A Unified Account of Three Apparent Motion Illusions

A. JOHNSTON,* C. W. G. CLIFFORD*

Received 20 October 1993; in revised form 18 April 1994

We discuss three motion illusions, the fluted square wave illusion, the reverse phi illusion and the Pante illusion. In these illusions reversed apparent motion is either induced or eliminated by the introduction of a blank inter-frame-interval between the frames of the apparent motion sequence. In order to simulate these effects with the multi-channel gradient model we had to introduce low-pass spatial filters and second-order temporal differentiating filters. These illusions have been used as evidence of multiple motion mechanisms. Here we demonstrate that they can be considered as emergent properties of a single computational strategy.

Motion Reverse phi Missing fundamental illusion Computational model

1. PLENOPTIC ARRAY: COMPUTING ORIENTATION

De Yoe and Van Essen (1988) and Adelson and Bergen (1991) point out that we can think of the input to the visual system as a multi-dimensional function describing radiant energy in the environment as a function of direction, time, wavelength and viewpoint. Adelson and Bergen call this formalism the plenoptic function. Since direction, or position in the visual field, requires two parameters and viewpoint three parameters, we have a seven-dimensional space which for most purposes can be reduced to five parameters by specifying the locations of the eyes in space and allowing the fifth parameter to index the two retinal images.

The advantage of this perspective is that it allows one to characterize many of the basic measurements involved in the visual process as the measurement of orientation within the multi-dimensional space. Thus, just as spatial orientation is characterized as orientation in the x-y plane, image motion is characterized as orientation in space-time and we can characterize binocular parallax or orientation in a space–eye solid in which the images from the two eyes are placed in register. This framework offers the possibility of a general theory of low-level motion which describes how orientation is measured and which is then interpreted within the context of a particular low-level vision domain. Of course any generalization would have to be qualified by an understanding that a particular implementation within a given domain must be limited by the physical constraints placed upon the biological system; nevertheless the notion of a general model offers a useful heuristic.

2. A COMPARISON OF COMPUTATIONAL STRATEGIES

There are three general strategies for computing orientation which have been adopted in the modelling of low-level motion mechanisms. It is revealing to compare the generic models, in terms of the filter kernels used to collect the initial measures and the strategies for combining results, within the space–time framework described above (Fig. 1). The correlational approach (Reichardt, 1961; van Santen & Speier, 1985) combines detectors for spatial pattern which are separated in both space and time. In order to detect motion in a particular direction the response of one of the filters is delayed or shifted in time so that the combined response can signal the presence of an oriented contour in space–time. It is clear that the response will be velocity tuned. Since the signals from the paired filters are multiplied the output is also dependent upon contrast. The final output is the difference between rightward and leftward detectors. Because the bilocal detector is velocity tuned and contrast dependent a single detector could not signal velocity. The correlation scheme is therefore only a partial theory of velocity encoding since it lacks a well specified scheme describing how elementary motion detectors could be combined to compute velocity.

The second approach uses energy measures. The first and critical step in the motion energy approach is the construction of filter kernels which are oriented in space–time (Adelson & Bergen, 1985; Freeman & Adelson, 1991). The oriented detectors signal the

*Department of Psychology, University College London, Gower Street, London WC1E 6BT, England [Email alan@psychol.ucl.ac.uk].
Reichardt Correlation

\[(A \times Bd) - (B \times Ad)\]

Motion Energy

\[R - L\]

Space-Time Gradients

\[\frac{T}{S}\]

The presence of motion in a particular direction. However, depending upon the velocity tuning of the filter it may also respond to static pattern. Thus, like the Reichardt model, the final output of the motion energy model is a comparison between rightward and leftward motion detectors [Fig. 1 (middle)]. In the full model, pairs of filters in quadrature phase (not shown) are applied to images and the outputs are squared and added to provide a strong response when there is significant variation in the image. Intuitively, the addition of the output of filters in quadrature phase ensures a response whether the image intensities are edge-like or bar-like and the squaring operation provides an indifference to the sign of the changes. This general technique was introduced by Adelson and Bergen (1985) in the motion domain and has also been used to model the computation of binocular disparity (Ohzawa, DeAngelis & Freeman, 1990), feature detection (Morrone & Burr, 1988) and spatial orientation (Bergen & Landy, 1991). However, the generic model is sensitive to contrast and various explicit normalization techniques need to be adopted to disambiguate the measure from image contrast (Heeger, 1992).

The alternative is to measure orientation using measures of the spatio-temporal gradients of intensity (Fennema & Thompson, 1979; Heeger & Simoncelli, 1995; Horn & Schunck, 1981; Johnston, McCowan & Buxton, 1992; Sobey & Srinivasan, 1991; Verri, Straforini & Torre, 1992; Young & Lesperance, 1993). In the case of one-dimensional moving patterns, which is the domain of interest here, the orientation is given by the ratio of the derivatives of intensity with respect to time and space

\[v = \frac{\partial t}{\partial I} \cdot \frac{\partial I}{\partial x}.\]

Figure 1 (bottom) shows a simple gradient scheme using first derivatives. Note the kernels are rotated in space–time relative to the energy scheme. Velocity is computed by forming the quotient of the filter outputs. This is much the simplest of the three generic models since it only requires two spatially overlapping, contiguous, space–time separable filters; the transient filter, T, which computes the temporal derivative and, the sustained filter, S, which computes the spatial derivative. Recall that in the energy model the inseparable space–time oriented filters are constructed from separable filters. For a static pattern, vertical lines in space–time, the output of the excitatory and inhibitory regions of T are equal and the quotient takes the value zero. A moving pattern gives rise to sloping contours. As the contour rotates, the value of T increases while S decreases in the correct proportion to signal velocity. With this simple quotient the sign depends upon both the direction of rotation and the polarity of the contour. The output is directly related to velocity and the result is contrast independent since increasing contrast affects the output of the T and S filters similarly (Johnston & Wright, 1985, 1986).
3. CONDITIONING IN THE GRADIENT APPROACH

However, there is a fundamental problem. In addition to the sensitivity to contrast polarity, the simple gradient scheme for coding orientation gives rise to infinities at extrema and at inflections with zero gradients where the first spatial derivative becomes zero. This problem can be overcome either by pooling over space (Heeger & Simoncelli, 1995; Lucas & Kanade, 1981; Simoncelli, Adelson & Heeger, 1991) or by including a series of measures of how the spatial derivatives from first order to order $n$ are changing with respect to space and time (Johnston et al., 1992). In either case the measures can be combined using a standard least squares formulation to arrive at the image speed. The disadvantage of pooling over space in order to condition the quotient in the least squares formulation is that it leads to a loss in spatial resolution. This can be reduced by scaling the filter outputs with respect to a Gaussian spatial weighting function (Heeger & Simoncelli, 1995; Simoncelli et al., 1991).

