THE ROYAL SOCIETY

doi 10.1098/rspb.2001.1727

Motion detection in human vision: a unifying approach based on energy and features

A. T. Smith* **and T. Ledgeway†**

Department of Psychology, Royal Holloway, University of London, Egham TW20 0EX, UK

Most studies of human motion perception have been based on the implicit assumption that the brain has only one motion-detection system, or at least that only one is operational in any given instance. We show, in the context of direction perception in spatially filtered two-frame random-dot kinematograms, that two quite different mechanisms operate simultaneously in the detection of such patterns. One mechanism causes reversal of the perceived direction (reversed-phi motion) when the image contrast is reversed between frames, and is highly dependent on the spatial-frequency content of the image. These characteristics are both signatures of detection based on motion energy. The other mechanism does not produce reversed-phi motion and is unaffected by spatial filtering. This appears to involve the tracking of unsigned complex spatial features. The perceived direction of a filtered dot pattern typically reflects a mixture of the two types of behaviour in any given instance. Although both types of mechanism have previously been invoked to explain the perception of motion of different types of image, the simultaneous involvement of two mechanisms in the detection of the same simple rigid motion of a pattern suggests that motion perception in general results from a combination of mechanisms working simultaneously on different principles in the same circumstances.

Keywords: vision; motion perception; motion energy

1. INTRODUCTION

One of the most fundamental tasks that must be performed by the human visual system is the detection of image motion. There are at least two general principles that can be used to extract a motion signal from the retinal image. An obvious method is to identify objects, features, edges or other pattern elements within the image and to infer motion from the changes in the spatial positions of these elements over time. This approach has appeared in several guises, the best known being the 'long-range' motion mechanism of Braddick (1980) and the more recent attention-dependent motion-tracking system of Cavanagh (1992). Such a mechanism necessarily involves some degree of spatial analysis of the image prior to motion detection. However, it has been known since the work of Exner (1888) that such a system cannot account for all instances of motion perception, and in recent years a popular alternative theory has been that motion detection is accomplished by a low-level process in which a motion signal is extracted without reference to explicitly represented spatial features within the moving image. Several computational models of this type have been produced, the most influential being the 'motion-energy' model of Adelson & Bergen (1985). This model employs receptive fields that are localized in space and orientated in both space and time (see also Fahle & Poggio 1981). Such receptive fields are constructed using spatial and temporal filters. As a result, only a limited range of spatial frequencies is passed by each motion detector. Therefore, the (local) spatial-frequency content of the image is an important determinant of perceived motion within such a framework. A third account, which may be seen as intermediate between the first two, is that motion is detected by a low-level mechanism whose performance is limited in some way by the spatial features of the image (e.g. the spacing of its elements) rather than by its spatial-frequency content, but which does not necessarily involve explicit tracking of individually identified features (Lappin & Bell 1976; Morgan 1992; Eagle & Rogers 1996). The nature, and even the number, of motion-detection mechanisms that are actually used in human vision remains controversial.

Most studies that are relevant to the question of how motion is detected report measures of perceptual judgements of one type of image and interpret the results in terms of a single framework. Authors advocating particular mechanisms have commonly used quite different stimulus classes in their search for supporting evidence from those used by investigators advocating alternative mechanisms. Cavanagh & Mather (1989) have argued that the effect of this is to differentiate the responses of the motion-detection system to different stimuli, rather than to distinguish between different underlying mechanisms. Where the existence of two mechanisms, one based on motion energy and one on features, has been advocated, it has usually been suggested that the two operate in different circumstances, for example in monocular and dichoptic viewing (Georgeson & Shackleton 1989), in foveal and peripheral viewing (Boulton 1987) or in motion with and without an interval between frames (e.g. Braddick 1980; Georgeson & Harris 1990). Few authors have considered the possibility that two or more mechanisms using different strategies may be involved in the detection of a single type of image motion, each mechanism underlying some subset of the data. Some

*Author for correspondence (a.t.smith@rhul.ac.uk).
†Present address: McGill Vision Research Center, McGill University, 687 Pine Avenue West, Montreal, Quebec, Canada H3A 1A1.

Received 22 January 2001 *Accepted* 30 April 2001

exceptions are the work of Boulton & Baker (1993), who found, using images consisting of multiple Gabor patches, two distinct limits to the detectability of motion depending on the density of the patches, and Mather *et al.* (1985), who developed a display that opposes energy-based motion and the motion of features in terms of direction, and found that each dominates in different circumstances. Still fewer authors have suggested that two mechanisms might be active simultaneously, although Anstis (1980, p. 165) described a display that he claims simultaneously stimulates a feature-tracking system, which determines perceived direction, and a low-level mechanism, which determines the direction of subsequent motion after-effects, while Hammett *et al.* (1993) demonstrated perceived transparency in a display in which feature motion and motion energy were opposed in direction. In this study, the involvement of two quite different mechanisms is demonstrated using a single commonly used class of images. The two mechanisms identified are shown to operate in heavily overlapping ranges of stimulus conditions, i.e. to operate simultaneously.

The image used is the random-dot kinematogram, in which a random pattern of dots, of the type shown in figure 1*a*, is rigidly displaced, and an observer is required to specify the direction of the displacement. The most common application of such patterns (e.g. Chang & Julesz 1983; Cleary & Braddick 1990) has been in measuring the largest displacement (termed $D_{\max}$) at which observers are able to detect the direction of the displacement correctly. The most relevant study of this kind to our work is that of Morgan (1992). In his study, two images, identical apart from a rigid displacement of the entire pattern, were presented in quick succession to give the appearance of motion, and $D_{\max}$ was measured. As the size of the elements of the pattern was increased, $D_{\max}$ was invariant up to a critical element size but increased thereafter (see also Sato 1990). Morgan *et al.* (1997) have recently found that a similar pattern of results is obtained for images from which low spatial frequencies have been removed by filtering; in this case the performance asymptote is at a lower value of $D_{\max}$, but is otherwise comparable. We have examined the perception of direction in two-frame random-dot kinematograms under two conditions. In the standard condition, pairs of high-pass filtered images, identical apart from the displacement, were presented in succession. In the reversed-contrast condition, the first image was the same as in the standard condition but the second image was a reversed-contrast version of its standard-condition counterpart. Reversing the contrast of one image in this way normally reverses the perceived direction of motion, for a variety of different patterns (Anstis & Rogers 1975). Such direction reversal, termed reversed-phi motion, is predicted by motion-energy models, although it can also be accommodated by some types of element-matching model. We have examined performance over an extensive range of displacements, element sizes and filter characteristics. We find that reversed-phi motion occurs in some cases but not in others, and infer the involvement of two mechanisms whose characteristics we discuss. A single-mechanism account (Morgan 1992; Eagle & Rogers 1996) that has previously been proposed to explain data of the same type is rejected.

2. MATERIAL AND METHODS

(a) *Subjects*

Two subjects were used. TL is one of the authors; TF is an experienced paid observer who was unaware of the purpose of the experiment. Both subjects had normal acuity (with optical correction in the case of TF).

