



The Effects of Spatiotemporal Integration on Maximum Displacement Thresholds for the Detection of Coherent Motion

JAMES T. TODD,*† J. FARLEY NORMAN*

Received 25 May 1994; in revised form 8 November 1994; in final form 2 December 1994

In a series of nine experiments, observers were required to identify the shapes of moving targets, and to discriminate regions of motion from regions of uncorrelated noise. Maximum displacement thresholds (D_{\max}) for performing these tasks were obtained under a wide variety of conditions. The stimulus parameters manipulated included the number of distinct frames in the motion sequences, the stimulus onset asynchrony between each frame, the size of the moving dots, and the shape, area and eccentricity of the target regions. For two-frame displays presented in alternation, the area of the target region was the only one of these variables to have any significant effect on D_{\max} . For longer length sequences, in contrast, D_{\max} varied dramatically among the different conditions over a range of 10 min arc to 10 deg. In an effort to isolate the specific processes of spatiotemporal integration, we also examined how performance is affected by having overlapping transparent motions in opposite directions, or by the presence of dynamic noise or limited dot lifetimes within the moving target regions. The overall pattern of results suggest that D_{\max} is primarily determined by the ability of the visual system to isolate motion signals from the noise produced by spurious false target correlations. As a general rule, D_{\max} will increase as a result of any stimulus manipulation that increases the number of local signal correlations detected relative to those arising from noise, and vice versa.

Motion perception Apparent motion Spatiotemporal integration Displacement thresholds

Most visually sensitive organisms, including humans, are especially attuned to changes in optical structure over time produced by their own actions or the motions of other objects. This sensitivity is highly adaptive. The visual information provided by motion is useful for a wide variety of functions, including the segregation of figure from ground, the control of locomotion and posture, and the visual specification of three-dimensional form. It should not be surprising, therefore, that biological organisms have evolved specialized mechanisms for exploiting this information.

The first specific model of how visual motion detectors might actually be constructed was the bilocal correlator proposed by Reichardt (1957, 1961). Since then numerous variations of this model have proliferated (e.g. Adelson & Bergen, 1985; Barlow & Levick, 1965; van Santen & Sperling, 1984, 1985; Watson & Ahumada, 1983), but they are all based on a similar delay and compare principle [see van Santen and Sperling (1985) for a detailed analysis of how these different implementations are formally related]. The minimal basic structure of a bilocal correlator is shown schematically in

Fig. 1(a). It consists of two spatially separated input channels with different delay times, whose outputs are compared through multiplication and temporal averaging. Note that the velocity tuning of this detector is determined by two parameters: The span σ between its front-end spatial filters, and the difference δ in their delay times—i.e. the detector would respond most vigorously to patterns moving at a velocity σ/δ .

It is important to recognize that a bilocal correlator such as the one shown in Fig. 1(a) can be affected by other stimulus properties in addition to velocity. For example, if the optical pattern undergoing motion does not contain components that match the tuning characteristics of the front-end spatial filters, then the detector would be unable to respond to any velocity. Conversely, there are other possible sources of stimulation, such as patterns flickering at a rate of $1/\delta$, for which the detector would respond vigorously even in the absence of motion. To prevent this latter type of response from occurring, Reichardt (1957, 1961) suggested a slightly more complex design, which we shall refer to as an opponent correlator [see Fig. 1(b)]. It consists of two nonopponent subunits tuned to motion in opposite directions, whose outputs are subtracted. Note that both of these designs have their own advantages and limitations. An opponent correlator will not respond spuriously to flickering patterns, but it will also not respond to overlapping

*Department of Psychology, 142 Townshend Hall, The Ohio State University, Columbus, OH 43210, U.S.A.

†The Center for Cognitive Science, The Ohio State University, Columbus, OH 43210, U.S.A.

