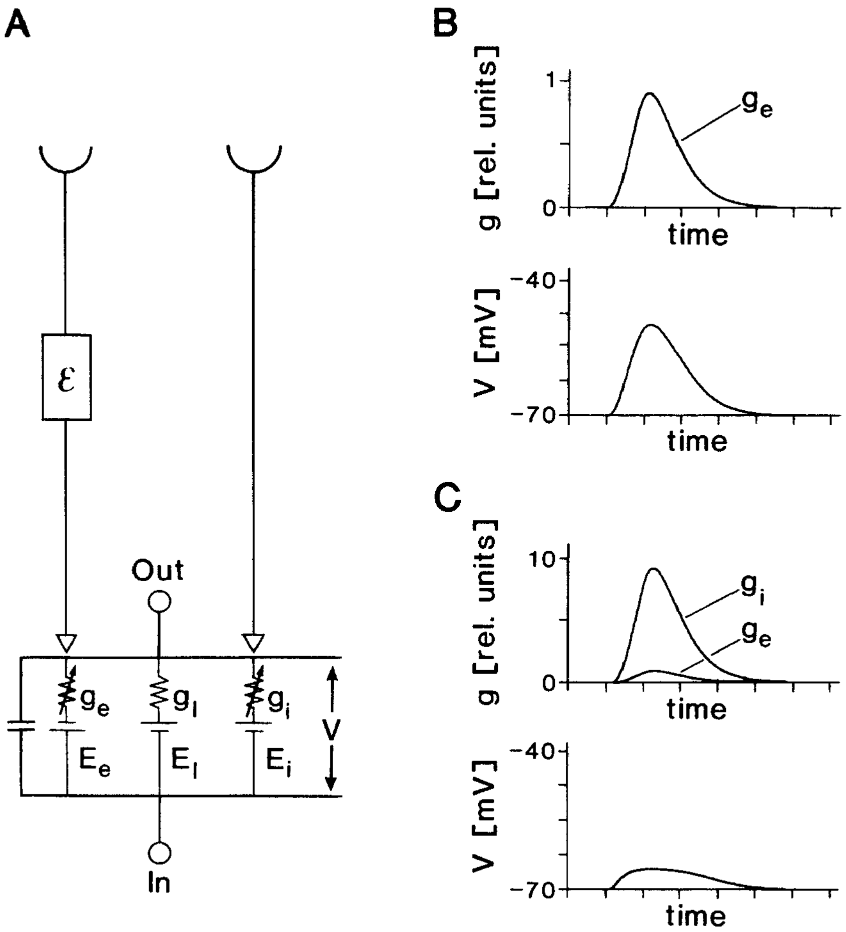

Fig. 9. Shunting inhibition model. A, The input signals derived from adjacent image points are assumed to change the conductances of different ion channels of a postsynaptic cell after one of the input signals is delayed with respect to the other. **B**, If the conductance of the inhibitory ion is low, activation of the excitatory channel leads to a pronounced EPSP. C, If simultaneously to the excitatory signal an inhibitory signal arrives, as is the case for motion in the "null direction", the EPSP is shunted. For motion in the "preferred direction" the depolarization caused by the excitatory input line is unaffected since the inhibitory signal arrives too late. The postsynaptic potential can be quantitatively determined according to the electrical equivalent circuit shown in A.

$$V = (E_l g_l + E_e g_e + E_i g_i)/(g_l + g_e + g_i)$$

E_l denotes the leak potential and E_e and E_i the reversal potential of the excitatory and inhibitory ion, respectively, g_l denotes the lumped leak conductances and g_e and g_i the conductances of the excitatory and inhibitory ion, respectively. As can be seen by a Taylor expansion of this expression the resulting postsynaptic potential approximates a multiplication of the input signals g_e and g_i if these are small compared with the leak conductance g_l and if the reversal potential of the inhibitory ion E_i is close to the leak potential E_l (Torre and Poggio, 1978).

the leak conductance of the postsynaptic cell. If these conditions are not satisfied, considerable deviations from a quadratic nonlinearity may appear (Torre and Poggio, 1978; Grzywacz and Koch, 1987).

The *threshold model* is another cellular possibility to account for the nonlinear interaction between the detector input signals (Fig. 10) (Srinivasan and Bernard, 1976; Grzywacz and Koch, 1987). A threshold operation is performed by any neuron which transmits information along its axon by means of action potentials. The threshold

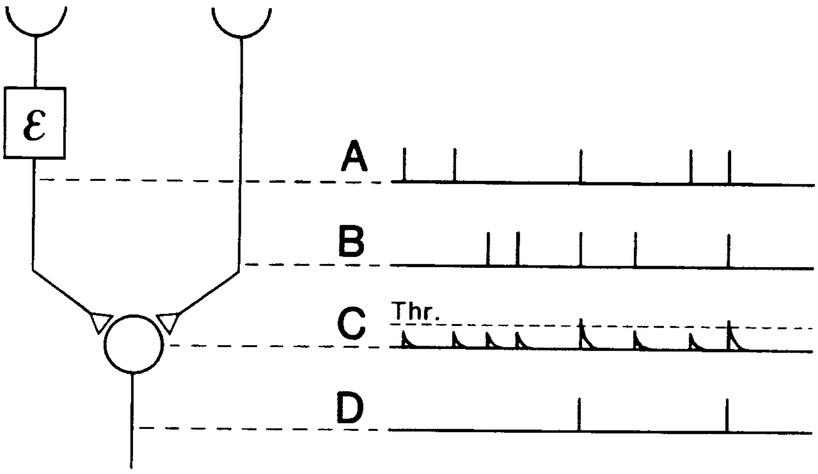

Fig. 10. Threshold model. The action potentials of two cells (A and B) induce excitatory potentials (C) in a postsynaptic cell. When arriving alone the EPSPs are too small to trigger an action potential. Only when two spikes arrive simultaneously the EPSPs are superimposed, exceed the threshold (Thr.: threshold as indicated by the dashed line in C) and give rise to an action potential (D). Note that the width of the temporal window within which the superposition of the two EPSPs causes an action potential is set by the time course of the EPSPs.

model proposes that both input lines of the detector induce excitatory postsynaptic potentials in a common follower neuron which may superimpose in a linear way. If the resulting membrane potential exceeds a given threshold, a spike will be triggered. Since the probability of coincidence of two events is the product of the probabilities of each single event alone, the output of this type of threshold model is equivalent, on average, to a multiplication (Srinivasan and Bernard, 1976), the core of the Reichardt model.

Both of these cellular models rely on the most common properties of neurons such as on changes in their input resistance or on spike thresholds. During recent years a number of more complex nonlinear synaptic interactions have been characterized electrophysiologically in various systems which may potentially qualify for the nonlinear interaction in biological movement detectors. The ionic channels subserving the NMDA receptor (e.g. Ascher and Nowak, 1987) may be such an example. However, there is no positive evidence so far that these cellular mechanisms play a role in motion detection. Moreover, how well these mechanisms are approximated by the different formal operations of the algorithmic movement detection models needs to be analyzed theoretically.