The gradient model is described in some detail in an earlier paper (Johnston et al., 1992) and the current version of the model is described in the Appendix. We assume simple cells act as blurred differential operators (Koenderink, 1987; Young, 1987; Young & Lesperance, 1993) and that cells in V1 represent image brightness in the form of a local Taylor series expansion (Koenderink, 1987). We can take any of the terms of the Taylor series and compute the ratio of its temporal and spatial derivatives to arrive a measure of the image speed, with the implicit assumption that velocity is constant locally. However, any single measure will be ill-conditioned so we combine measures of the spatial and temporal derivatives of a number of terms of the Taylor expansion using a least squares formulation to find the speed of the motion. We refer to this version of the gradient model as the Multi-channel Gradient Model version 1 (MCGM v.1).

It is not clear whether spatial filters in the visual system have sufficient overlap in their receptive fields to implement convolution, or represent the output as an image at the resolution found in the retinal receptors, although this spatial organization is implicit in virtually all spatio-temporal filter models of low-level visual mechanisms, including the model presented here. Pooling over space leads to a blurred representation in the output at the same sampling rate as the image. A Taylor series representation provides a means of conditioning the quotient in the gradient model and also allows the visual system to generate an accurate estimate of image brightness or brightness change in a neighborhood of the point at which the Taylor series is computed; the more terms used the better the estimate, or the greater the region that fails below some criterion error.

Thus from local measures the visual system could represent the brightness changes in a region out with the spatial confines of the constituent receptive fields. Alternatively, the representation would allow the generation of an estimate of image motion for the region taken as the spatial support of the filter kernels. We see this capacity for estimation at a distance and interpolation between sampling points as a particularly advantageous characteristic of the Taylor series approach. It would also support the construction of a sparse representation of the image in visual cortex but this feature has not been incorporated into the model. The MCGM v.2 used the local approximation to estimate brightness gradients in the neighborhood around the point of interest in order to compute the spatial and temporal derivatives of intensity expressed relative to the local mean. This change to the original model smoothed the output and improved the ability of the model to detect the motion of a contrast modulation signal carried by band-limited, spatio-temporal noise.

4. MULTIPLE MOTION SYSTEMS: CLASSIFICATION VS UNIFICATION

Braddick (1974) drew a distinction between two ways in which the motion system operates which he referred to as the short range process and the long range process. This classification led to the idea of two types of motion mechanism, one based on the responses of simple motion sensitive filters which had a limited spatial support and an additional mechanism based on matching features which operates over greater distances and longer time intervals. Cavanagh and Mather (1989) have criticized the use of this distinction and have argued that the criteria used to distinguish between the action of the short range and long range process simply reflect the kinds of stimuli (random dot kinematograms vs isolated patterns) investigators have used to study the motion system. They proposed an alternative classification which rests on a distinction between first-order and second-order motion patterns. A first-order motion mechanism might detect the space-time orientation of a first-order luminance measure. A second-order mechanism might detect the orientation of motion patterns defined by their second-order statistics. Cavanagh and Mather argue for many independent motion detectors at each point in the image which encode motion patterns on the basis of various attributes like luminance, texture, colour or motion boundaries, one for each possible attribute. The results of these motion detectors are integrated at a later stage. Chubb and Sperling (1988) make a similar distinction between Fourier and non-Fourier motion systems. The Fourier system provides a motion energy or correlational analysis of the motion signal. They argue second-order or drift-balanced stimuli cannot be detected by a first-order mechanism and thus we require an additional non-Fourier system which applies motion energy analysis after full wave rectification of the signal. Zanker (1993) detects second-order motion by cascading two layers of first-order motion detectors. Other suggestions have included the proposal of a band-pass temporal pre-filter in order to explain particular motion illusions (Shioiri & Cavanagh, 1990). The general flavour of the classificatory approach is that specific models are proposed to deal with specific classes.
of motion phenomenon. Although this may be valid it lacks parsimony and it introduces problems of integration. In this paper we demonstrate that a number of illusions which have suggested the existence of multiple motion mechanisms can be accommodated by a simple single-stage model.

5. INCREMENTAL MODELLING

We have adopted an approach we refer to as incremental modelling. There are an infinity of models which could be used to compute motion. Any gradient model based on first derivatives can provide an estimate of velocity when the spatial intensity gradients are non-zero. However the simplest gradient formulation considered above leads to infinities at local maxima and minima. Similarly, we found that there are a number of formulations of the gradient model which accurately compute the motion of gratings. To constrain the modelling process further, and to guide further developments, we need to select more complex motion patterns. The approach is incremental in that changes made to accommodate results with new patterns should not impair the models performance on previously studied patterns. The main aim of the model is to provide a rapprochement between strategies for computing speed which have been successfully applied in computer vision and the detailed evidence from visual psychophysics and neurophysiology on the properties of the spatio-temporal channels in the human and primate motion analysis systems, and how these spatio-temporal channels are combined to compute motion.