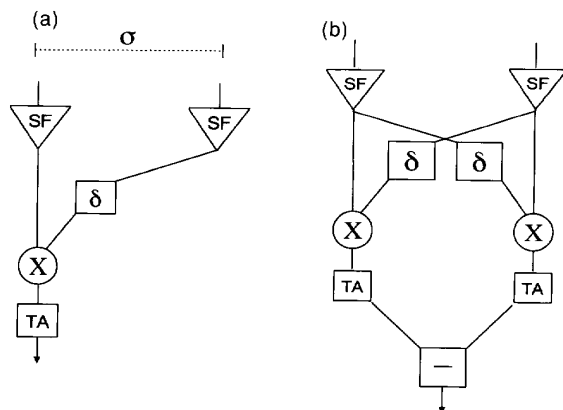


FIGURE 1. A schematic model of a bilocal correlator adapted from Reichardt (1957, 1961). The configuration in (a) contains two spatial filters separated by a span σ , and a differential time delay δ , whose outputs are multiplied and temporally averaged. The configuration in (b) is an opponent correlator in which the outputs of two subunits tuned to motion in opposite directions are subtracted from one another. The advantage of an opponent correlator is that it does not respond to flicker.

transparent motions in opposite directions. A nonopponent correlator will respond spuriously to flickering patterns, but it is also able to cope with motion transparency (see van Doorn & Koenderink, 1982a, b; van de Grind, van Doorn & Koenderink, 1986).

Because bilocal correlators have localized receptive fields and are selective to the spatial structure of a pattern as well as its velocity, an effective motion detection system would have to contain a large population of correlators with different tuning characteristics in order to be sensitive to a reasonably broad range of possible environmental events. In an ideal system, multiple detectors would be assigned to each location within the visual field. Detectors tuned to higher velocities would have relatively long spans and relatively short delay times, while those tuned to slower velocities would have relatively short spans and relatively long delay times.

There have been numerous attempts to estimate the range of these parameters within the human visual system using psychophysical techniques. For example, Braddick and his colleagues at Cambridge University (e.g. Baker & Braddick, 1982a, b, 1985a, b; Braddick, 1974, 1980; Cleary & Braddick 1990a, b; Snowden & Braddick, 1989a, b, 1990), have attempted to estimate the spans and delay times of low level visual motion detectors from the acceptable values of displacement and interstimulus interval (ISI) respectively, with which observers can detect the apparent motion of a complex noise pattern. Their data suggest that motion detectors throughout the visual field have a constant delay time of approx. 40 msec. The estimated span values, in contrast, vary with eccentricity from a maximum value of approx. 15 min arc near the fovea to just under 2 deg for patterns presented 10 deg in the periphery.

A similar series of experiments with dramatically different results has been performed by researchers at the University of Utrecht (see van Doorn & Koenderink,

1982a, b, c, 1983, 1984; van Doorn, Koenderink & van de Grind, 1985; Fredericksen, Verstraten & van de Grind, 1993, 1994a, b, c; van de Grind, van Doorn & Koenderink, 1983, 1987; van de Grind *et al.*, 1986; van de Grind, Koenderink & van Doorn, 1992; van de Grind, Koenderink, van Doorn, Milders & Voerman, 1993; Koenderink, van Doorn & van de Grind, 1985). They estimate the spans and delays of bilocal correlators by measuring the threshold signal-to-noise ratios for detecting continuous motions of complex noise patterns at different velocities within viewing windows of variable size and shape. Their results suggest that if displacements are scaled by the cortical magnification factor, then the properties of motion detectors are remarkably homogeneous throughout the visual field. For detectors tuned to slow velocities the estimated delay times range 60–2000 msec, but they all have similar estimated spans of approximately 16 cortical receptive field widths. For detectors tuned to high velocities, in contrast, the estimated spans range from 10 min arc to 16 deg arc, but they all have similar estimated delay times of approx. 50 msec.

What possible explanation could account for why two such respected laboratories could differ by more than an order of magnitude in estimating the ranges of δ and σ within the human visual system? We believe that this discrepancy may be due to differences in the stimulus displays employed by each group, which varied dramatically in the amount of redundant information available to be pooled over multiple motion detectors. It is important to keep in mind that bilocal correlators do not respond exclusively to motion. They can also respond to spurious correlations between different elements of a moving or scintillating pattern that happen by chance to align appropriately in space and time. Because an individual detector cannot distinguish these spurious correlations from actual motion, a reliable segregation of signal from noise can only be achieved by pooling information over a sufficiently large population of detectors. Moreover, if there were systematic variations in the number of neurons tuned to different velocities, then some velocities could require a greater amount of redundant information than others in order to be perceived as coherent motion.