2.4. Common properties

Despite the apparent differences between the movement detector models discussed in Section 2.1–2.3, all of them satisfy three requirements which any motion detection mechanism has to meet in order to signal motion in a directionally selective way (Poggio and Reichardt, 1973; Buchner, 1984; Borst and Egelhaaf, 1989). (i) They have at least two input channels. A single input channel could not distinguish, for instance, a dark bar crossing its receptive field in one direction from a dark bar moving in the opposite direction, or from a transient dimming of the light. For the correlation method in its basic form two spatially separate measurements form the signals which are correlated; for the gradient method at least two spatially separate measurements have to be used to approximate the spatial gradient. (ii) The processing of the movement detector input signals has to be asymmetrical in some way. If it were symmetrical, the detector input channels could be interchanged without affecting the output. In this case, it would no longer be possible to tell which channel was excited first and which later, and the detector would no longer be directionally selective. In the different versions of correlation detectors the asymmetry is due to the fact that the detector input signals which are correlated are temporally filtered in a different way. In gradient detectors the asymmetry is inherent in the determination of the spatial brightness gradient. (iii) To compute motion, the input channels of a detector have to be related to each other in a nonlinear way. Otherwise, the mean output of a detector would be equal to the detector's response to its averaged input signals; in the averaged input signals, however, all information about the temporal sequence is lost. Thus, the result of a linear interaction cannot be directionally selective. The logical gates used by the Barlow-Levick and Marr-Ullman model, as well as the different analog operations such as the multiplication in the Reichardt model, the squaring operation in the energy model and the division used by the gradient detector represent such nonlinear interactions.

All the models of biological motion detection discussed here operate on continuous time-dependent signals of discrete neighboring input channels. In this respect, they differ from many motion detection algorithms used in machine vision. These algorithms derive motion information from a sequence of temporally discrete consecutive image frames. In a particular class of algorithms, certain features, so-called "tokens" are identified in subsequent image frames in a first processing step. In the next processing step, the displacement of these tokens is determined. From this and the interframe time interval the velocity of the token can be calculated. As one might expect, the biggest computational problems introduced by this method are the identification of the object or token in a frame and the discovery of which token in one frame corresponds to which in the next frame. This problem is the well-known *correspondence problem.* Since the token has to be identified before its displacements can be detected, it has to be known in advance which objects are the relevant ones.

3. Performance of motion detectors

How well are the various motion detectors suited to signal visual motion? Detector models which use a logical gate as their essential nonlinear interaction can only respond with binary signals; therefore, they are only able to signal in an all-or-none fashion motion of a particular velocity (Barlow-Levick model) or to distinguish motion in one direction from non-motion or motion in the opposite direction (Marr-Ullman model). In contrast, motion detectors with analog interactions (multiplication, division, addition followed by a squaring or threshold operation) can signal motion in a graded way. Therefore, only these models will be further considered and the dependence of their responses on the various stimulus parameters will be examined.

3.1. Velocity dependence

An ideal motion sensor should monitor the velocity, for instance, like the speedometer of a car: Within its operating range, the higher the velocity the higher should be the output signal. Yet, most of the models show a different behavior. Consider, for instance, the response of a single subunit of the Reichardt model to an object moving in its preferred direction. It responds optimally to a certain velocity, i.e. when the maxima of the two input signals coincide at the multiplication stage; this optimal velocity thus depends on the movement detector delay. If the object is moving at a lower or higher velocity the result of the multiplication will be smaller. Consequently, such a detector shows a peak in its velocity dependence (Reichardt and Varjú, 1959) which makes the interpretation of a given output signal ambiguous with respect to the velocity of the object.

Depending on their realization, gradient detectors may also show a velocity dependence which differs from an ideal velocity sensor. For instance, when the spatial and temporal brightness changes are approximated by the difference of the pattern brightness at neighboring points of the retinal image and by the temporally high-pass filtered input signal, respectively (see Fig. 7), the detector response shows an optimum at a particular velocity and then decreases again. Hence, at least this simple gradient detector exhibits a velocity dependence similar to correlation-type movement detectors and does not signal the correct pattern velocity.

The response optimum of correlation detectors is not located at a constant velocity. It rather depends on the spatial properties of the pattern in such a way that the ratio of the optimum velocity and the spatial wavelength of the pattern, i.e. the temporal frequency, is constant (Reichardt, 1961; Götz, 1964; Buchner, 1984; Reichardt, 1987). This relation only holds for steady-state conditions and fully opponent correlation-type movement detectors. Otherwise deviations from the linear relationship between the optimum velocity and the pattern wavelength may appear.

3.2. Pattern dependence

The response of several motion detector models is not exclusively determined by the velocity but depends also on the contrast and the spatial frequency content of the stimulus pattern. The reason for this may be most intuitive for the Reichardt detector. The output of the multiplicative interaction between the detector input channels not only depends on how well the signals coincide temporally at the multiplication stage, but also on the shape and size of the signals themselves. Since higher contrasts lead to higher amplitudes of the detector input signals, the outcome of the multiplication will be the larger the higher the contrast of the moving stimulus. Thus, the detector "mixes" up velocity with contrast. This is also true for the energy model. An ideal gradient detector does not show this contrast dependence: Although a pattern with higher contrast leads to a larger spatial gradient, the temporal brightness change becomes larger accordingly. Thus, when dividing spatial and temporal brightness changes, the contrast factors out. However, this may be no longer true if the gradient scheme is realized by technical or biological hardware. Due to the fact that both the spatial gradient and the temporal brightness change have to be approximated by certain filters and care has to be taken to prevent the spatial gradient from being zero (the denominator of the division in Eqn. 2), the output signal even of gradient detectors may also become, at least to some extent, contrast sensitive.

The response amplitude of a movement detector is also affected by the spatial frequency content of the stimulus pattern. This is a consequence of the geometry of the movement detector and the spatial filters in its input channels. The angular distance between the two input channels is often referred to as the "sampling base" and determines the spatial resolution of the motion detection system and its dependence on the spatial frequency components of the stimulus pattern. According to Shannon's sampling theorem (Shannon and Weaver, 1949), the smallest spatial wavelength that is resolved adequately amounts to twice the sampling base. For smaller wavelengths the response may become inverted, signalling the wrong direction of motion. This phenomenon is known as "geometrical interference" or "spatial aliasing" (Varjú, 1959; Götz, 1964; van Santen and Sperling, 1984). Maximum responses are expected for four times the sampling base. Towards higher spatial wavelengths the response amplitude decreases again. This means that, owing to its finite sampling base, a motion detection system has an intrinsic spatial band-pass characteristic, even if there are no additional spatial filters in its input channels (Götz, 1964, 1965, 1972). This fact is sometimes neglected when the spatial frequency dependence of motion detection systems is interpreted (Burr et al., 1986). This intrinsic spatial band-pass characteristic of movement detectors may be modified by spatial filters in their input channels. Depending on the spatial frequency of the pattern, the effect of these filters on the movement detector input is the same as reducing the pattern contrast. Accordingly, the detector output depends on the spatial frequency of the pattern (as it depends on pattern contrast). An important consequence of filtering out the high spatial frequency components in the movement detector input signals is that spatial aliasing can be prevented (For a more detailed discussion, see Borst and Egelhaaf, 1989).