Of particular interest in the context of the present paper we found that the McGM predicted a clear reversal of perceived direction in a reverse phi stimulus used by Chubb and Sperling (1989) as strong evidence of two motion systems. When a grating made up of one-quarter cycle black and three-quarter cycle grey regions is contrast reversed and shifted forward one-quarter cycle at a fixed temporal rate [Fig. 2(a)], we see forward motion from close to the display but reversed apparent motion from a distance. The McGM predicts motion in the expected direction for wide bars but predicts motion in the reverse phi direction as the bars are narrowed, simulating a change in viewing distance. Shioiri and Cavanagh (1990) have reported that reversed apparent
motion can be induced by introducing a blank grey interval into a standard temporal sequence and also that the reverse phi effect can be eliminated by the same manipulation. We decided to investigate this effect and some other related illusions which have led to the proposal of multiple motion mechanisms to see whether these conclusions were justified.

6. TRANSLATING BETWEEN THE MODEL SPACE AND REAL SPACE

The results of the model are in the form of images which describe speed as a function of space and time. They follow a convention that the border of each frame is set to zero and then each frame is scaled to the full brightness range. We are only interested in modelling the perceived direction of motion rather than perceived speed, which is not measured in experiments of the type discussed here. Motion to the left is signalled by points which are lighter than the border and motion to the right is signalled by points which are darker than the border. Where appropriate we report a directional index (DI) based on the sum of the positives and negatives (absolute values) taken over space and time (positives — negatives/positives + negatives). The DI provides a useful global output measure for the model but we are aware that the visual system and the observer may utilize various measures of the velocity field to perform a particular direction discrimination task.

![Diagram](image.png)

**FIGURE 2.** (a) This figure demonstrates the usual explanation of the reverse phi effect, that there is a significant amount of motion energy in the reverse direction which can be detected by filters which are oriented in space-time. (b) The results of the McGv v.3 for the Chubb and Sperling reversed apparent motion stimulus. The space-time stimulus is shown on the left and the results, in the form of a space-time plot of speed, are shown on the right. The stimulus is presented for 8 frames then shifted by a quarter cycle. In Chubb and Sperling's experiment the stimulus was present for 66.6 msec prior to displacement. We have estimated (see text) that each frame in our space-time pattern is equivalent to around 7.8 msec and so the space-time plot corresponds quite closely to the conditions used by Chubb and Sperling. The model correctly predicts motion in the direction of the travelling contrast reversing bars for near-view conditions (top row) and reversed motion for far-view conditions (middle row). In each case the stimulus is shown on the left and the output is shown on the right. The power spectra of the stimuli are shown centrally. The output follows the convention that the border of each window is always set to zero and then each window is scaled to the full brightness range. Leftward motion is signalled as brighter than the border and rightward motion as darker than the border. When we introduce a blank IFI (12 frames) to the Chubb and Sperling stimulus (bottom row) the reversed motion effect is replaced with forward motion.
In order to provide a means of translating between the space–time units of the modelling space and those of the real world we measured the high frequency limits of the envelope of our spatio-temporal differentiating filters. We assume a high spatial frequency limit of 60 c/deg and a critical flicker fusion limit of 60 Hz for the human visual system. The image has a time span of 1 sec with each pixel representing a frame duration of 7.8 msec. To estimate the amount of visual space represented by 1 pixel we assumed differentiating filters up to order 8 and matched the upper spatial frequency limit with that of the human system. This procedure gave a value of 32.5 sec arc per pixel making 1 c/image equivalent to 0.86 c/deg. Details of the form of the zero-order filters and filter parameters are given in the Appendix.

7. THE REVERSE PHI MOTION ILLUSION

(a) Results for the McGM v.2

The reverse phi illusion described by Anstis and Rogers (1975) results from the combination of a discrete displacement and a contrast reversal. It is generally believed that the reverse phi effect can be explained on the basis of motion energy analysis (Adelson & Bergen, 1985; Chubb & Sperling, 1989), since the space–time patterns have energy at the opposite direction to that followed by the contrast reversing stimulus element [Fig. 2(a)]. This energy could be detected by a receptive field with an appropriate orientation in space–time (Adelson, 1991; Chubb & Sperling, 1989). However, Chubb and Sperling describe a reverse phi stimulus which appears to move in the reverse direction when viewed from a distance but in the forward direction when viewed close to. The effect of changing spatial scale in this way is to stretch the Fourier transform parallel to the spatial frequency axis. There is no effect on the temporal frequencies in the pattern and there is no indication in the transform that there should be a reversal of perceived direction of motion [Fig. 2(b) middle]. It would appear that an intuitive analysis of the stimulus, based on the motion energy approach, does not lead to a satisfactory account of these effects. Chubb and Sperling argue that the far view stimulus is processed by a Fourier pathway and the near view stimulus is processed by a non-Fourier pathway, which includes full wave rectification on input. However, the data could result from a single mechanism which exhibits this scale-dependent behaviour, and, in fact, the McGM v.2 correctly predicts perceived motion in the Chubb and Sperling stimulus [Fig. 2(b)], including the rather paradoxical prediction that motion is seen in the grey regions.

Using a different kind of stimulus, random dot kinematograms, Shioiri and Cavanagh (1990) report that the reverse phi effect disappears if a blank inter-frame-interval (IFI) is introduced between the discrete steps of the apparent motion stimulus. We decided to see if this effect could be simulated and whether the presence of a grey IFI would influence perception in the Chubb and Sperling display. Since the McGM accepts only space–time patterns our simulations used one-dimensional binary noise patterns. One-dimensional spatial patterns are considered to give similar results to two-dimensional patterns in psychophysical tasks (Baker & Braddick, 1982), however we found it was necessary to scale down the bar width and displacement distance relative to the Shioiri and Cavanagh values in order to unambiguously detect the instantaneous displacement of 1D noise stimuli.