The pooling of information over multiple motion detectors to separate signal from noise could be achieved in several different ways. One possibility, first described by Lappin and Bell (1976), involves summing the outputs of an entire population of bilocal correlators tuned to a given velocity (see also van Doorn & Koenderink, 1982a, 1984; Fredericksen *et al.*, 1993, 1994c; van de Grind *et al.*, 1992). This sum is then compared with what would be expected from a stimulus composed of uncorrelated dynamic noise to determine if they are significantly different. The sensitivity of such a process would increase proportionally as the square root of the number of pooled correlators, which is generally consistent with the available psychophysical data. Increasing the number of elements in a moving pattern decreases signal-to-noise thresholds (van Doorn & Koenderink, 1982b, 1984;

Fredericksen *et al.*, 1993, 1994c; Koenderink *et al.*, 1985) and increases displacement thresholds (Lappin & Bell, 1976) as predicted by the model, though the interpretation of these findings is complicated by the fact that large area patches must necessarily stimulate more peripheral regions of the retina that may have greater sensitivity to higher velocities of motion (see Baker & Braddick, 1982a, 1985a).

It is also possible to pool information over time. This could be achieved quite easily by modifying the Lappin and Bell model to summate the number of local correlations over an arbitrarily extended period of time, as suggested by van Doorn and Koenderink (1982a, 1984), or, perhaps more plausibly, by filtering the outputs of many motion detectors through a leaky integrator (Fredericksen *et al.*, 1994a, b, c). Another possibility suggested by Snowden and Braddick (1989a) is to connect a chain of local correlators parallel to the preferred direction of motion, such that the stimulation of any given subunit produces a facilitative priming effect on the next correlator in the chain. Still other variations of temporal pooling could involve higher-order correlations of motions or features over an arbitrary number of time intervals. There is considerable evidence that the human visual system is capable of employing some form of temporal pooling or recruitment in the detection of coherent motion (e.g. see van Doorn & Koenderink, 1982a, 1984; McKee & Welch, 1985; Nakayama & Silverman, 1984; Snowden & Braddick, 1989a, b, 1990; Watamaniuk, Sekuler & Williams, 1989; Williams & Sekuler, 1984), but the evidence is somewhat contradictory concerning the overall temporal extent over which this pooling can occur. For example, Snowden and Braddick (1989a) report that displacement thresholds for multiple frame apparent motion sequences asymptote after just a few hundred milliseconds, yet van Doorn and Koenderink (1984) have found that signal-to-noise thresholds continue to improve for presentation periods as long as 1 or 2 sec.

In an effort to resolve some of the apparently conflicting findings on these topics that have been reported in the literature, the research described in the present article was designed to investigate how the processes of spatio-temporal integration are influenced by a wide variety of stimulus variables. Our basic strategy was to adopt the D_{\max} paradigm developed by Braddick, but with a broad range of stimulus parameters similar to those used by van de Grind, van Doorn and Koenderink.

EXPERIMENT 1

Methods

The experiment was controlled by a Masscomp 5600 computer with a Lexidata Lex-90 graphics system. The stimuli were presented within a 33 × 26 cm display screen with a spatial resolution of 1280 × 1024 pixels and a raster refresh rate of 60 Hz. The displays were viewed through a monocular viewing hood at a distance of 76 cm, such that each pixel spanned a visual angle of

approx. 1.15 min arc. All responses were recorded by pressing keys on the computer keyboard.

Observers were presented with apparent motion sequences composed of 2–16 distinct frames of a small random check rectangle moving coherently against a background of uncorrelated noise. Their task on each trial was to indicate whether the moving rectangle had a horizontal or vertical orientation.

The background pattern was composed of a 320 × 256 grid of checks that filled the display screen (i.e., each check spanned a visual angle of 4.6 min arc). The structure of this grid was determined at random on every frame, such that each check had a 50% probability of being black or white. The only temporally stable feature of this background pattern was a small 9.2 min arc red square at its center that served as a fixation point. The moving rectangle was composed of a randomly structured 20 × 40 grid of checks, whose long axis could be oriented either horizontally or vertically with equal probability. This rectangle could appear either above or below the fixation point, such that its closest edge had a vertical eccentricity of 36.8 min arc.

Within a given apparent motion sequence the rectangular figure was shifted horizontally in discrete jumps with a fixed frame-to-frame displacement that varied over trials, and a fixed stimulus onset asynchrony (SOA) of 50 msec (i.e. 3 raster refresh cycles). Within a given block of trials, there were five different frame-to-frame displacements presented 100 times each in a random order. The number of distinct frames in the apparent motion sequences varied across blocks in a random order from possible sequence lengths of 2, 4, 8 or 16 frames. After each sequence was completed it was repeated again in a continuous cycle until the observer made an appropriate response. The rectangles could move either rightward or leftward with equal probability. In most cases their trajectories were centered horizontally in the display screen, such that the maximum target eccentricity varied with sequence length. However, in an effort to disentangle these variables, we also included an additional two-frame condition at the maximum horizontal eccentricity for the 16-frame displays (i.e. 5.5 deg).