Although the output of the ideal gradient detector does not depend on the spatial frequency of the stimulus pattern (for the same reason that it does not depend on pattern contrast), its discrete implementations (such as shown in Fig. 7) may depend to some extent on the structure of the stimulus pattern, too.

3.3. Time course of the response

Imagine a pattern with sinusoidal brightness distribution moving at a constant velocity across the receptive field, for instance, of a Reichardt detector. Let us further assume that the spatial wavelength of the grating is large compared to the sampling base of the detector. This implies that during some time of the stimulation cycle the two detector input signals are large (though different), then they decrease, reach a minimum and finally increase again. Since these signals are multiplied, the output signal of a Reichardt detector is expected to be modulated over time (Grzywacz and Koch, 1987; Egelhaaf et al., 1989b). In the case of a multiplicative interaction between the movement detector input channels (or any other type of quadratic nonlinearity) the response is modulated with the temporal frequency of the stimulus pattern and its second harmonic (for a detailed discussion of this feature, see Egelhaaf and Borst, this volume). Temporal response modulations will also arise if the nonlinear interaction cannot be approximated by a multiplication. Depending in a characteristic way on the nature of the nonlinearity, the time-dependent detector response will then also contain higher-order frequency components.

In principle, the temporal response modulations can be prevented by filtering the movement detector input in a specific way. This is true for both Reichardt and energy models if the fully opponent detectors are concerned. Then the modulations disappear if the detector input signals are filtered, either spatially or temporally, in such a way that the background brightness is no longer represented at the movement detector input. However, as soon as there are any time-independent components in the detector input signals or the detectors are no longer perfectly mirror-symmetrical, the response modulations will persist. In the case of the energy model, the temporal modulations can already be eliminated at the level of a detector subunit. However, this imposes even more severe constraints on the filters in the detector input channels. If the signals are filtered in such a way that the filter responses to sinewave gratings moving at a constant velocity differ in their phase by exactly one quarter of a temporal cycle (i.e. they are in "quadrature phase", see Section 2.1), the temporal response modulations disappear (Adelson and Bergen, 1985; van Santen and Sperling, 1985). However, as soon as the phase shift deviates from this value, response modulations in the detector output appear.

A gradient detector, when realized in a mathematically ideal form, i.e. when the temporal and spatial brightness changes are always strictly proportional (see Eqn. 2), does not show modulated responses to pattern motion with a constant velocity; instead, the time course of its response immediately reflects the time course of pattern velocity. However, hardware implemented gradient detectors such as shown in Fig. 7 may deviate from this ideal performance. In particular, when the spatial brightness gradient becomes small, pronounced changes in the time course of the response may occur.

4. Further processing of motion detector responses

Unless implemented in a very specific way, most detector models discussed here do not encode faithfully the velocity of the stimulus pattern. We have seen that this is reflected in two response characteristics: (i) The output signal of a detector may be modulated over time, even if the pattern moves with a constant velocity. (ii) The mean response amplitude may depend on the textural properties of the pattern. Different ways of coping with these limitations will be discussed in the following.

4.1. Spatial integration

One way to get rid of the temporal response modulations is to integrate spatially over an array of movement detectors looking at neighboring points in the visual field. Ideally, this array has to cover an integer multiple of spatial periods of a periodic stimulus pattern (Fig. 11A). Since then all local detectors look at different spatial phases of the pattern, their output signals are phase-shifted, too. Spatially integrating these output signals leads to a smooth time course of the overall response which is, within certain limits (see below), proportional to the time course of pattern velocity. However, this useful consequence of spatial integration has to be paid for by a reduced spatial resolution of the motion detection system. It is obvious that the minimum range of spatial integration which is necessary for the response modulations to disappear depends on the spatial frequency content of the stimulus pattern which has access to the motion detection system. This implies that if the retinal image is spatially band-pass filtered peripheral to the site of movement detection, the minimum range of spatial integration depends on the low frequency cut-off of the spatial filters. Of course, the temporal modulations in the response of individual movement detectors are also eliminated when the signals are integrated over time instead of space. This suggests a trade-off between spatial and temporal resolution of a movement detection system. However, spatial or temporal integration only eliminates the temporal modulations of the detector signal. The mean response amplitude still depends on the structure of the stimulus pattern.

The time course of the spatially integrated movement detector response is proportional to pattern velocity only within a certain dynamic range. If the pattern velocity changes too rapidly, for instance at the onset of motion, characteristic response transients may occur. This has been studied in some detail for the Reichardt model (see Egelhaaf and Borst, this volume; Borst and Bahde, 1986; Egelhaaf and Reichardt, 1987; Egelhaaf and Borst, 1989, 1990). However, the response transients are neither a special feature of this model, nor are they due to adaptational changes of some parameters of the motion detection system. They rather reflect the consequence of response transients of the temporal filters which are a constituent part of any motion detection system.

4.2. Recovery of pattern velocity in one spatial dimension

As discussed above with correlation-type motion detectors, the spatial and temporal frequencies of the stimulus pattern are confounded to some extent. This problem has an analogue in color vision where the output of a photoreceptor depends on both the spectral wavelength and the light intensity. In color vision this ambiguity can be circumvented by combining the output of photoreceptors with different spectral sensitivities. Hence, the ambiguities are resolved by a sufficient number of independent measurements. There are two ways to accomplish this in motion vision: (i) By combining the output of different types of movement detectors and (ii) by determining particular features of the pattern structure and by combining these measurements with the movement detector output.

The first of these methods is similar to the strategy used in color vision. Here motion is detected in parallel by a population of motion detectors with different spatial and temporal frequency optima (Fig. 11B). By appropriately wiring up these detectors, units can be constructed which respond best to a particular pattern velocity. This basic idea has been proposed with slight modifications in a number of publications (e.g. Heeger, 1987; Grzywacz and Yuille, 1990; Glünder, 1990). It is related to approaches which have