For the Shioiri and Cavanagh motion pattern, the model predicted a reversed phi effect for zero IFI but

![Figure 3](image)

**FIGURE 3.** A schematic of MeGM v.3. The model is described in detail in the Appendix. The elliptical elements represent the lobes of the spatial filters used to compute the spatial and temporal derivatives. Pairs of filter outputs are multiplied and added to form the numerator and denominator. Only the spatial filters up to fourth order are shown. Filters enclosed in dotted lines were added to those in McGM v.2 in order to model correctly some of the illusions considered in the current paper. The ratio N/D provides the motion signal. For one member of the pair on the numerator (N) we have higher orders of temporal differentiation than for those on the denominator (D). The model computes the least squares estimate of velocity given measures of how the higher order derivatives of intensity are changing with respect to space and time. Velocity is assumed to be locally constant.
with the introduction of a blank grey IFI the results tended to ambiguity with equal amounts of motion signalled in both directions. Similar results were found in simulations using the Chubb and Sperling (1989) stimulus, although we could clearly see the elimination of the reverse phi effect when we observed the motion sequence. In fact, we were unable to produce a clear prediction of reversed motion with grey IFIs for any of the motion patterns considered here.

(b) The McGM v.3

Two changes to the McGM v.2 were required to produce a reversed motion signal for grey IFIs. In order to simulate any of the motion illusions considered here we had to include an additional set of filter pairs to those used in version 2 (Fig. 3). These pairs were formed by increasing the order of the temporal derivatives of the original filters in McGM v.2 by one. We also took the opportunity to introduce causal filters in the temporal domain, Gaussians of log time (see Koenderink, 1988), rather than the delayed impulse and its derivative used in McGM v.2, although this change in the form of the differentiating temporal filters was not critical for the prediction of reversed motion. The second order temporal filters are more narrow-band than the first-order filters, peak at a higher temporal frequency, and would constitute a third temporal “channel”. Foster, Gaska, Nagler and Pollen (1985) reported the existence of low-pass, broad-band and narrow-band filters in monkey visual cortex. Low-pass and broad-band filters are more prominent in V1 whereas the broad-band and narrow-band filters are prominent in V2. There is also some psychophysical evidence for three temporal filters. Figure 4 shows the normalized amplitude of the log Gaussian filter and its derivatives along with empirical data from Hess and Snowden (1992) on the shape and sensitivity of the psychophysically determined temporal filters for the human visual system. The functions were fitted to the Hess and Snowden data some time after the simulations were carried out and therefore the value of one of the filter parameters differs slightly from that used in the simulations. Nevertheless there is an excellent correspondence between the form of the functions used for the temporal filters in the model and the psychophysically determined functions.

The second change involved the inclusion of the first-order spatial and temporal derivatives of our zero-order kernel in the set of spatio-temporal filters (Fig. 3). Figure 5 shows the power spectra of the derivatives of Gaussians for the filters used in the simulation. The introduction of the first-order spatial and temporal derivatives was necessary to model an effect of changing the IFI brightness on the Pantele illusion discussed below. In McGM v.2 we omitted the first-order temporal derivative kernel as we were unaware at the time of good physiological evidence for linear filters with transient temporal properties which had spatially unstructured (Gaussian) receptive fields in primary visual cortex. However cells with Gaussian spatial fields which change their polarity over time have been found by Pollen, Gaska and Jacobson in monkey V1 (Daniel Pollen, personal communication). Since we have introduced the first-order derivatives of space and time the model now computes the direction in which brightness is conserved rather than the direction in which the difference in brightness from the local mean is conserved as was the case in the McGM v.2.

(c) Results of McGM v.3

The results of applying the revised model, McGM v.3, are shown in Figs 2(b) and 6(a, b). Now we find the reversed apparent motion effect disappears and is replaced by forward motion as described by Shioiri and Cavanagh. In Fig. 7 we plot the direction index as a function of IFI duration expressed in frames. Each point is the average of three determinations of the index for three different random patterns. For the reverse contrast patterns the reversed apparent motion effect is eliminated for IFIs greater than 4 frames.

FIGURE 4. The sensitivity of the three temporal filters—the log Gaussian and its derivatives. $\alpha = 10.0$, $\tau = 0.268$. The data are psychophysical measurements of the three temporal filters in the human visual system taken from Hess and Snowden (1992).

FIGURE 5. The power spectrum of the derivative of Gaussian filters used in the simulations. Only the even orders are shown. The eighth spatial derivative of the Gaussian with the spatial parameters used here would intersect the spatial frequency axis at 70 c/deg. The filters shown above have been scaled so that the highest filter has a high spatial frequency cut off at 60 c/deg.
8. MOTION ILLUSIONS INDUCED BY A GREY IFI

(a) Random binary 1D noise

Shioiri and Cavanagh also report that the introduction of a grey IFI is sufficient in itself to generate a reverse apparent motion illusion. However, we should note that subjects do report ambiguous motion percepts in these studies and they were asked to respond on the basis of the stronger of the two motion percepts when unsure. We found in simulations using the McGM v.3 that for one-dimensional noise there was a bias towards seeing motion in the reverse direction for IFIs greater than 4 frames in duration [Figs 6(c, d) and 7]. They argue this effect is evidence for the existence of a band-limited temporal filter placed before the motion analysis stage. The simulation demonstrates this two-stage approach is unnecessary although the essential element of Shioiri and Cavanagh’s interpretation is supported, since the presence of band-limited filters in both the numerator and denominator of the gradient model is a requirement for this effect. There is an interesting symmetry between the results for conventional and contrast reversed apparent motion. If a sine wave gratings is contrast reversed and then shifted by $\phi$ deg the stimulus is identical to that produced by a $\pi - \phi$ degree shift in the opposite direction. It is presumably this physical constraint which gives rise to these symmetrical effects.

The maximum detectable displacement for one-dimensional random binary noise patterns was estimated to be $< 1.75$ min arc. This value was much lower than the values of 15–20 min arc found for $D_{\text{max}}$ in random dot kinematograms (Braddick, 1974). Since we used the upper limit of spatial resolution to scale the outputs of the model, one possibility is that there are other banks of linear spatial filters at a lower scale. Alternatively, there may be some further pooling of signals over space in addition to that present in the current version of the model.