(b) *Stimuli*

The stimuli were generated using an Apple Macintosh computer (Cupertino, CA, USA) and were displayed on an Apple monochrome monitor (P4 phosphor) with a frame rate of 67 Hz. A set of images were drawn and stored on disk. Each image was divided into square pixels or elements, each of which was randomly assigned one of two luminance values with equal probability (figure 1*a*). Seven different element sizes were used, ranging from 1.17 arcmin to 75 arcmin. The mean luminance was 48 cd m^{-2} and the Michelson contrast was 50%. Each of the images was then high-pass spatial-frequency filtered using conventional Fourier techniques to remove low spatial frequencies (the limited bandwidth of cathode ray tube (CRT) displays inevitably results in the presence, to some degree, of low frequencies in the displayed image, particularly when combined with the inherent brightness non linearity of the display). The filter used was isotropic with a sharp cut-off (ideal filter). The resulting filtered images, which had eight-bit brightness resolution, were rescaled to the original Michelson contrast. Three different filter cut-off frequencies (the frequency below which all power was removed) were used: 1 cycle degree^{-1} (figure 1*b*), 2 cycle degree^{-1} (figure 1*c*) and 4 cycle degree^{-1} (figure 1*d*). This yielded a set of four images (including the unfiltered version) for each element size, all derived from the same original broadband image. For each image in the series, a reversed-contrast version with the same mean luminance was produced by subtracting the image from a homogenous image of maximum luminance (96 cd m^{-2}).

(c) *Procedure*

In any one trial, subjects were presented with a two-frame sequence, each frame lasting for 75 ms, with no interval between the two frames. The images presented in the two frames always had the same element size and filter cut-off. The two-frame sequences were created as follows. The stored images described in § 2b were large, relative to the displayed images. The first frame displayed was a $5^\circ \times 5^\circ$ square section of the parent image; its location within the parent image was chosen at random. There were two conditions, referred to as the 'standard' and 'reversed-contrast' conditions. In the standard condition, the second frame was selected from the same parent image; in the reversed-contrast condition it was selected from the reversed-contrast version of the same parent image (light becomes dark and vice versa). In both cases, the location of the second frame within the parent image was the same as that of the first frame except for a shift in horizontal position. The two frames were displayed at the same location on the screen and appeared as a displacement of the entire pattern within a static square window; elements that exceeded the boundary disappeared and were replaced by new elements on the opposite side of the image. The square image was presented on a uniform background of the same mean luminance. The image was gamma-corrected with a look-up table. The correction factor was determined by spatially modulating the contrast of the image and adjusting for minimum luminance modulation (Smith 1994).

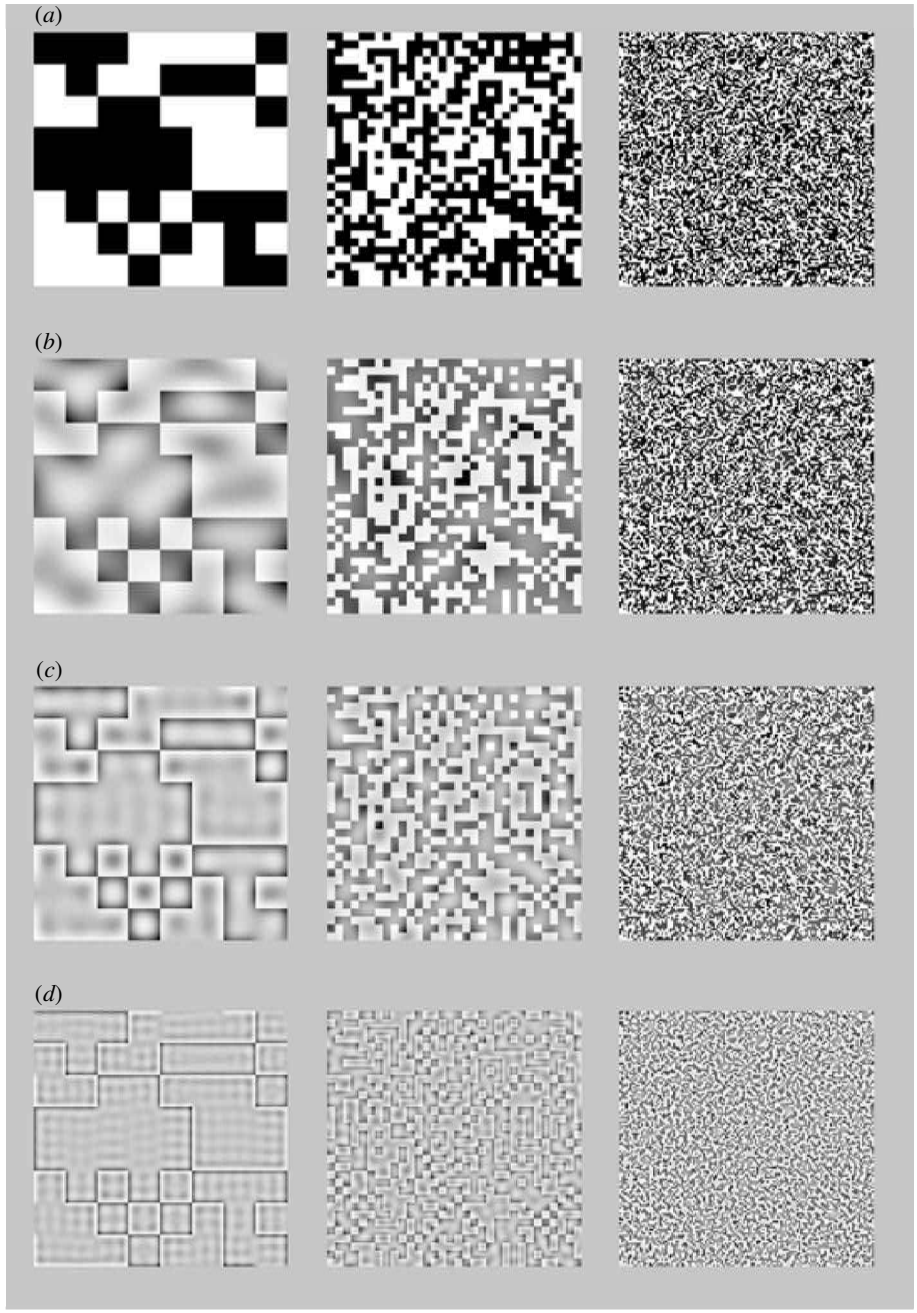

Figure 1. Images typical of those used in the experiment. Part (*a*) shows binary noise drawn with three different element sizes. The other parts show the same images after high-pass spatial-frequency filtering using a filter cut-off of (*b*) 1 cycle degree^{-1}, (*c*) 2 cycle degree^{-1} and (*d*) 4 cycle degree^{-1}. These spatial frequencies assume an image size of $5^\circ \times 5^\circ$. The contrasts of the filtered images have been rescaled to that of the original image.