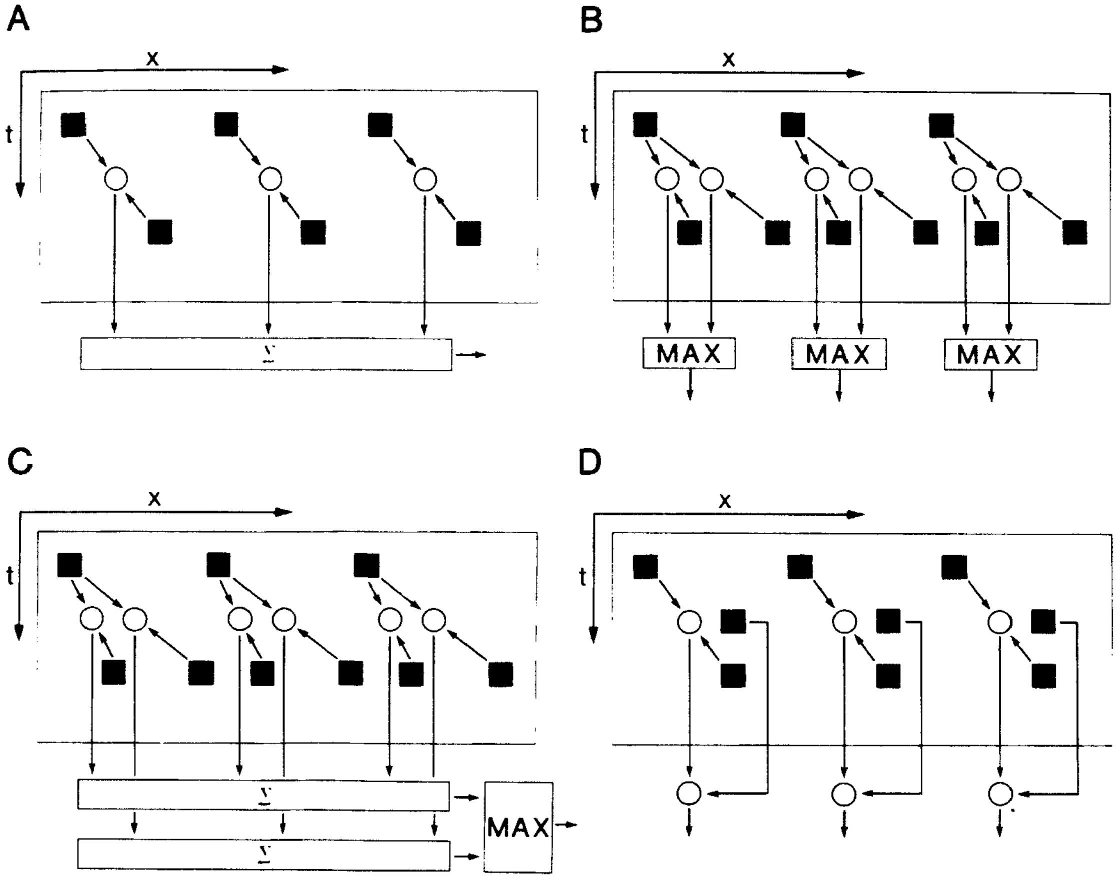

Fig. 11. Various ways to process the output signals of local correlation detectors. Each correlation detector is symbolized in the space-time domain as in Fig. 3A with two black squares indicating points of measurement and an open circle indicating the nonlinear interaction. The image has to be thought to be covered by a dense array of detectors. Only three detectors are shown in the scheme. **A**, The output signals of spatially distributed local motion detectors of a given type are pooled across space. If this is done over an area covering integer multiples of one period of the largest spatial wavelength of the pattern, the temporal modulations of the output signals of single detectors disappear. The advantage of pooling has to be paid for by a concomitant loss of spatial resolution of the motion detection system. **B**, If populations of detectors with different preferred velocities, i.e. different orientations in the space-time domain, are used, the ambiguities in the output signal of a single detector with respect to velocity and other pattern parameters can be reduced. The local velocity may then be determined by a simple winner-takes-all method ("MAX"). **C**, A combination of **A** and **B** where the output of detectors with the same filter characteristic are spatially pooled before their responses are compared and the maximum is determined. As in **A** spatial resolution is lost by spatial pooling. **D**, The inherent ambiguity of the detector response can also be resolved when the relevant pattern parameters are determined by independent measurements (symbolized by an additional black square). From an appropriate combination of the detector output signal with the local pattern parameters (e.g. the spatial derivatives of the local image brightness) the local image velocity can be calculated.

been put forward in the context of computer vision. However, in contrast to biological vision, motion information is extracted here from a sequence of frames taken in discrete time steps. In one algorithm patches of the image in a given frame are shifted with respect to the preceding frame in all directions over various distances, and for each shift the cross-correlation of the intensity values in the corresponding patches is calculated. The shift at which a maximum correlation is obtained is taken as the local displacement and the resulting vector is interpreted as the corresponding image velocity (Fig. 11B) (Bülthoff et al., 1989). If the whole image is shifted and cross-correlated in this way, the method is called a "global correlation algorithm" and has been successfully applied, for instance, to automatically track moving clouds (Fig. 11C) (Smith and Phillips, 1972).

The other strategy to recover pattern velocity is based on measuring the relevant properties of the stimulus pattern at the same location of the retinal image where the velocity has to be determined (Fig. 11D). This pattern information can then be used to derive the pattern velocity from the ambiguous motion detector response. For the Reichardt model this possibility has been worked out in formal terms. When the temporal filter is approximated by a constant delay ε, the output signal of a detector $r(x,t)$ is related, at a first approximation, to the pattern speed along the x-axis v_x through the first and second spatial derivatives ($\partial F/\partial x$ and $\partial^2 F/\partial x^2$, respectively) of the local brightness distribution F (Reichardt, 1987):

$$r(x,t) = -\varepsilon \cdot [(\partial F/\partial x)^2 - F \cdot \partial^2 F/\partial x^2] \cdot v_x \qquad (3)$$

If the spatial derivatives are known from separate measurements, the equation can be solved and the pattern speed can be recovered from the detector output signal. This method will be dealt with in more detail below when the problem of motion detection in two spatial dimensions is treated.

5. Motion in two dimensions

So far, the computation of motion information has been considered only for one spatial dimension. However, for the visual system any motion in three-dimensional space is reflected in the time-dependent brightness changes of the projected two-dimensional retinal image. Therefore, we now extend our analysis to motion in two spatial dimensions, thereby considering two major questions. (i) How is the two-dimensional motion information represented at the output of arrays of motion detectors such as were discussed in the previous sections? (ii) To what extent is it possible to recover from these output signals the correct two-dimensional retinal velocity field as given by the geometrical projection of the three-dimensional motion vectors onto the retina?

5.1. Output of a pair of motion detectors

It may be intuitively clear that there is no way for a single movement detector to represent the two components of a two-dimensional retinal velocity vector. At best, a single motion measurement can encode one of its components. Hence, two independent measurements are required. These can be combined in a response vector which may then be regarded as a two-dimensional representation of the retinal velocity vector.

In general, the response vector determined in this way does not coincide with the corresponding velocity vector. The following considerations will show us why. Assume a horizontally oriented movement detector. If a stimulus pattern, say some sort of curved obliquely oriented edge, moves horizontally, brightness changes are induced along the detector axis; these lead to responses of the detector (Fig. 12A). Motion of the same pattern orthogonally to the detector axis also leads to time-dependent brightness changes along the detector axis and thus to responses (Fig. 12B). This means that the response of a movement detector to translatory motion in an arbitrary direction of the image plane, apart from special