(b) Instantaneous three-sixteenths cycle phase shifts

Pantle and Turano (1992) describe a similar motion illusion to the Shioiri and Cavanagh effect which in our hands appears to be more robust than the noise displacement effects. They found that a grating displaced by thirteen-sixteenths of a cycle with the presentations separated by a grey IFI appears to move in the “long path”, thirteen-sixteenths of a cycle, direction. Figure 8(a, b) shows the results of simulation using McGM v.3. The introduction of a blank grey IFI leads to a clear reversed apparent motion signal. Figure 9 shows how the direction index changes with IFI duration. The reversal in direction occurs for IFIs greater than 4 frames. Note there is no reason to suppose the visual system is matching corresponding points in the “short path” or “long path” conditions. Long path motion simply reflects a motion reversal. Contrast modulated gratings in which the modulation frequency matches the grating

\[ \text{Direction index} \]

\[ \text{Interstimulus interval (frames)} \]
AN ACCOUNT OF THREE MOTION ILLUSIONS

9. THE FLUTED SQUARE WAVE MOTION ILLUSION

The fluted square wave or missing fundamental motion illusion was introduced by Edward Adelson (Adelson, 1982; Adelson & Bergen, 1985). A square wave grating moving to the right in discrete quarter-cycle steps appears to move in the forward direction but, if the fundamental component frequency is removed to leave a fluted or scalloped square wave grating, the pattern appears to move in reverse direction. It would appear we detect the motion of the most prominent component, the third harmonic, which, if isolated, would indeed be moving in the reverse direction. The motion energy model indicates reversed motion which mirrors human perception. However, Mather (1990) has shown that the simple gradient model gives similar results in two-frame displays.

Georgeson and Harris (1990) reported that the introduction of a blank IFI between discrete quarter-cycle displacements of the fluted square wave eliminated the illusion and reinstated the percept of forward motion. They concluded that the reversed motion indicated the operation of the short range process or simple motion filter operations whereas the forward percept indicated the operation of a long range feature matching system.

We investigated the effects of introducing a blank IFI on the response of the McCGM. Version 2 of the model signalled reverse motion for an instantaneous displacement. As we increased the duration of the IFI the response of the model reduced in amplitude and tended to ambiguity with equal amount of motion signalled in each direction. The results for McCGM v.3 were quite different. We found that the direction of motion signalled by the model reversed for IFIs greater than 4 frames indicating forward motion (Figs 10 and 11). Thus there is no need to posit two motion mechanisms on the basis of their observations. The McCGM v.3 can accommodate the time dependent behaviour. The results for

Frequency used in the three-sixteenths-cycle displacement display do not give rise to reversed motion (Pantle & Turano, 1992) and this effect is predicted by the McCGM v.3 [Fig. 8(c)].

(c) Changing IFI brightness

For a dark IFI Pantle and Turano report no reversed motion illusion. In simulations they found that the motion energy model gave a “warped” response surface but the mean level of the output did not differ from the grey IFI condition. With the McCGM v.3 we find that the reversed motion effect is reduced to ambiguity for a dark IFI [Fig. 8(d)]. This result is dependent upon the inclusion of the zero-order spatial filters. If we remove these filters the IFI brightness level has no effect on the output of the model.

FIGURE 8. The space-time stimulus is shown on the left and the results are shown on the right. The grating pattern is presented for 16 frames prior to displacement which we estimate to be equivalent to 125 msec. In the Pantle and Turano experiments the presentation frame duration was 288 msec and the direction of displacement altered from trial to trial. We stepped the grating forward to simplify the calculation of the direction index. (a) A discrete three-sixteenths cycle step appears to move to the left. The results of the model show this effect. DI = 1.0. (b) When a blank IFI is introduced the model correctly signals reverse motion. DI = -0.48. (c) This does not occur for a contrast modulation. DI = 0.09. (d) The results tend to ambiguity with a dark IFI. DI = -0.04.

FIGURE 9. The direction index plotted against IFI duration in frames for displaced sine wave patterns. The input images were as described in Fig. 8. The data show a clear reversal of motion direction for IFIs of greater than 4 frames.

(c)
10. MOTION ILLUSIONS AS DIAGNOSTIC EPIPPHENOMENA

A number of visual paradigms have been considered in this paper, each of which have given rise to the notion of multiple motion mechanisms. We have shown that much of the psychophysical data can be accommodated by a single motion model. In our view these illusions are diagnostic epiphenomena which should allow us to select between various competing motion models. There is little to be gained from a modelling approach with the specific aim of generating any of these particular behaviours or a theoretical approach which attempts to explain them in isolation. Here, the motion illusions are seen to be emergent properties of a particular computational strategy. Clearly, quite complex behaviour can emerge from a simple approach to motion computation and one should be cautious in attributing visual experience to multiple causal agents or mechanisms on the basis of complex or perplexing psychophysical data. In particular, reversals of perceived motion are not indicative of the presence of two motion mechanisms.

All versions of the model perform equally well for moving gratings. In the first version of the McGM we utilised the fact that speed could be computed by forming a quotient with any pair of spatio-temporal differentiating filters in which the order of the temporal derivative on the numerator exceeded the temporal derivative on the denominator by one and the order of the spatial derivative on the denominator exceeded the spatial derivative on the numerator by one (Koenderink, 1987),

$$v = \frac{D_{(x,j+1) i}}{D_{(i+1) x,j} i}$$
However for the reasons stated above this formulation is ill-conditioned and therefore we included many pairs of measures combined using a least squares formulation. In McGM v.2 and v.3 the weights are derived from the Taylor series expansion of the image at the point of interest and they are used to estimate the values of the spatio-temporal derivatives in a region around that point. In version 2 we calculate the speed using the rates of change of image brightness expressed relative to the local mean with respect to time and space. This provided a smoother output and made the model more sensitive to the motion structure in second-order motion patterns like those generated by the translation of a contrast modulation signal acting on static sine gratings or band-limited noise.