The displacement could be to either the right or the left, and was determined at random from trial to trial with equal probability. The subject's task was to say, after each trial, in which direction, left or right, the image appeared to be displaced. No feedback was given. Consecutive trials were separated by a 3 s interval during which the screen was homogeneous and had the same mean luminance as the patterns. The displacement distance was varied over a wide (2°) range, in most cases extending well beyond D_{max}, using the method of constant stimuli. The different displacement values were randomly interleaved but the different element sizes and filters were tested in separate blocks. Subjects were first given a brief practice in each condition, and the results were discarded. A total of 40 judgements were then obtained for each displacement value in each stimulus condition, over a period of three or four weeks.

3. RESULTS

Figure 2 shows some of the results obtained with spatially broadband (unfiltered) images. For clarity, results are shown for only four out of the seven element sizes used. The results are shown for both the standard condition (figure 2*a*) and the reversed-contrast condition (figure 2*b*). In the standard condition, performance is near perfect at the smallest displacements (up to about 20 arcmin) for all element sizes. As the displacement

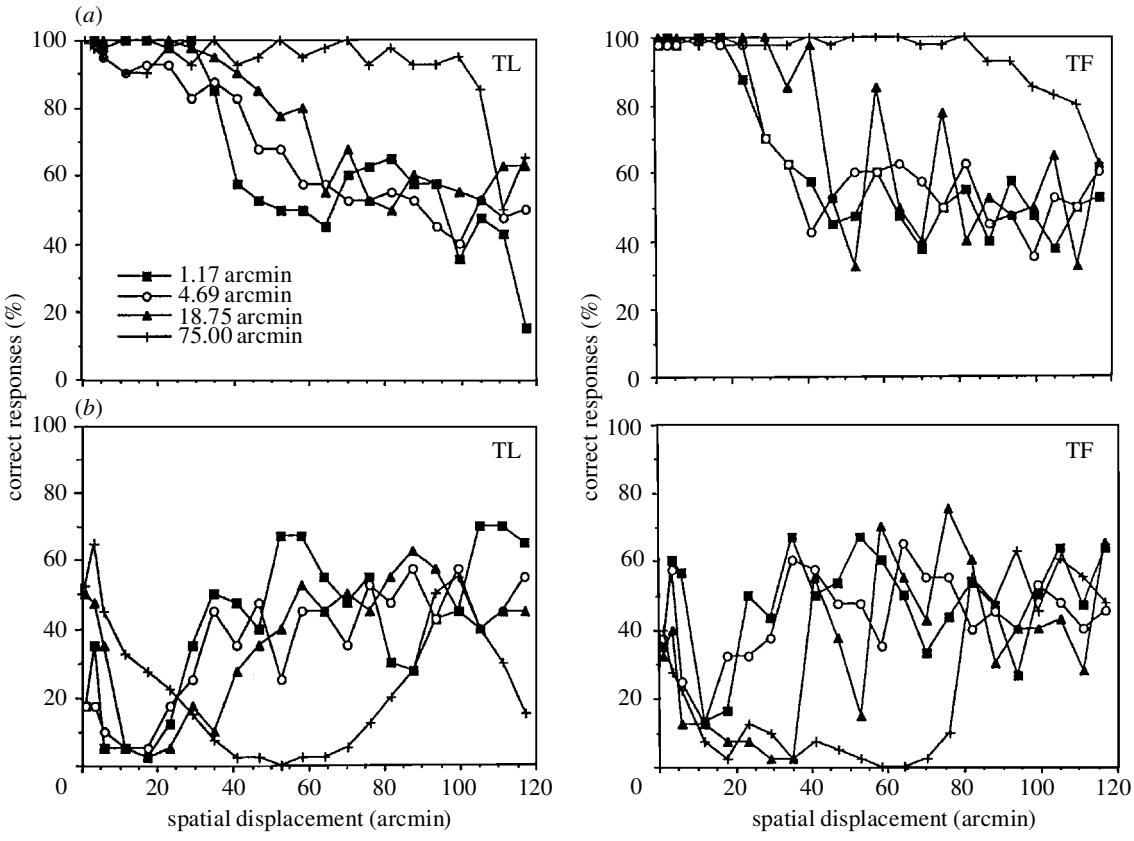

Figure 2. Results obtained using unfiltered (spatially broadband) images. Results are shown separately for the two subjects (left and right plots) and for the standard and reversed-contrast conditions ((*a*) and (*b*), respectively). The plots show direction judgements in terms of the percentage of correct decisions as a function of the spatial displacement between the two frames of the motion sequence. A value of 100% indicates correct identification of the direction of displacement, 50% indicates chance performance and 0% indicates that the subject consistently chose the direction opposite to that of the displacement, reflecting reversed motion perception. Results are shown for four different element sizes.

increases, performance starts to decline towards chance levels. The displacement at which the decline occurs is related to the element size, being greatest for the largest elements. This pattern of results replicates the findings of other authors (e.g. Morgan 1992). In the reversed-contrast condition, a different pattern of results is evident. To a first approximation, the functions relating performance to displacement are inverted with respect to those obtained in the standard condition. At displacements below 20 arcmin, the performance is predominantly below chance, i.e. the observers consistently saw motion in the direction opposite to the displacement (reversed-phi motion). As displacement increased, performance gradually approached chance levels. However, the reversed-phi motion at small displacements is less compelling than the 'forward' motion seen at the same displacements in the standard condition. Although all the functions come close to 0% correct (reliable reversed-phi motion) at some displacement value, they usually do so at an intermediate displacement rather than at the smallest displacements (the most striking example being subject TL, element size 75 arcmin). At the smallest displacements (< 20 arcmin) performance is variable: the trend is usually towards reversed-phi motion, but often the trend is weak. The significance of this aspect of the data will be addressed in § 4c.

The results obtained with high-pass filtered images are shown in figures 3–5. Figure 3 shows the results obtained with a filter cut-off of 1 cycle degree^{-1}. In the standard condition, performance is again near perfect at the smallest displacements. For the smallest element sizes, performance declines sharply at a displacement of about 25 arcmin and reversed motion is seen (though not with perfect reliability) between *ca.* 30 arcmin and 50 arcmin. Forward motion again tends to be seen at displacements of *ca.* 60 arcmin and reversed motion at *ca.* 100 arcmin. Thus, the functions show a cyclical variation between the correct detection of direction and perception of motion in the opposite direction to the displacement. The period of the cycle is very close to one cycle of the filter cut-off frequency, i.e. 60 arcmin (1°). For the largest element size, performance does not show this cyclical behaviour: forward motion is consistently seen up to *ca.* 70 arcmin, then performance declines to chance levels.