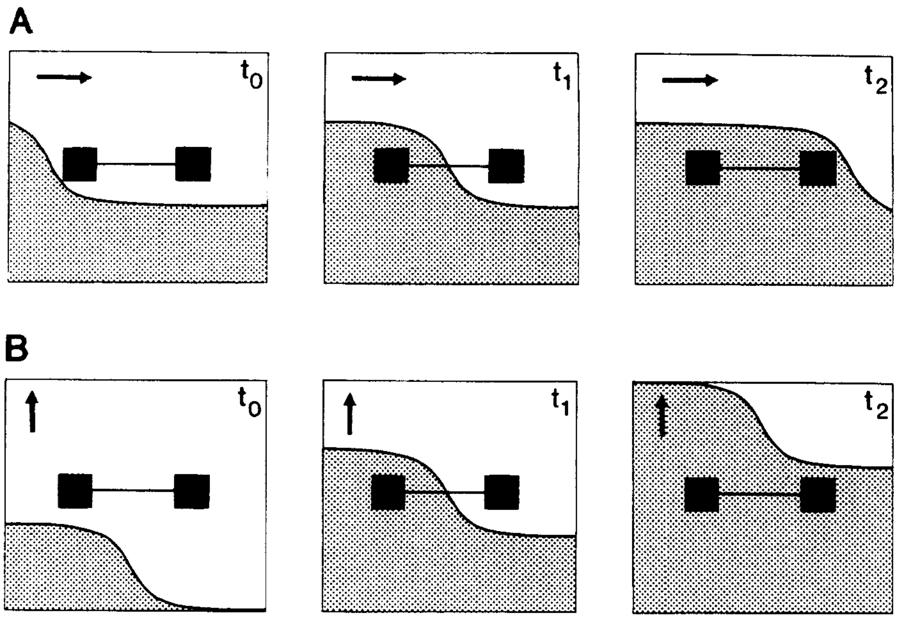

Fig. 12. A motion detector oriented along the horizontal axis of the image is shown to respond to pattern motion in both the x-direction and the y-direction. **A**, A pattern moving horizontally is shown at three instants of time ($t_0 - t_2$). Both input channels of the motion detector are activated in sequence. **B**, The same sequence of activation is obtained if the pattern is moving vertically. Thus, a motion detector will confound in its output signal, depending on the structure of the pattern, both velocity components v_x and v_y.

stimulus patterns, does not depend exclusively on the velocity component aligned with its axis but also on the local structure of the pattern as well as on the velocity component orthogonal to the

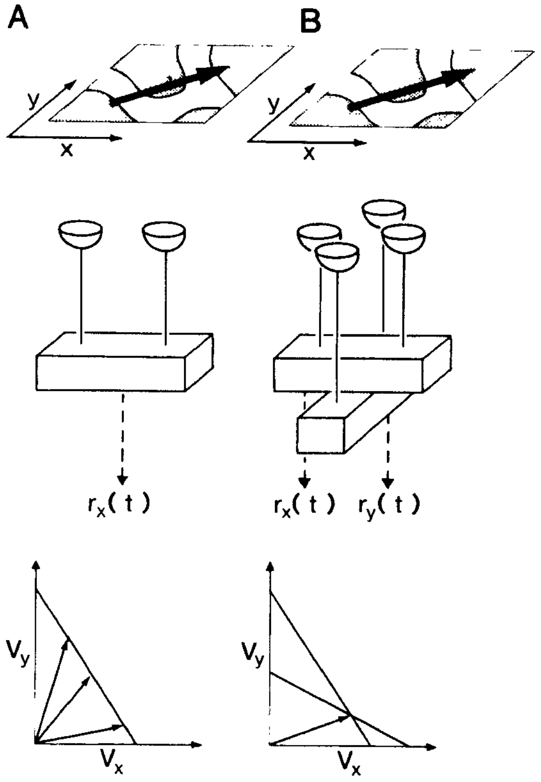

Fig. 13. **A**, When the image is moving in a given direction (top), its two-dimensional motion vector cannot be calculated on the basis of the output signal of a single motion detector. The output of a single movement detector depends on both the x- and the y-component of the image velocity in the general form of $r_x(t) = a \cdot v_x + b \cdot v_y$ with a and b being functions of the local brightness distribution. This equation can be pictured as the so-called *line of constraint* (bottom) showing that an infinite number of vectors v (v_x,v_y) satisfy this equation. **B**, However, if motion is detected in parallel by a pair of differently oriented motion detectors, an additional equation can be obtained: $r_y(t) = c \cdot v_x + d \cdot v_y$. If these two equations are linearly independent and if the pattern-dependent terms a,b,c and d are known from independent measurements, the equations can be solved and a unique motion vector can be calculated locally. The solution can be pictured as the intersection point of both lines of constraint obtained from the output signals of both detectors. Note that the crucial requirement here is that both measurements are linearly independent.

detector axis. This may be expressed by the following equation:

$$r_x(t) = a \cdot v_x + b \cdot v_y, \tag{4}$$

with v_x and v_y corresponding to the horizontal and vertical velocity components; a and b are functions of the local pattern structure. In accordance with the intuitive notion that a single motion measurement is not sufficient for a two-dimensional representation of the retinal velocity vectors, this equation with two unknown variables, v_x and v_y, is compatible with an infinite number of velocity vectors and cannot be solved. Geometrically, the tips of these vectors coincide with a straight line in a v_x, v_y diagram, the so-called *line of constraint* (Horn and Schunck, 1981; Adelson and Movshon, 1982) (Fig. 13A). The slope of the constraint line depends on both the motion detection mechanism under consideration as well as on the local structure of the stimulus pattern.

A second motion measurement provided by a vertically oriented detector leads to an equivalent equation

$$r_y(t) = c \cdot v_x + d \cdot v_y \tag{5}$$

with c and d again representing functions of the structure of the stimulus pattern. Since r_x and r_y, in general, depend on both the components of the retinal velocity vector and the structure of the stimulus pattern, the response vector which is obtained by combining r_x and r_y is expected not to encode the retinal velocity vector correctly in terms of direction and magnitude.

For a pair of orthogonally oriented, mathematically ideal gradient detectors the responses are given, as in the one-dimensional case (see Eqn. 2), by the relation between the temporal change of the pattern brightness at a given location and the corresponding spatial change either along the x- or y-axis

$$r_x(t) = -\,(\partial F/\partial t)\;/\;(\partial F/\partial x) \tag{6}$$

$$r_y(t) = -\,(\partial F/\partial t)\;/\;(\partial F/\partial y) \tag{7}$$

Equation 1 can now be extended for the two-dimensional case to the so-called *optical flow constraint equation* (Horn and Schunck, 1981).

$$-\,\partial F/\partial t = \partial F/\partial x \cdot v_x + \partial F/\partial y \cdot v_y \tag{8}$$

which describes the relation between the pattern velocity, spatial gradient and temporal brightness change. Using Eqns. 6–8, the pattern dependent coefficients in Eqns. 4 and 5 can be specified leading to the following pair of equations for the responses of a pair of gradient detectors

$$r_x(t) = v_x + (\partial F/\partial y)/(\partial F/\partial x) \cdot v_y \tag{9}$$

$$r_y(t) = (\partial F/\partial x)/(\partial F/\partial y) \cdot v_x + v_y \tag{10}$$

The resulting response vector, with r_x and r_y representing its x- and y-component, respectively, depends on the brightness change of the retinal image along both the x- and y-axis. Hence, in general, it does not coincide with the corresponding retinal velocity vector.