However earlier versions did not predict the reversals of motion seen in the apparent motion illusions described above. In order to predict these effects it was necessary to extend the Taylor representation of the motion signal to include first derivatives with respect to time and to include the image brightness at the point of interest. As a consequence of this when the filters are constructed to compute the spatial and temporal derivatives of this representation we now have three temporal filters implementing the zero-, first- and second-order partial derivatives with respect to time. The partial derivatives which are first order with respect to both space and time are also now included in the set of filters used in version 3. We show in the Appendix that McGM v.3 is equivalent to computing speed as the direction in space–time which minimises the change in brightness for that region of interest.

It is expected that the model will undergo further revision in order to accommodate other properties of the human and primate motion analysis system. One failing of the current version of the model is that it is insensitive to large instantaneous displacements of random noise. We suspect that this is due to the fact that, although the set of spatial derivatives cover the full spatial frequency range, the filter kernels have a limited spatial support. It would of course be possible to include additional sets of filters with larger spatial support in future revisions of the model.

11. ON THE EQUIVALENCE OF MOTION MODELS

In the psychophysical literature there has been a tendency to argue for multiple motion systems. In contrast, in the computational modelling literature, there has been efforts to establish a formal equivalence for some versions of the generic motion models. Although correlation models, energy models and the spatio-temporal gradient approach are conceptually quite different, the claims for formal equivalence might lead to the impression that all motion models are essentially the same. However, it should be noted that the energy model shown to be equivalent to one version of the Reichardt correlational model (Adelson & Bergen, 1985) is different to the model compared with the gradient scheme (Adelson & Bergen, 1986) and, of course, it is a sensitivity to the differences between predictions of the various motion models which allow them to be tested against empirical evidence. Emerson, Bergen and Adelson (1992) found some evidence for the pre-opponent stage of the energy model in cat complex cells but no evidence for the correlational model. They did not compare their results against a gradient model. In this paper we compared two versions of the McGM which provided different predictions in three psychophysical tasks.

Adelson and Bergen (1986) demonstrated a connection between a gradient scheme and an energy measure constructed by the difference signal taken between space–time oriented motion detectors scaled by the response of a static filter. For an equivalent, the oriented filters have to be constructed from the derivatives of the same kernels used in the gradient model. Adelson and Bergen (1986) work with derivatives of Gaussians. The key insight is that, for first derivatives of Gaussians, a derivative in any direction can be computed from a weighted sum of the results of taking a derivative in any two different directions. So in Fig. 1 each of the filters in the gradient scheme (bottom) can be replaced with a weighted sum of the oriented filters used in the energy model (middle). But in the energy formulation the oriented filters are constructed from the very space–time separable filters whose action we are attempting to duplicate. The gradient scheme is equivalent to the motion energy scheme if the key stage in the energy model, the construction of space–time oriented inseparable filters from separable filters, is rendered redundant. Heeger and Simoncelli (1995) use a similar argument to Adelson and Bergen to express a gradient scheme based on third-order derivatives in terms of space–time oriented filters but the approach is not easily extended to high orders of differentiation, since to “steer” derivatives of Gaussians in this way you need a set of oriented filters, tuned to different orientations, one element greater than the order of the operator whose action you wish to compute (Freeman & Adelson, 1991).

In addition, for a two-dimensional Gaussian kernel, the derivative in any direction will be oriented in
space–time. If a filter with an asymmetric temporal profile is used then the derivative in a given space–time direction will not generally be oriented along a single, dominant direction. Thus it is not generally the case that the spatial and temporal derivatives of a filter kernel can be linearly combined to produce a spatio-temporally oriented filter suitable for motion energy calculation. In particular, the equivalence does not hold for asymmetric temporal filter profiles necessary in the design of causal filters, like the filters used in the McGM v.3. Thus, while both energy and gradient models can employ causal filters, the two classes of model will not be equivalent in such cases. Since the problem of motion perception demands temporal causality, any physically realizable gradient scheme will not be equivalent to a motion energy approach based upon spatio-temporally oriented filters.

The motivation for the rightwards and leftwards motion detectors in the energy model comes, in part, from the observation that adaptation to motion in one direction leads to an after effect of motion seen in the opposite direction. The motion energy model involves opponency between velocity tuned motion detectors. However, the magnitude of the motion after effect is temporal frequency tuned not velocity tuned (Wright & Johnston, 1985) and recently Smith and Edgar (1994) have shown adaptation can increase perceived motion in patterns moving in the same direction as the adapting stimulus, a finding which sits more easily with a ratio model than within the motion energy framework. Formal relationships between models are interesting but there are clear conceptual and procedural differences between the various approaches. A complete model of the human motion system will require the specification of a particular algorithm for computing motion and an explicit implementation.

12. THE SPATIO-TEMPORAL FILTERS

The challenges to the McGM provided by the data considered here led to two modifications. The first modification involved the inclusion of first-order temporal differentiating filters. These filters were needed to model the effects of changing the brightness of the IFI in the Pantle illusion. With hindsight we can see that filters with no inhibitory regions in their spatial receptive fields are necessary to explain the now classical observations (Robson, 1966) that the contrast sensitivity function is low pass at high temporal frequencies. However, these linear non-oriented transient “cells”, which also lack any centre–surround organization, are not, to our knowledge, generally included in any classification of simple cells in V1. They have however been seen with white noise cross-correlation techniques (Daniel Pollen, personal communication). The inclusion of first order spatial differentiating filters, sustained “edge detectors”, is uncontroversial.

Although there is general agreement that there exists a relatively large number of spatial channels at each point in the visual field the question of whether there are two temporal filters or three spanning the visible range of temporal change has been the subject of recent debate (Hammett & Smith, 1992). Watson and Robson (1981) showed that, at threshold, low temporal frequencies could be distinguished from high temporal frequencies but no finer discriminations could be made, suggesting just two temporal channels. That temporal frequency discrimination can improve at high temporal frequencies (Hess & Plant, 1985; Mandler & Makous, 1984) has been taken as evidence for a third temporal filter but Hammett and Smith have argued that this may be due to subjects responding on the basis of the salience of the fading of contrast seen at high temporal frequencies.