In the reversed-contrast condition, the functions, like those for broadband images, are essentially inverted versions of their standard-condition counterparts. Direction perception tends to be reversed below 25 arcmin, veridical between 30 arcmin and 50 arcmin, reversed between 50 arcmin and 70 arcmin, and so on. Again, however, the inversion of the functions is not perfect. In particular, the function for 75 arcmin elements shows the same cyclical

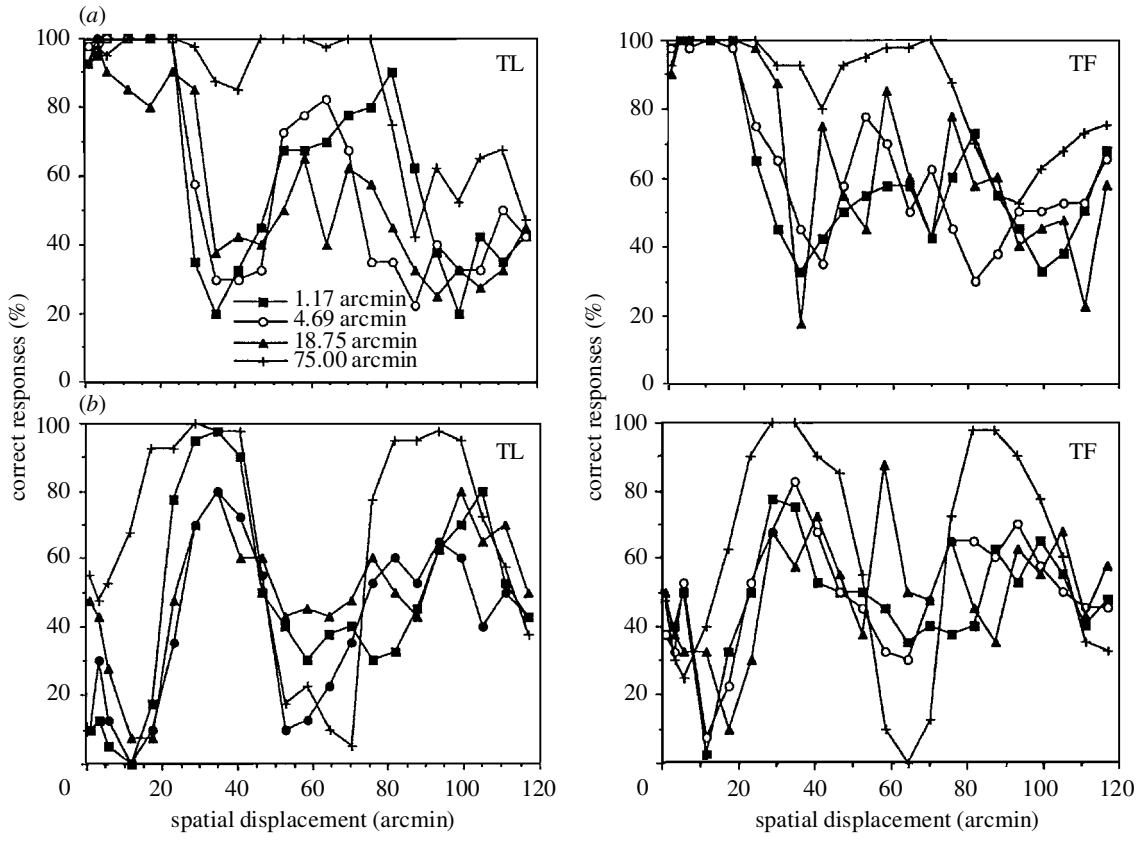

Figure 3. Results obtained using images that were high-pass spatial-frequency filtered using a filter cut-off of 1 cycle degree^{-1}, plotted in the same format as figure 2.

behaviour as the others, whereas in the standard condition it does not.

Thus, two different types of motion reversal are apparent in the data. To avoid confusion, we refer to direction reversal that results from contrast reversal as 'reversed-phi' and to displacement-related direction reversal of the cyclical kind, which occurs both with and without contrast reversal, as 'cyclical reversal'.

With a filter cut-off of 2 cycles degree^{-1} (figure 4) the cyclical performance variations are even more pronounced, and have twice the frequency. In the standard condition direction perception is veridical at the lowest displacements and, for most element sizes, cycles between forward and reversed motion with a period of *ca.* 30 arcmin, again equal to one period of the filter cut-off frequency. However, at the largest element size, as in the 1 cycle degree^{-1} cut-off case, the performance does not show compelling cyclical reversal. In the reversed-contrast condition (figure 4*b*) the functions are again inverted with respect to those obtained in the standard condition. For the largest element size, cyclical reversal is more compelling than it is in the standard condition, particularly for observer TF. To the extent that it does not cycle regularly, performance tends towards veridical, rather than reversed, motion.

The pattern of results for a filter cut-off of 4 cycles degree^{-1} (figure 5) is similar in many respects. The cyclical variation in performance has a period of *ca.* 15 arcmin, again equal to one period of the filter cut-off frequency, and is not seen for the largest element size. In the reversed-contrast condition, the phase of the cyclical psychometric function is again inverted for small element sizes. But this time, for the largest element size there is no sign of reversal. Motion is consistently seen in the forward direction in the reversed-contrast condition as well as in the standard condition; neither reversed-phi nor cyclical reversal is seen. Looking back to figure 4, it can be seen that the results for large elements are intermediate between the non-cyclical forward motion seen in figure 5 and the cyclical performance seen in figure 3.

The results obtained with high-pass filtered images can be summarized as follows. For small element sizes performance cycles between forward and reversed motion as displacement increases, with a period equal to one period of the filter cut-off frequency. The cycle has the opposite phase (forward motion is seen in place of reversed motion and vice versa) when the contrast polarity is reversed between the two frames of the display. For large element sizes, particularly with high filter cut-offs, a quite different pattern of results is seen. Motion is almost always seen in the forward (veridical) direction and is typically unaffected by contrast reversal between frames. Importantly, the transition between the two types of behaviour is not sharp: intermediate element sizes show signs of both behaviours.

4. DISCUSSION

We interpret the data by postulating the involvement of two separate motion-detection mechanisms employing very different principles and operating in heavily overlapping ranges. Both types of mechanism have been