In contrast to the gradient scheme the response vectors obtained from a pair of Reichardt detectors depend not only on the first spatial derivatives of the pattern brightness but also on the second spatial derivatives (Reichardt, 1987; Reichardt et al., 1988; Reichardt and Schlögl, 1988):

$$\begin{aligned} r_x(t) = -\varepsilon \cdot \big[&((\partial F/\partial x)^2 - F \cdot \partial^2 F/\partial x^2) \cdot v_x + \\ &(\partial F/\partial x \cdot \partial F/\partial y - F \cdot \partial^2 F/\partial x \partial y) \cdot v_y\big] \end{aligned} \tag{11}$$

$$\begin{aligned} r_y(t) = -\varepsilon \cdot \big[&(\partial F/\partial y \cdot \partial F/\partial x - F \cdot \partial^2 F/\partial y \partial x) \cdot v_x + \\ &((\partial F/\partial y)^2 - F \cdot \partial^2 F/\partial y^2) \cdot v_y\big] \end{aligned} \tag{12}$$

The pattern dependence of the response vectors of Reichardt detectors is illustrated by the computer simulation of Fig. 14. Although the pattern moves with a constant velocity along the x-axis of the detector array, the response vectors shown here at a particular instant of time point into different directions depending on the local structure of the stimulus pattern. Thus, the local instantaneous response vectors may not coincide with the direction of pattern motion.

From all this we can conclude that local motion measurements by the various types of movement detectors discussed here do not directly yield reliable information on the speed and direction in which the different pattern segments of the two-dimensional retinal image are moving.

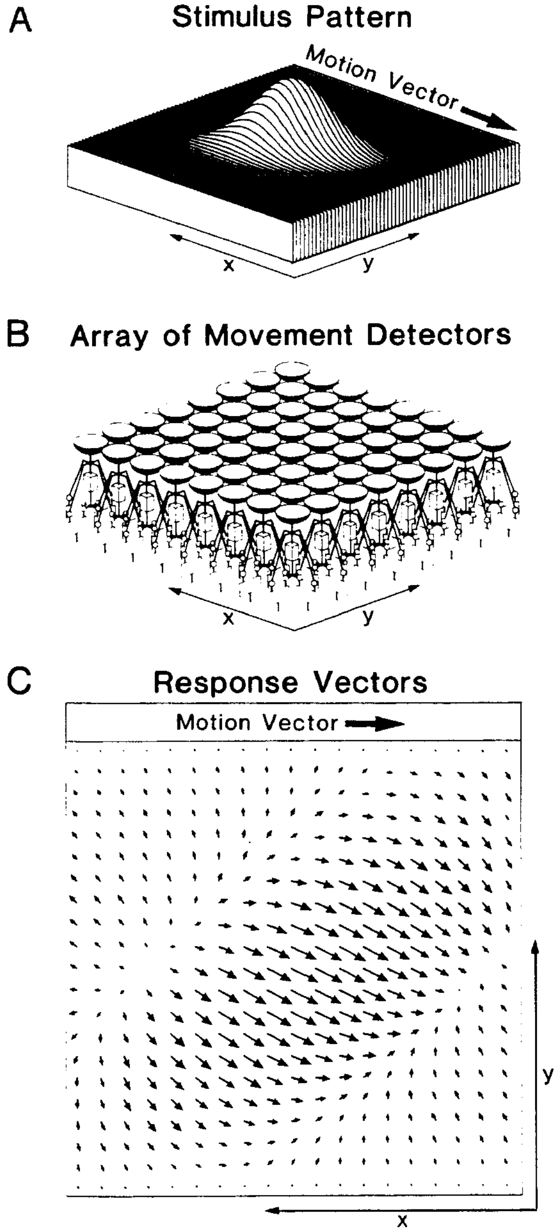

Fig. 14. Representation of the two-dimensional image velocity at the output of an array of correlation detectors of the Reichardt type. **A**, An elongated Gaussian-shaped brightness distribution moving along the x-axis with constant velocity represents the input pattern. **B**, This stimulus is fed through a two-dimensional array of pairs of orthogonally oriented Reichardt detectors. **C**, If the output signal of each member of such a pair is interpreted as the velocity component along the x and y-axis, respectively, a response vector field is obtained at each instant of time. An instantaneous snapshot of this vector field is shown here. Depending on the local structure of the pattern, the local response vectors all point into different directions, although the entire pattern moves along the x-axis. Consequently, most of the response vectors deviate significantly from the input vector. (Model simulation courtesy W. Reichardt.)

5.2. Methods for recovering the 2D velocity field

Is it possible to compute the correct retinal velocity field from the response profiles of an array of movement detectors? In analogy to motion in one spatial dimension (Section 4.2), there are various ways in which this can be done. It is beyond the scope of this article to review this field extensively (see also Heeger, 1987; Ogata and Sato, 1989; Grzywacz and Yuille, 1990). Instead, we just concentrate on one important aspect.

The correct retinal velocity field can be recovered, at least for translatory motion in the image plane, on the basis of local mechanisms alone, without global constraints (Reichardt et al., 1988; Uras et al., 1988). If the pattern-dependent coefficients of Eqns. 4 and 5 are known from other measurements, we are confronted with two equations of the two unknown variables v_x and v_y. If these equations can be solved for v_x and v_y, the retinal velocity vector can be recovered. In geometrical terms, this means that the two corresponding lines of constraint are not colinear but intersect. The coordinates of the intersection point then represent the x- and y-components of the retinal velocity vector (Fig. 13B).

However, such an intersection point and, thus, an unambiguous velocity vector does not exist under all circumstances. Two conditions have to be met. (i) The brightness distribution of the pattern has to be truly two-dimensional. This condition refers to a property of the stimulus pattern: If the pattern brightness varies only along one axis, it is impossible to detect its direction and speed in two dimensions. Assume, for instance, a pattern consisting of parallel stripes moving in a given direction. The motion vector can be decomposed formally into two components, one being orthogonal and the other parallel to the stripes. The orthogonal component is accompanied by a brightness change in this direction, whereas the parallel component is not. Hence, the latter component cannot be detected by any mechanism whatsoever. Thus, the pattern brightness has to vary along two orthogonal directions. Only then can the two-dimensional velocity be recovered. (ii) The two motion measurements have to depend on the local pattern structure in such a way that the corresponding Eqns. 4 and 5 are, in general (i.e. apart from the types of patterns just mentioned), linearly independent. As will be shown below, this condition is not met by all motion detection mechanisms.

Movement detectors of the correlation type satisfy this requirement. Equations 11 and 12 which describe the output of a pair of orthogonally oriented detectors are in general linearly independent. Thus, the correct two-dimensional motion vector can, at least in principle, be recovered from them, if the pattern-dependent coefficients in the equations are determined by independent measurements (Reichardt et al., 1988). The movement detectors to be combined need not be oriented orthogonally. In principle, any orientations will do as long as they are different. Of course, it is not necessary to use a pair of equivalent detectors; two linearly independent equations may also be obtained if the detectors to be combined have temporal and/or two-dimensional spatial filters in their input channels which differ in an appropriate way.

In the case of a pair of gradient detectors the correct retinal velocity vector cannot be recovered by local mechanisms alone. This is because Eqns. 9 and 10 are not linearly independent and, therefore, do not have a unique solution. Hence, the correct retinal velocity vector cannot be derived from two independent measurements by differently oriented gradient detectors at a given location. However, it is possible to derive the retinal velocity vector, for instance, by combining measurements of gradient detectors at different parts of the retinal image (see below).