However we have evidence for the third temporal filter from other techniques. Hess and Snowden (1992) measured the masking effect of flickering one-dimensional spatial noise on the detection of near-threshold, contrast-reversing sinusoidal gratings and found tentative evidence for a third temporal filter which peaked above 8 Hz but which had similar temporal characteristics to the mid-range band-pass filter. We found that a third temporal filter, the second-order temporal differentiating filter, was necessary to simulate the changes in perceived direction found with blank IFIs in the apparent motion displays studied here. The simulations therefore provide further evidence for a third temporal filter. We fitted averaged data presented in Hess and Snowden using the log Gaussian temporal differentiating filters used in this study (Fig. 4). There is a remarkably good fit between the filter functions and the data although for the low-pass filter the data points at high temporal frequencies have higher values than predicted. The best least-squares fit was calculated by adjusting the spread of the log Gaussian while keeping the σ parameter at 10.0. The normalized data was also adjusted vertically. This procedure resulted in a slight increase in the value of τ over that used in the simulations. What is interesting is that the model parameters are constant for the three filters. The variation in the shape and position of the filters depends only upon the processes of differentiation. We take this as strong empirical evidence for the specific filters used in this study and the multi-channel gradient strategy.

13. SUMMARY

We demonstrated that the current version of the McGM model was able to predict the appearance and elimination of reversed motion in a number of motion tasks. It was also possible to predict accurately the duration of grey IFIs above which we should expect to find reversals of motion by scaling the results of the model using measures of the spatio-temporal limits of the filters in the human visual system. In the temporal domain the log Gaussian and its derivatives provided an excellent fit to published data on shape of temporal tuning curves. The current model did however provide an underestimate of the upper limit on the displacement of one-dimensional noise. In order to extend the upper
limit it may well be necessary to introduce additional filters tuned to a lower spatial scales.

REFERENCES


APPENDIX


Acknowledgements—The research was supported by grants from the Joint Councils Initiative in Cognitive Science/HCI and the Wellcome Trust. We would like to thank Daniel Pollen, James Gaska and Lowell Jacobson for allowing us to refer to their work prior to publication.

$\text{McGM} v.3$

We assume that velocity is locally constant. The image brightness around a point of interest, $f(x + c, t + h)$, can be represented by the equivalence class of functions which agree in their two-dimensional Taylor expansion up to first order in time and order $n$ in space i.e.

$$f(x + c, t + h) = \sum_{i=0}^{n} \sum_{j=0}^{i-j} \frac{c^h}{(i-j)!} D_{i,j} f(x, t),$$  \hspace{1cm} (A1)

where $D_{i,j}$ denotes a differential operator which, when applied to the image brightness at $(x, t), f(x, t)$, yields the $i$th partial derivative with respect to $x$ and the $j$th partial derivative with respect to $t$. These 2D $n$th derivative values can be computed either by convolving the image with a smoothing kernel and then differentiating, or, as is usual, convolving the image with the derivatives of the smoothing kernel, taking advantage of the fact that the derivative of a convolution is the convolution of either of the two functions with the derivative of the other (Bracewell, 1965). Details of the zero-order smoothing kernel used in the McGM are given below. The Taylor expansion about a point can be used to predict the brightness within a neighbourhood of that point. Indeed using this approach we can calculate the average brightness over the spatio-temporal region which provides the support for the zero order, undifferentiated filter kernel. For a cortical cell, this would correspond to the area of its receptive field.

If we were to integrate the contributions of the Taylor series over a region around the point of interest $-a \leq c \leq a$, $-b \leq h \leq b$, we would have

$$g(x, t) = \frac{1}{(4ab)} \int_{-a}^{a} \int_{-b}^{b} f(x + c, t + h) dh \, dc.$$  \hspace{1cm} (A2)

In practice this integral value would be computed approximately as a discrete summation,

$$g(x, t) = \frac{1}{(2a + 1)(2b + 1)} \sum_{-a}^{a} \sum_{-b}^{b} f(x + c, t + h)$$  \hspace{1cm} (A3)

We then might decide to compute velocity simply by taking the ratio of this average brightness value

$$v = \frac{\partial g(x, t)}{\partial t}$$  \hspace{1cm} (A4)

but this quotient would be ill-conditioned as the denominator can take zero values at extrema. This formulation would be equivalent to computing the partial derivatives of the smoothed input image. If we maintain the dimensionality inherent in the original measures we may achieve a well conditioned estimate of the speed. Therefore, we consider each of the terms associated with the Taylor series approximation to the average brightness, as given in equation (A3), and take spatial and temporal derivatives of all of these terms to form two vectors $x$ and $t$ which we can use to find the speed. Writing out the terms of $x$ and $t$ explicitly, and ignoring the normalization factor $1/(2a + 1)(2b + 1)$, which cancels later, we have

\[
\begin{align*}
\left( D_{x\cdot f}(x, t) \right) = & \left( \frac{a^{2}}{2!}, \frac{a^{3}b}{3!}, \frac{a^{4}b^{2}}{4!}, \ldots, \frac{a^{n+1}b^{n}}{n!} \right), \\
\left( D_{t\cdot f}(x, t) \right) = & \left( \frac{b}{1!}, \frac{a^{2}b}{2!}, \frac{a^{3}b^{2}}{3!}, \ldots, \frac{a^{n}b^{n-1}}{n!} \right), \\
\end{align*}
\]

\[
\begin{align*}
\left( D_{x\cdot f}(x, t) \right) \cdot \left( D_{t\cdot f}(x, t) \right) = & \left( \frac{a^{2}b}{2!}, \frac{a^{3}b^{2}}{3!}, \frac{a^{4}b^{3}}{4!}, \ldots, \frac{a^{n+1}b^{n}}{n!} \right), \\
\end{align*}
\]