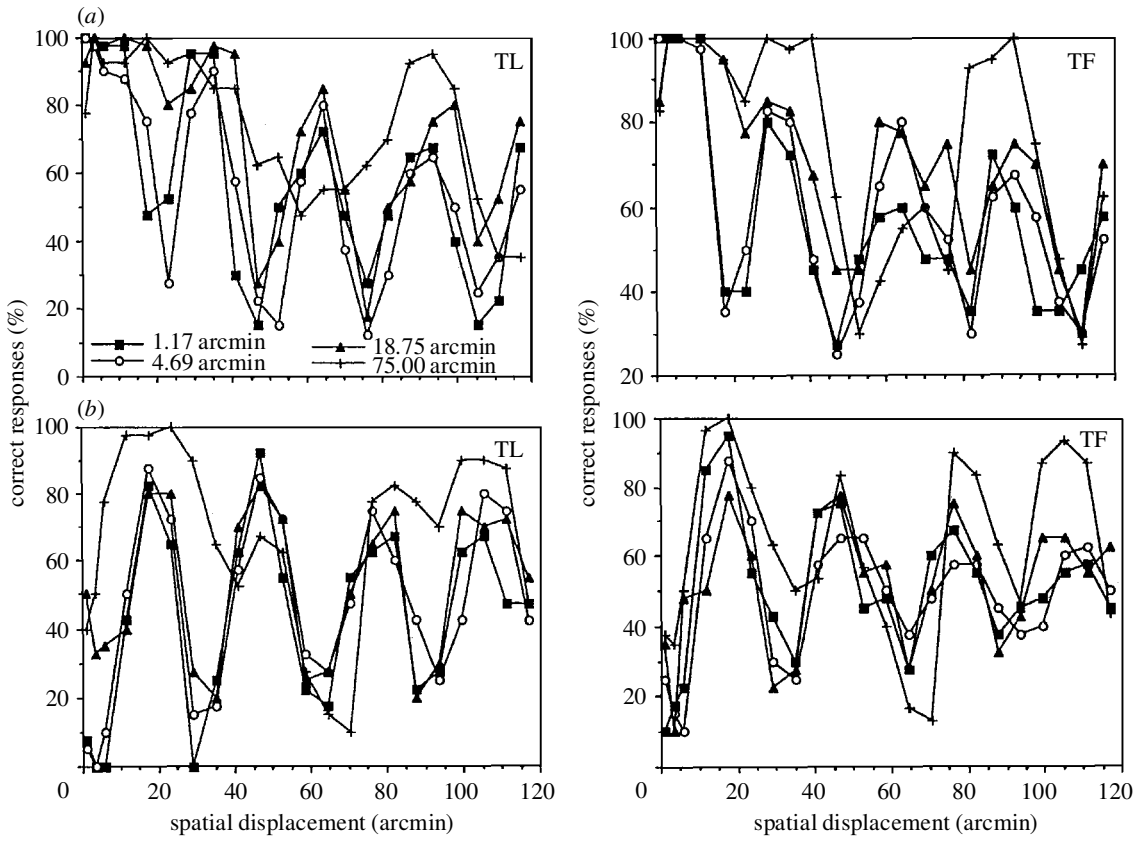

Figure 4. Results obtained using images that were high-pass spatial-frequency filtered using a filter cut-off of 2 cycles degree^{-1}, plotted in the same format as figure 2.

proposed before, but they are usually thought of as operating in different circumstances rather than in parallel.

(a) *Motion energy*

Over the range of element sizes and filter values for which cyclical variations in perceived direction are seen, our data are consistent with the notion that motion is detected by a system operating along the lines of the motion-energy mechanism of Adelson & Bergen (1985). It is generally assumed that there are multiple motion-energy detectors, which operate in different spatial ranges. Cyclical variation of performance with high-pass filtered images would be expected if detection were simply based on the lowest spatial-frequency component present in the filtered image. This component would be detected veridically for displacements up to half its spatial period, but would alias for displacements between 0.5 and 1.0 periods, giving a reversed perception of direction. Displacements between 1.0 and 1.5 periods would be indistinguishable from those of between 0 and 0.5 periods, and so on for larger displacements; hence the cyclical behaviour. Somewhat similar reversals of performance have been observed for bandpass-filtered images by Cleary & Braddick (1990). In the reversed-contrast condition, the psychometric functions are essentially inverted versions of those obtained in the standard condition in the range in which cyclical direction perception occurs, indicating that the perceived direction of motion is reversed by contrast reversal for all displacements. This behaviour, too, is expected if the motion-energy principle is used, since contrast reversal is equivalent to a 180° phase shift or adding 0.5 periods to the displacement. In this model, both types of direction reversal (cyclical and reversed-phi) have the same fundamental basis, both being determined by the relative phases of (local) spatial-frequency components in the two frames.

Although the above reasoning relates to filtered images, much of it also applies to the data obtained with broadband images (figure 2). But here the pattern of the results is rather different from that found with filtered images. In particular, although reversed-phi motion is seen in the reversed-contrast condition, there is no cyclical reversal of direction perception. Consider the notion, developed above, that the lowest spatial frequency in the image determines performance (in conditions where motion energy is used). In a broadband image, the lowest spatial frequency present is very low. We would, therefore, expect that aliasing of the lowest effective component would never occur, at least in the 2° range of displacements that we used. The lowest component used would need to be no lower than about 0.25 cycle degree^{-1} in order to avoid aliasing. It is, therefore, no surprise that cyclical reversal does not occur. However, if detection were based on local components with a spatial frequency of 0.25 cycle degree^{-1}, performance should be good over much of the range of displacements used. In fact, it is good only up to displacements of *ca.* 40 arcmin (for the smaller element sizes, where reliable reversed-phi motion occurs; see figure 2). This would suggest that detection is based on a component of *ca.* 0.8 cycle degree^{-1}, but if this were the case then aliasing would occur at a displacement

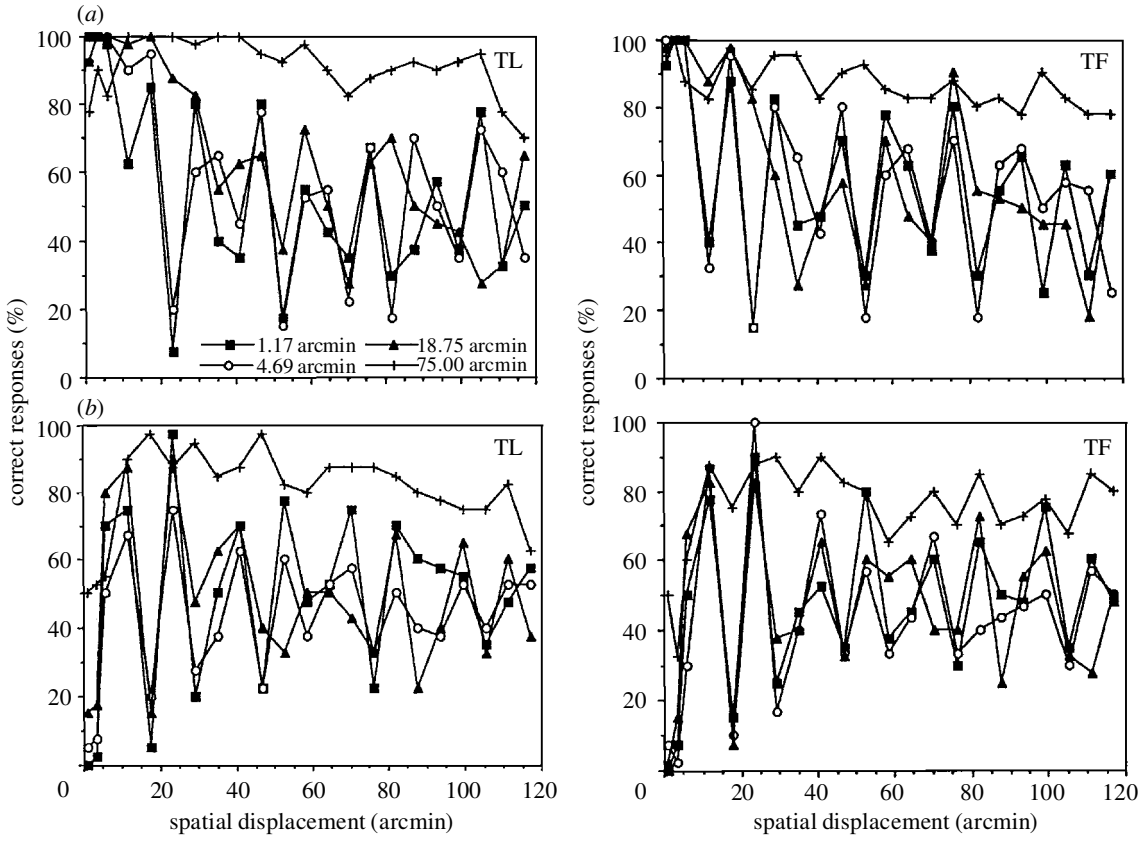

Figure 5. Results obtained using images that were high-pass spatial-frequency filtered using a filter cut-off of 4 cycles degree^{-1}, plotted in the same format as figure 2.