Different types of approaches are used in computer vision to estimate velocity fields. These algorithms are based on the formal description of the relation between the pattern velocity, the spatial brightness gradient and the temporal brightness change as given by the optical flow constraint equation (Eqn. 8) (Horn and Schunck, 1981). From the infinite number of vectors solving this equation a particular one, i.e. the vector directed along the brightness gradient, is selected deliberately. Some methods take into account global constraints, i.e. assumptions about properties of the velocity field, to obtain the information which is necessary in addition to Eqn. 8 for calculating unique velocity vectors. For instance, one might assume that the velocity vectors should vary smoothly across the image (Horn and Schunck, 1981) or along the contour line of an object (Hildreth, 1984). The rationale behind these smoothness assumptions is that nearby points tend to move with similar velocities. This sort of algorithm is not restricted to translatory motion of rigid objects but allows also for general motion of nonrigid objects (Hildreth, 1984). However, it does not compute the image velocity by a local mechanism but resorts to global constraints imposed by the physical nature of the two-dimensional retinal velocity field.

However, the optical flow constraint equation (Eqn. 8) can be the basis for algorithms which allow the retinal velocity to be computed locally. There are various ways to derive two linearly independent equations from it. One such possibility has been proposed in a recent study (Uras et al., 1988). By differentiating Eqn. 8 with respect to the spatial coordinates x and y, respectively, two linearly independent equations are obtained which contain second derivatives of the local image brightness (Uras et al., 1988):

$$-\partial^2 F/\partial t\partial x = \partial^2 F/\partial x^2 \cdot v_x + \partial^2 F/\partial x\partial y \cdot v_y \tag{13}$$

$$-\partial^2 F/\partial t\partial y = \partial^2 F/\partial y\partial x \cdot v_x + \partial^2 F/\partial y^2 \cdot v_y \tag{14}$$

If all spatial and temporal derivatives are known the equations can be solved for the two unknowns v_x and v_y and the local image velocity can be calculated. Recently, this approach has been generalized (Srinivasan, 1990). Instead of differentiating the optical flow constraint equation, it can alternatively give rise to two linearly independent equations if it is filtered by different linearly independent spatio-temporal filters.

All these considerations only show under what conditions the correct retinal velocity vectors can be recovered. They do not indicate how well the different algorithms perform, if they are implemented in technical or biological systems. However, it has been realized only recently that it is possible to compute the correct retinal velocity on a local basis (Reichardt et al., 1988; Uras et al., 1988). Before, it was assumed that it is impossible to measure locally velocity components other than that parallel to the brightness gradient, a problem often referred to as the *aperture problem* (e.g. Hildreth and Koch, 1987; Grzywacz and Yuille, 1990). However, as explained above there is no such aperture problem apart from the trivial fact that the correct image velocity cannot be detected if the stimulus pattern does not provide sufficient information, i.e. in the case of patterns with brightness variations along only one of its axes (Reichardt et al., 1988; Uras et al., 1981).

6. Conclusion: The use and usefulness of models

In general, models are of twofold advantage: (i) In a purely synthetic, engineer-like approach models help to realize the problems. (Think about the difficulties in machine vision of accomplishing even apparently simple tasks.) (ii) In conjunction with an analytical approach, models help to make explicit what has been assumed only implicitly before. (Everybody has a model in mind!) Models force the experimenter to formulate precisely the explanation of a particular experimental result. Because even simple mechanisms, such as some of the motion detection mechanisms discussed here, provide serious pitfalls to intuition it is often necessary to calculate the responses either by analytical methods or by use of a computer simulation.

Interestingly, only a small number of models of motion detection have been proposed so far in the literature. These different models can be grouped, according to the computational principle they are built on, into two classes, the so-called correlation- and gradient-detectors. Despite the apparent overall difference in the structure of the various model versions, some of them show a similar performance, for instance, with respect to their velocity tuning and their dependence on pattern structure. Thus, the dichotomy between, at least, some of the models starts to fade away, and this is all the more true when the "sloppy workmanship" of the neuronal hardware is taken into account. In order to decide between different model variants it becomes most important to derive predictions that are robust against these neuronal imperfections. In addition, a closer look at the internal structure of the models is necessary, because predictions may differ significantly depending on the processing level which is considered. For instance, since unambiguous velocity information can be recovered from various types of motion detectors by further processing steps, the finding of neurons exhibiting a strict velocity tuning (i.e. responses which depend exclusively on the velocity of the pattern irrespective of its structure) does not speak against or in favor of any primary mechanism of local motion detection. Here, the problem for the experimenter will always be to decide at which processing level the investigated neuron is located. Hence, it is often not as easy as one might expect at first glance to find out what mechanism underlies motion detection in a particular biological system. Therefore, modelling is an indispensable tool for designing more specific experiments which may eventually allow us to distinguish between different mechanisms. Thus, in research on motion vision models help to guide the experimenter to a better understanding of the system under investigation.

Acknowledgements

We are grateful to K.G. Götz, R. Hengstenberg, W. Reichardt and J.M. Zanker for helpful comments on the manuscript and to F. Buchstäber and B. Pilz for drawing the figures.

References

Adelson, E.H. and Bergen, J.R. (1985) Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A 2, 284–299.

Adelson, E.H. and Movshon, J.A. (1982) Phenomenal coherence of moving visual patterns. Nature 300, 523–525.

Ascher, P. and Nowak, L. (1987) Electrophysiological studies of NMDA receptors. Trends Neurosci. 10, 284–288.

Ballard, D.H. and Brown, C.M. (1982) Computer Vision. Prentice-Hall, Englewood, NJ.

Barlow, H.B. and Levick, W.R. (1965) The mechanism of directionally selective units in rabbit's retina. J. Physiol. 178, 477–504.

Borst, A. and Bahde, S. (1986) What kind of movement detector is triggering the landing response of the housefly?. Biol. Cybern. 55, 59–69.

Borst, A. and Egelhaaf, M. (1989) Principles of visual motion detection. Trends Neurosci. 12, 297–306.

Borst, A. and Egelhaaf, M. (1990) Direction selectivity of fly motion-sensitive neurons is computed in a two-stage process. Proc. Natl. Acad. Sci. USA, in press.

Buchner, E. (1984) Behavioural analysis of spatial vision in insects. In: M.A. Ali (Ed.), Photoreception and Vision in Invertebrates, Plenum Press, New York, London, pp. 561–621.

Bülthoff, H., Little, J. and Poggio, T. (1989) A parallel algorithm for real-time computation of optical flow. Nature 337, 549–553.

Burr, D.C., Ross, J. and Morrone, M.C. (1986) Seeing objects in motion. Proc. R. Soc. Lond. B 227, 249–265.

Cavanagh, P. and Mather, G. (1989) Motion: The long and short of it. Spatial Vision 4, 103–129.

Chubb, C. and Sperling, G. (1988) Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. J. Opt. Soc. Am. A 5, 1986–2006.

Doorn, A.J. van and Koenderink, J.J. (1982a) Temporal properties of the visual detectability of moving spatial white noise. Brain Res. 45, 179–188.