In pilot experiments with this general task it became immediately obvious that the maximum displacement thresholds for segregating the moving figure from the background noise varied dramatically with the number of distinct frames in each sequence. To avoid ceiling and floor effects it was therefore necessary to use a different set of possible displacements for each sequence length in order to obtain a reasonable measure of the 75% threshold. Based on our pilot observations, we used displacements of 4.6, 9.2, 13.8, 18.4, and 23.0 min arc for the two-frame sequences; 13.8, 18.4, 23.0, 27.6, and 32.2 min arc for the 4-frame sequences; 23.0, 27.6, 32.2, 36.8, and 41.4 min arc for the 8-frame sequences; and 32.2, 36.8, 41.4, 46.0, and 50.6 min arc for the 16-frame sequences.

Two observers participated in the experiment including the first author (JT) and a paid graduate student

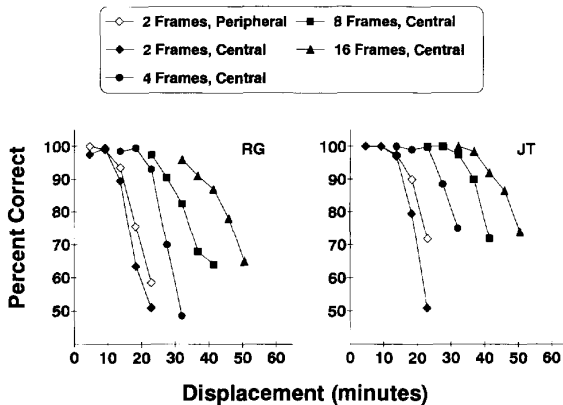


FIGURE 2. The percentage of correct responses as a function of target displacement for each observer in Expt 1. The different psychometric functions were obtained for apparent motion sequences that varied in length and retinal eccentricity.

volunteer (RG). Both observers performed two blocks of trials for each of the five sequence length conditions (16-frames, 8-frames, 4-frames, 2-frames central, and 2-frames peripheral), resulting in a total of 200 judgments per displacement in each condition.

Results

The individual psychophysical functions for both observers are shown in Fig. 2. It is clear from these data that the number of distinct frames in the apparent motion sequences had a large effect on the observers' abilities to distinguish the moving patterns from the uncorrelated background noise. The maximum displacement thresholds (D_{max}) increased from approx. 15 min arc for the 2-frame displays to over 45 min arc for the 16-frame displays.

One potentially relevant difference between these conditions is that the maximum retinal eccentricity of the moving targets increased proportionately with the length of the apparent motion sequences. Note, however, that this cannot by itself account for the changes in D_{max} . The 2-frame displays were presented at 0 and 5.5 deg retinal eccentricities, but this had no detectable effect on the observers' thresholds. This result has been reported previously by Chang and Julesz (1983a), though Baker and Braddick (1985a) have shown that D_{max} does increase with retinal eccentricity if the target size is adjusted appropriately to compensate for the cortical magnification factor. It would appear from these findings that some minimal number of bilocal correlators must be pooled over space and/or time in order to trigger the perception of coherent motion, and that the required amount of pooling may increase with target displacement or velocity.

It is interesting to compare the results shown in Fig. 2 with those obtained by Snowden and Braddick (1989a) in a closely related series of experiments on the effects of temporal integration. Snowden and Braddick found that they could increase D_{max} by approximately 50% by increasing the number of frames in an apparent motion

sequence, but that this effect leveled off after about four or five frames. Note that this contrasts sharply with the 300% increases in D_{max} obtained in the present experiment, which did not level off even after 16 frames. On the basis of these contrasting results we hypothesized that the effectiveness of temporal integration may be affected by other stimulus factors, and we set out to investigate some of the more likely candidates. In Expt 2, for example, we attempted to determine the optimal SOA between each successive frame over an extended sequence.