of around 80 arcmin, which is not the case. The various facts can be reconciled by postulating that performance is indeed based on a very low spatial frequency, but that the presence of higher frequencies in the image reduces performance, as previously shown by Cleary & Braddick (1990). There are several possible reasons for such a reduction in performance. One possible explanation, suggested by Cleary & Braddick (1990), is that detection is performed in a simple way by a low-spatial-frequency channel, but that this channel is inhibited by high-spatial-frequency channels, reducing its sensitivity. A second possible interpretation is that the perception of motion of a broadband image is determined not by the lowest active spatial-frequency channel but simply by integrating the various motion signals across spatial ranges. Some high-frequency components will alias, and the addition of these aliased signals could impair performance, even though the balance remains in favour of forward motion, so that aliasing is not perceived. A third interpretation is that the interference comes not from motion-energy mechanisms operating at high spatial frequencies but from the feature-matching mechanism (see §4b). Above a certain displacement, the correspondence problem will be too great for a feature-based mechanism, and so it will contribute only noise to the computation.

(b) *Matching of complex unsigned spatial features*

For larger element sizes, particularly but not exclusively when a high (e.g. 4 cycle degree^{-1}) filter cut-off is used, performance is very similar in the standard and reversed-contrast conditions. Reversing the contrast of the image has little or no effect, and neither compelling reversed-phi motion nor reliable cyclical reversed motion is seen. We suggest that in these conditions, motion is detected by a mechanism that identifies and labels complex spatial features formed by the random juxtaposition of pixels of the same polarity, and tracks their positions over time. The matching of such features predicts simply that motion will be seen veridically up to some limiting displacement and will fail at all displacements above this value. Cyclical performance variations are not expected because false correspondences, although common for individual square elements, will be rare for more complex features. The mechanism used is apparently indifferent to contrast polarity (i.e. it matches unsigned features), since reversed-phi motion does not occur either. Several demonstrations have been made of 'forward' motion despite polarity reversal (e.g. Anstis 1980; Mather *et al.* 1985). The indifference to spatial-frequency content seen in these conditions is consistent with a feature-tracking mechanism since the positions of complex features are little affected by spatial filtering.

(c) *Simultaneous involvement of both mechanisms*

Although we have provided clear evidence of two motion mechanisms that operate on images of the same class, there is no sharp demarcation of the stimulus conditions in which the two mechanisms are active. On the contrary, there is an extensive range of stimulus conditions over which both mechanisms appear to influence performance. For example, at the smallest displacements

performance is, in all cases, close to 100% in the standard condition, but in the reversed-contrast condition it fits neither the motion-energy result (0% correct) nor the feature-matching result (100% correct), but instead hovers uncertainly at *ca.* 30%. This is true for both broadband (figure 2) and filtered (figures 3–5) images. We interpret this as follows. In the standard condition, both motion energy and feature matching predict forward motion and so this is reliably perceived. In the reversed-contrast condition, motion energy predicts reversed-phi motion while the matching of (unsigned) features predicts forward motion, and so performance is intermediate. Performance is below 50%, suggesting a dominance of motion energy, but the fact that reversed motion is not reliably seen presumably reflects the influence of the feature-matching mechanism.

Similar examples of mixed influences are evident for larger displacements, over a range of element sizes and filter cut-offs. In the standard condition, although cyclical reversal occurs in many instances, totally reliable reversed motion (i.e. 0% correct) is never seen. This is presumably because the feature-matching mechanism is always arguing for forward motion even when motion energy is dominant and is signalling the incorrect direction. Another example may be seen in figure 3. In the standard condition, both subjects saw forward motion up to *ca.* 75 arcmin displacement for the largest element size, reflecting feature matching, but both also showed a clear dip in the function at *ca.* 40 arcmin displacement, reflecting the influence of motion energy, which predicts aliasing (motion reversal) at this point. Thus, the two mechanisms operate simultaneously and in substantially overlapping ranges.

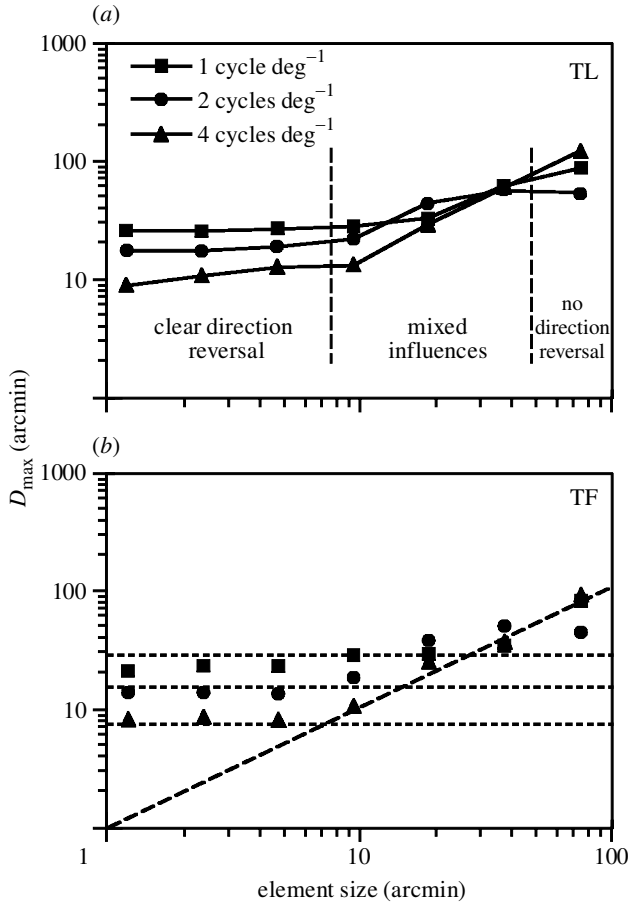

Figure 6. Values of D_{max} derived from the data shown in the preceding figures and from additional data obtained using other element sizes. Data for the two subjects are shown separately. D_{max} is invariant for element sizes up to *ca.* 10 arcmin and increases thereafter. The plots also include two other pieces of information. In part (*a*), labels indicate the conditions in which reversed-phi motion occurs in the reversed-contrast condition, those in which it did not and those in which clear signs of both behaviours were evident. In part (*b*), predictions of a simple model explaining the data are illustrated (see § 4e). The empirical data follow the envelope of the sensitivities of two distinct detection mechanisms.