Doorn, A.J. van and Koenderink, J.J. (1982b) Spatial properties of the visual detectability of moving spatial white noise. Brain Res. 45, 189–195.

Egelhaaf, M. and Borst, A. (1989) Transient and steady-state response properties of movement detectors. J. Opt. Soc. Am. A 6, 116–127.

Egelhaaf, M., Borst, A. and Reichardt, W. (1989) The computational structure of a biological motion detection system. J. Opt. Soc. Am. A 6, 1070–1087.

Egelhaaf, M. and Borst, A. (1990) Transient and steady-state response properties of movement detectors: errata. J. Opt. Soc. Am. A 7, 172.

Egelhaaf, M. and Reichardt, W. (1987) Dynamic response properties of movement detectors: Theoretical analysis and electrophysiological investigation in the visual system of the fly. Biol. Cybern. 56, 69–87.

Enroth-Cugel, C. and Robson, J.D. (1966) The contrast sensitivity of retinal ganglion cells in the cat. J. Physiol. 187, 517–522.

Fennema, C.L. and Thompson, W.B. (1979) Velocity determination in scenes containing several moving objects. Comp. Graph. Im. Process. 9, 301–315.

Foster, D.H. (1969) The response of the human visual system to moving spatially-periodic patterns. Vision Res. 9, 577–590.

Foster, D.H. (1971) The response of the human visual system to moving spatially-periodic patterns: Further analysis. Vision Res. 11, 57–81.

Glünder, H. (1990) Correlative velocity estimation: visual motion analysis, independent of object form, in arrays of velocity-tuned bilocal detectors. J. Opt. Soc. Am. A 7, 255–263.

Götz, K.G. (1964) Optomotorische Untersuchungen des visuellen Systems einiger Augenmutanten der Fruchtfliege Drosophila. Kybernetik 2, 77–92.

Götz, K.G. (1965) Die optischen Übertragungseigenschaften der Komplexaugen von *Drosophila*. Kybernetik 2, 215–221.

Götz, K.G. (1972) Principles of optomotor reactions in insects. Bibl. Ophthal. 82, 251–259.

Grzywacz, N.M. and Koch, C. (1987) Functional properties of models for direction selectivity in the retina. Synapse 1, 417–434.

Grzywacz, N.M. and Yuille, A.L. (1990) A model for the estimate of local image velocity by cells in the visual cortex. Proc. R. Soc. Lond. B 239, 129–161.

Hassenstein, B. and Reichardt, W. (1956) Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers *Chlorophanus*. Z. Naturforsch. 11b, 513–524.

Heeger, D.J. (1987) Model for the extraction of image flow. J. Opt. Soc. Am. A 4, 1455–1471.

Hildreth, E. (1984) The computation of the velocity field. Proc. R. Soc. Lond. B 221, 189–220.

Hildreth, E.C. and Koch, C. (1987) The analysis of motion: From computational theory to neuronal mechanisms. Ann. Rev. Neurosci. 10, 477–533.

Horn, B.K.P. (1986) Robot Vision. MIT Press, Cambridge, MA.

Horn, B.K.P. and Schunck, B.G. (1981) Determining optical flow. Artificial Intelligence 17, 185–203.

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

Lelkens, A.M.M. and Koenderink, J.J. (1984) Illusory motion in visual displays. Vision Res. 24, 1083–1090.

Limb, J.O. and Murphy, J.A. (1975) Estimating the velocity of moving images in television signals. Comp. Graph. Im. Process. 4, 311–327.

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

Marr, D. and Ullman, S. (1981) Direction selectivity and its use in early visual processing. Proc. R. Soc. Lond. B 211, 151–180.

Ogata, M. and Sato, T. (1989) Motion perception model with interactions between spatial frequency channels. Proc. Int. Soc. Opt. Eng. 1077, 292–299.

Poggio, T. and Reichardt, W. (1973) Considerations on models of movement detection. Kybernetik 13, 223–227.

Reichardt, W. (1961) Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In: W.A. Rosenblith (Ed.), Sensory Communication, MIT Press and Wiley, New York, London, pp. 303–317.

Reichardt, W. (1987) Evaluation of optical motion information by movement detectors. J. Comp. Physiol. A 161, 533–547.

Reichardt, W. and Varjú, D. (1959) Übertragungseigenschaften im Auswertesystem für das Bewegungssehen. Z. Naturforsch. 14b, 674–689.

Reichardt, W., Egelhaaf, M. and Schlögl, R.W. (1988) Movement detectors provide sufficient information for local computation of 2-D velocity field. Naturw. 75, 313–315.

Reichardt, W.E. and Schlögl, R.W. (1988) A two dimensional field theory for motion computation. Biol. Cybern. 60, 23–35.

Rodieck, R.W. and Stone, J. (1965) Analysis of receptive fields of cat retinal ganglion cells. J. Neurophysiol. 28, 833–849.

Santen, J.P.H. van and Sperling, G. (1984) Temporal covariance model of human motion perception. J. Opt. Soc. Am. A 1, 451–473.

Santen, J.P.H. van and Sperling, G. (1985) Elaborated Reichardt detectors. J. Opt. Soc. Am. A 2, 300–320.

Shannon, C.E., and Weaver, W. (1949) The Mathematical Theory of Communication. University of Illinois Press, Chicago.

Smith, E.A. and Phillips, D.R. (1972) Automated cloud tracking using precisley aligned digital ATS pictures. IEEE Trans. Comput. C-21, 715–729.

Sperling, G. (1989) Three stages and two systems of visual processing. Spatial Vision 4, 183–207.

Srinivasan, M.V. (1990) Generalized gradient schemes for measurement of image motion. Biol. Cybern., 63, 421–431.

Srinivasan, M.V. and Bernard, G.D. (1976) A proposed mechanism for multiplication of neural signals. Biol. Cybern. 21, 227–236.

Thorson, J. (1965) Small-signal analysis of a visual reflex in the locust. II. Frequency dependence. Kybernetik 3, 53–66.

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

Ullman, S. (1983) The measurement of visual motion. Trends Neurosci. 6, 177–179.

Uras, S., Girosi, F., Verri, A. and Torre, V. (1988) A computational approach to motion perception. Biol. Cybern. 60, 79–87.

Varjú, D. (1959) Optomotorische Reaktionen auf die Bewegung periodischer Helligkeitsmuster. Z. Naturforsch. 14b, 724–735.

Wang, H.T., Mathur, B. and Koch, C. (1989) Computing optical flow in the primate visual system. Neural Computation 1, 92–103.

Watson, A.B. and Ahumada, A.J. (1985) Model of human visual-motion sensing. J. Opt. Soc. Am. A 2, 322–342.

Watson, A.B., Ahumada, A.J. and Farrell, J.E. (1986) Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. J. Opt. Soc. Am. A 3, 300–307.

Wilson, H.R. (1985) A model for direction selectivity in threshold motion perception. Biol. Cybern. 51, 213–222.

Zanker, J.M. (1990) Theta motion: A new psychophysical paradigm indicating two levels of motion detection. Naturw. 77, 243–246.