\[
\begin{align*}
\left( D_{x\cdot f}(x, t) \right) \cdot \left( D_{t\cdot f}(x, t) \right) = & \left( \frac{a^{2}b}{2!}, \frac{a^{3}b^{2}}{3!}, \frac{a^{4}b^{3}}{4!}, \ldots, \frac{a^{n+1}b^{n}}{n!} \right), \\
\end{align*}
\]

\[
\begin{align*}
\left( D_{x\cdot f}(x, t) \right) \cdot \left( D_{t\cdot f}(x, t) \right) = & \left( \frac{a^{2}b}{2!}, \frac{a^{3}b^{2}}{3!}, \frac{a^{4}b^{3}}{4!}, \ldots, \frac{a^{n+1}b^{n}}{n!} \right), \\
\end{align*}
\]

\[
\begin{align*}
\left( D_{x\cdot f}(x, t) \right) \cdot \left( D_{t\cdot f}(x, t) \right) = & \left( \frac{a^{2}b}{2!}, \frac{a^{3}b^{2}}{3!}, \frac{a^{4}b^{3}}{4!}, \ldots, \frac{a^{n+1}b^{n}}{n!} \right), \\
\end{align*}
\]

\[
\begin{align*}
\left( D_{x\cdot f}(x, t) \right) \cdot \left( D_{t\cdot f}(x, t) \right) = & \left( \frac{a^{2}b}{2!}, \frac{a^{3}b^{2}}{3!}, \frac{a^{4}b^{3}}{4!}, \ldots, \frac{a^{n+1}b^{n}}{n!} \right), \\
\end{align*}
\]

\[
\begin{align*}
\left( D_{x\cdot f}(x, t) \right) \cdot \left( D_{t\cdot f}(x, t) \right) = & \left( \frac{a^{2}b}{2!}, \frac{a^{3}b^{2}}{3!}, \frac{a^{4}b^{3}}{4!}, \ldots, \frac{a^{n+1}b^{n}}{n!} \right), \\
\end{align*}
\]

\[
\begin{align*}
\left( D_{x\cdot f}(x, t) \right) \cdot \left( D_{t\cdot f}(x, t) \right) = & \left( \frac{a^{2}b}{2!}, \frac{a^{3}b^{2}}{3!}, \frac{a^{4}b^{3}}{4!}, \ldots, \frac{a^{n+1}b^{n}}{n!} \right), \\
\end{align*}
\]

\[
\begin{align*}
\left( D_{x\cdot f}(x, t) \right) \cdot \left( D_{t\cdot f}(x, t) \right) = & \left( \frac{a^{2}b}{2!}, \frac{a^{3}b^{2}}{3!}, \frac{a^{4}b^{3}}{4!}, \ldots, \frac{a^{n+1}b^{n}}{n!} \right), \\
\end{align*}
\]
The dots indicate the numerous levels of implicit nesting of the terms in equation (A3). Horizontal dots indicate increments in the order of spatial differentiation, the vertical dots indicate increments in the spatial location, c, from \(-a\) to \(t\). The diagonal dots show increments in the temporal index \(h\) from \(-b\) to \(h\). Recall that halfway through forming both vectors \(x\) and \(t\) the additional differentiation with respect to \(t\) is required by equation (A1), is implemented. The ratio of each pair of components of the two vectors, (5a) and (5b), taken in order, is a weighted measure of velocity as it comprises a differentiation with respect to time divided by a differentiation with respect to space. The weights given to these pairings are derived from the Taylor series expansion as shown above.

The least squares approximation to the image velocity, \(v\), based on the weighted measures of how the image brightness and its derivatives are changing with respect to space and time (Johnston et al., 1992; Kreider, et al., 1966) is given by the dot product of \(x\) and \(t\) over the dot product of \(x\) with itself, \(x \cdot x\), a term which is equivalent to the square of the magnitude of \(x\).

\[
v' = \frac{x \cdot t}{x \cdot x}.
\]

Since the denominator is a squared magnitude it can only be zero when all the terms in \(x\) are zero, and in that case the numerator is also zero and we have \(0/0\), the indeterminate case. All the spatial derivatives will only be zero when there is no spatial structure in the image, in which case speed is undefined. Since we have assumed velocity to be locally constant, \(x\) and \(t\) are linearly dependent (i.e. parallel) and therefore the value, \(v'\), is the relative length of the two vectors. In addition, because the vectors are parallel the value of \(v'\) which minimizes the distance \(|v'x - t|^2\) also minimizes

\[
\langle (v'x - t) \rangle = \langle (v'x - t) \rangle^2.
\]

Thus the choice of weights provides an estimate of the space-time orientation which minimizes the change of brightness averaged over a region around the point of interest.

As in other gradient models the MCGM provides the direction along which image brightness is conserved. The generic model provides a least squares estimate of image speed based on measures of how the image brightness and its derivatives are changing with respect to space and time. The inclusion of higher order derivatives has the effect of conditioning the quotient. In x.3 we have an additional property, the choice of derivatives and weights has the effect of implementing a brightness constraint based on a Taylor series estimate of the average brightness within a region. This is only strictly true under the assumption that velocity is locally constant but the results of the model should not be very sensitive to any violations of this assumption. In the simulations values of \(a = 5\) and \(b = 1\) were used giving a spatial region of 11 pixels and a temporal region of 3. The derivatives were calculated using blurred differentiating filters. The zero order kernel used in these simulations was the product of a Gaussian in space and a Gaussian of log time, given by

\[
K(x, t) = \frac{1}{\sqrt{4 \pi \sigma}} \exp\left(-\frac{x^2}{4\sigma}\right) \frac{1}{\sqrt{\pi \tau}} \exp\left(-\frac{\tau^2}{4\tau}\right)
\]

with \(\sigma = 1.5, \tau = 10\) and \(\tau = 0.2\). The kernel is scaled so that the integral over its spatio-temporal support is equal to 1.0.