(d) *Behaviour of D_{max}*

Figure 6 shows values of D_{max} derived from the data, to allow comparison of our results with those of other investigators, particularly Morgan (1992). The values of D_{max} shown reflect the displacements at which performance in the standard condition first fell to 75% correct, for each element size and filter cut-off. Figure 6*a* also indicates the conditions in which reversed-phi motion occurred in the reversed-contrast condition and those in which it did not. Because the transition between the two behaviours (as element size was varied) was not abrupt, a statistical procedure was used to categorize the data. For each combination of filter and element size, the data from the standard condition were correlated with those from the reversed-contrast condition. The correlation coefficients are shown in table 1. A strong positive correlation indicates that motion tended to be seen in the same direction irrespective of contrast polarity (suggesting matching of unsigned elements), while a strong negative correlation indicates that reversed-phi motion tended to be seen in the reversed-contrast condition (consistent with motion-energy detection). Strong negative correlations were found only on the flat portion of the curve, and strong positive correlations only on the rising portion, indicating that the two portions of the curve reflect the dominances of the two different mechanisms. Thus, the conditions in which cyclical performance variations occur coincide with those where D_{max} is invariant with element size, and those where they do not occur are coincident with those where D_{max} increases with element size. A sizeable intermediate category is evident, in which correlations (whether positive or negative) are weak, suggesting that both mechanisms are operating simultaneously. We conclude that the behaviour of D_{max} is better explained by two mechanisms with quite different properties than by a single mechanism of whatever type. Specifically, Morgan's (1992) interpretation of the function relating D_{max} to element size in terms of the activity of a single mechanism may be incorrect, since it cannot easily be reconciled with the fact that reversed-phi motion and cyclical reversal both occur in some conditions but not others.

(e) *Quantitative modelling*

Also shown in figure 6*b* are some simple quantitative predictions concerning the behaviour of D_{max} for filtered

Table 1. Coefficients showing the correlation between performance in the standard and the reversed-contrast conditions for each element size and each of the filter cut-offs used.

(The coefficients shown are the means of those obtained for the two observers.)

	element size (arcmin)						
	1.2	2.4	4.7	9.4	18.7	37.5	75
1 cycle degree^{-1}	−0.66	−0.27	−0.59	−0.14	−0.29	−0.05	+0.32
2 cycle degree^{-1}	−0.75	−0.64	−0.66	−0.14	−0.59	0.00	+0.30
4 cycle degree^{-1}	−0.68	−0.36	−0.66	−0.50	−0.05	−0.14	+0.91

images when performance is governed by the two mechanisms we suggest. In the case of motion energy, we assume that for filtered images motion is detected by a detector tuned to the lowest spatial frequency remaining in the image after high-pass filtering. If this is the case, then D_{max} has a theoretical maximum of half the period of that spatial frequency. For example, for a filter cut-off of 1 cycle degree^{-1}, a displacement of 0.5 will be ambiguous and performance will be at chance levels. Thus, D_{max} is expected to be just less than 0.5 (30 arcmin), assuming high efficiency. This value and the corresponding values for the other filters used are shown in figure 6*b* (dotted horizontal lines). In the case of feature matching we apply a related logic in order to predict performance. Consider a pattern like those we have used but with regular elements (i.e. a checkerboard). Based on a matching of signed features (in this case checks), direction will be detected correctly for displacements of less than the size of one element. When the displacement equals the element size, direction will be ambiguous because each check in the first image is equally likely to correspond to the check to the right or the left of it in the second image. If detection is based on unsigned features, as suggested by the lack of reversed-phi motion, the theoretical limit is half the size of an element. For noise patterns (as opposed to checkerboards), adjacent elements of the same polarity combine to yield larger features. The average one-dimensional extent of these features is two elements, doubling the theoretical maximum value of D_{max} to one element in the case of unsigned features. The function D_{max} = element size is shown in figure 6*b* (dashed line of unit slope). Importantly, the prediction is unaffected by high-pass filtering, which does not greatly affect the positions of features (see figure 1).

It can be seen from figure 6*b* that our empirical data are consistent with the supposition that D_{max} is determined by the more sensitive of the two mechanisms in every case. Each empirical function relating D_{max} to element size follows the envelope of the theoretical predictions of the two mechanisms for the relevant filter value. For small elements, motion energy yields higher D_{max} values than does feature matching, and so motion energy is used. As a result, the functions are flat and follow the filter-cut-off-dependent predictions for motion energy. For larger elements, feature matching yields a better performance than does motion energy, and so it is used. As a result, D_{max} increases with element size but is unaffected by high-pass filtering.

(f) *Role of second-order motion mechanisms*

Before invoking a motion mechanism that tracks the positions of complex spatial features, it is important to consider whether performance in the conditions where feature tracking might be invoked can be explained in terms of a second-order motion mechanism in which motion energy in the contrast, rather than the luminance, domain is detected (Chubb & Sperling 1988). The existence of such a mechanism is now widely accepted. In a random-dot kinematogram local contrast is high at the edges of the elements and low elsewhere, so the pattern contains variations in contrast whose motion might be detected. With small pattern elements, these contrast variations occur on a scale that is probably too fine to be useful; but as element size increases, they might become involved in the detection of motion in our experiments.

Two factors favour an explanation in terms of the use of contrast energy for large elements in our experiments. The first is that high-pass filtering in the luminance domain does not remove low-spatial-frequency components in the contrast domain. After filtering, the locations of the edges are essentially unchanged and contrast remains high at the edges and low elsewhere. Although we are normally much more sensitive to first-order than to second-order motion (e.g. Smith & Ledgeway 1997), this may not be the case after high-pass filtering in the luminance domain. In this case, low frequencies in the contrast domain could be used to avoid aliasing, giving second-order motion-energy mechanisms a marked advantage over first-order mechanisms. The second factor is that luminance polarity reversal between frames does not affect the polarity of contrast variations in the image. Again, contrast remains high at the element edges and low elsewhere after luminance reversal, and the locations of the edges are unaffected by contrast reversal. Use of second-order motion energy therefore predicts immunity to luminance contrast reversal, as observed for large element sizes (figures 3–5). Only if reversal occurs in the contrast domain itself (high contrast becomes low contrast and vice versa) will reversed-phi motion be seen with second-order motion (Nishida 1993).

However, one crucially important factor goes against an explanation in terms of second-order motion energy. Our stimuli were brief (75 ms per frame, 150 ms in total) and it has been shown (Derrington *et al.* 1993) that direction of motion of second-order patterns (beats) cannot be detected at durations of less than 200 ms, probably because of inferior temporal sensitivity. To test whether second-order motion could be used with our

own stimuli at short durations, we conducted further experiments in which the elements themselves were contrast-defined and there was thus no possibility of using first-order motion energy. The stimulus was fine two-dimensional noise divided into square elements in which the noise contrast was either high (*ca.* 90%) or zero; the mean luminance of the element was the same in both cases. The images were high-pass filtered at 1 cycle degree^{-1} in the contrast domain, and the same range of element sizes were used as in the main experiment. The two subjects failed to achieve performances significantly different from chance for any element size or step size. This makes it highly unlikely that second-order motion energy could have been used in our main experiment.

(g) *Relation to other studies*

Ledgeway & Hess (2000) have recently conducted a related experiment in which they measured the perceived direction of multiple Gabor patterns, defined either by luminance or by contrast modulation, and displaced with various step sizes. In some circumstances they found direction aliasing for both first-order and second-order patterns, consistent with detection of motion energy in the sine grating carrier. When motion energy was rendered unusable by random phase, spatial-frequency or orientation shifts, the perceived motion was consistent with feature tracking of the Gabor envelopes, again for both first-order and second-order patterns. This reinforces the distinction between second-order motion and feature tracking, but importantly it does not show that the two mechanisms operate simultaneously and influence perception conjointly.

The results of several other studies using two-frame random-dot kinematograms are consistent with our conclusion that motion-energy and feature-tracking mechanisms operate simultaneously in overlapping ranges. Morgan *et al.* (1997) found that removing low spatial frequencies reduces $D_{\max}$ for small element sizes but not for large ones. This is readily explained within our framework: removing low frequencies impairs a motion-energy system but has little effect on the locations of spatial features. Similarly, Morgan & Fahle (1992) and Eagle & Rogers (1996) have both reported that when dot density is reduced, causing an increase in the average dot spacing without markedly changing the spatial-frequency spectrum, there is an increase in $D_{\max}$ only for low dot densities. We suggest that for low densities features are tracked, but that for higher densities motion energy is increasingly used. Motion energy is used in the high-density range because it is efficient, whereas the feature-tracking system faces an overwhelming correspondence problem. At very low densities the correspondence problem is minimal but energy levels are low, so feature tracking is used because it is more efficient. Boulton & Baker (1993) have also argued for two different motion mechanisms at different stimulus densities, while Ito (1999) has advocated a similar distinction in the context of stereo-defined motion stimuli. Nishida & Sato (1992) measured the perceived direction of bandpass-filtered random-dot kinematograms, using various step sizes, and also the direction of after-effects of adaptation to such patterns. They found that direction was seen correctly at step sizes greater than 0.5 cycles of the lowest spatial frequency in the image, even though most of the motion energy is in the opposite direction (this was also found by Cleary & Braddick (1990)). This is consistent with the use of feature tracking (or second-order motion; they do not attempt to resolve this distinction). However, the motion after-effect behaved as expected on the basis of first-order motion energy: positive motion after-effects were seen at step sizes of 0.5–0.6 cycles. This suggests that motion-energy mechanisms, although they may fail to dominate perception, are nonetheless simultaneously active.

This work was funded by a grant from the Biotechnology and Biological Sciences Research Council to A.T.S.

REFERENCES

Adelson, E. H. & Bergen, J. R. 1985 Spatiotemporal energy models for the perception of motion. *J. Optical Soc. Am.* A **2**, 284–299.

Anstis, S. M. 1980 The perception of apparent movement. *Phil. Trans. R. Soc. Lond.* B **290**, 153–168.

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

Boulton, J. C. 1987 Two mechanisms for the detection of slow motion. *J. Optical Soc. Am.* A **4**, 1634–1642.

Boulton, J. C. & Baker, C. L. 1993 Different parameters control motion perception above and below a critical density. *Vis. Res.* **33**, 1803–1811.

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

Cavanagh, P. 1992 Attention-based motion perception. *Science* **257**, 1563–1565.

Cavanagh, P. & Mather, G. 1989 Motion: the long and short of it. *Spatial Vis.* **4**, 103–129.

Chang, J. J. & Julesz, B. 1983 Displacement limits for spatial frequency filtered random-dot cinematograms in apparent motion. *Vis. Res.* **23**, 1379–1385.

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

Cleary, R. F. & Braddick, O. J. 1990 Direction discrimination for band-pass filtered random dot kinematograms. *Vis. Res.* **30**, 303–316.

Derrington, A. M., Badcock, D. R. & Henning, G. B. 1993 Discriminating the direction of second-order motion at short stimulus durations. *Vis. Res.* **33**, 1785–1794.

Eagle, R. A. & Rogers, B. J. 1996 Motion detection is limited by element density not spatial frequency. *Vis. Res.* **36**, 545–558.

Exner, S. 1888 Einige beobachtungen über bewegungs-nachbilder. *Centralblatt für Physiologie* **1**, 135–140.

Fahle, M. & Poggio, T. 1981 Visual hyperacuity: spatiotemporal interpolation in human vision. *Proc. R. Soc. Lond.* B **213**, 451–477.

Georgeson, M. A. & Harris, M. G. 1990 The temporal range of motion sensing and motion perception. *Vis. Res.* **30**, 615–619.

Georgeson, M. A. & Shackleton, T. M. 1989 Monocular motion sensing, binocular motion perception. *Vis. Res.* **29**, 1511–1523.

Hammett, S. T., Ledgeway, T. & Smith, A. T. 1993 Transparent motion from feature- and luminance-based processes. *Vis. Res.* **33**, 1119–1122.

Ito, H. 1999 Two processes in stereoscopic apparent motion. *Vis. Res.* **39**, 2738–2784.

Lappin, J. S. & Bell, H. H. 1976 The detection of coherence in moving random dot patterns. *Vis. Res.* **16**, 161–168.

Ledgeway, T. & Hess, R. 2000 The properties of the motion-detecting mechanisms mediating perceived direction in stochastic displays. *Vis. Res.* **40**, 3585–3597.

Mather, G., Cavanagh, P. & Anstis, S. 1985 A moving display which opposes short-range and long-range signals. *Perception* **14**, 163–166.

Morgan, M. J. 1992 Spatial filtering precedes motion detection. *Nature* **355**, 344–346.

Morgan, M. J. & Fahle, M. 1992 Effects of pattern element density upon displacement limits for motion detection in random binary luminance patterns. *Proc. R. Soc. Lond.* B **248**, 189–198.

Morgan, M. J., Perry, R. & Fahle, M. 1997 The spatial limit for motion detection in noise depends on element size, not on spatial frequency. *Vis. Res.* **37**, 729–736.

Nishida, S. 1993 Spatiotemporal properties of motion perception for random-check contrast modulations. *Vis. Res.* **33**, 633–646.

Nishida, S. & Sato, T. 1992 Positive motion after-effect induced by bandpass-filtered random-dot kinematograms. *Vis. Res.* **32**, 1635–1646.

Sato, T. 1990 Effects of dot size and dot density on motion perception with random dot kinematograms. *Perception*, **19**(Suppl.), 329.

Smith, A. T. 1994 Correspondence-based and energy-based detection of second-order motion in human vision. *J. Optical Soc. Am.* A **11**, 1940–1948.

Smith, A. T. & Ledgeway, T. 1997 Separate detection of moving luminance and contrast modulations: fact or artifact? *Vis. Res.* **37**, 45–62.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.