EXPERIMENT 2

Methods

The apparatus and procedure were identical to those used in Expt 1. Within each block of trials, the target displacement between successive frames of a motion sequence was held constant, while the SOA was varied across trials from possible values of 16.7, 50.0, 83.3, 116.7, and 150.0 msec presented 100 times each in a random order. The number of distinct frames in the apparent motion sequences varied across blocks in a random order from possible sequence lengths of 2, 4, 8, or 16 frames. The target displacement for each block was selected to achieve an accuracy of approximately 90% at a 50 msec SOA based on the results of Expt 1. Thus, the displacements used were 13.8, 18.4, 27.6, and 36.8 min arc, respectively, for the 2-, 4-, 8-, and 16-frame displays. The same two observers who participated in Expt 1 ran two blocks for each condition, resulting in a total of 200 trials per SOA for each sequence.

Results

The individual results for both observers are shown in Fig. 3. These data show clearly that the effect of varying SOA interacts strongly with the length of an apparent motion sequence. For 2-frame displays performance remained approximately constant at all SOAs except for the degenerate 16.7 msec condition. With a 16.7 msec SOA, the individual frames of a 2-frame motion sequence were alternated back and forth at 60 Hz. Because this is above the threshold for flicker fusion, they appeared perceptually to be temporally unchanging, and above chance performance could only be achieved by detecting spatially correlated energy in the fused composite image. A very different pattern of results emerged, however, as the number of distinct frames in the apparent motion sequences was increased beyond two. The increased displacement thresholds for these longer sequences that were obtained in Expt 1 could only be achieved for SOAs of approx. 50 msec, and performance deteriorated rapidly as the display timing was increased or decreased relative to this optimal value.

These findings are quite consistent with an earlier series of experiments by van de Grind *et al.* (1986), in which signal-to-noise thresholds obtained under varying conditions were used to estimate the spans and delays of

bilocal correlators within the human visual system. It is important to keep in mind that the velocity tuning of a bilocal correlator is determined by the ratio of its span and delay. Based on their results from several converging paradigms, van de Grind *et al.* argued that neurons tuned to low velocities have a wide variety of delays, but that they all have approximately the same span when scaled by the cortical magnification factor. Neurons tuned to higher velocities, in contrast, can have a wide variety of spans, but they all have similar delays of approx. 50 msec. For a population of correlators with these particular properties, the optimal SOA for detecting a moving pattern would become more tightly tuned around 50 msec as velocity is increased—just as we observed in the present experiment (see also the similar results of Fredericksen *et al.*, 1993, 1994a, b).

There is, however, another previously reported experiment by Snowden and Braddick (1989a) that does not appear to support this conclusion. They measured D_{max} for discriminating directions of motion using sequences composed of 2–8 displacements with SOAs ranging from 20 to 100 msec. The results revealed that for each possible sequence length, performance remained approximately constant for all SOAs except for a slight attenuation of D_{max} in the 20 msec condition.

We were again perplexed by why our results should be so different from those obtained by Snowden and Braddick (1989a). There were several differences between the methods employed in these studies that could potentially have influenced the observers' performance. Snowden and Braddick used a direction discrimination task, in which sparse arrays of random dots were viewed within a fixed aperture for limited periods of time (i.e. display

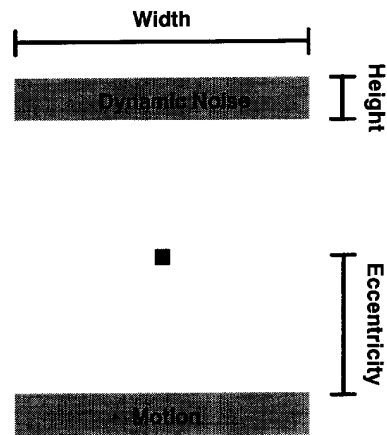


FIGURE 4. The basic stimulus configuration employed in Expts 3–9 included two rectangular apertures whose height, width and retinal eccentricity were varied. One of the apertures contained a pattern of coherent motion, while the other contained temporally uncorrelated noise.

durations were all less than 1 sec). Our experiments, in contrast, involved a form discrimination task, in which densely textured objects moving over temporally uncorrelated backgrounds were viewed for unlimited periods of time. In an effort to determine how various task or display differences could influence the spatiotemporal integration of local motion measures, we set out to discover if we could replicate either of these reported patterns of results using a different experimental paradigm.

EXPERIMENT 3

Methods

The basic stimulus configuration for this experiment is shown in Fig. 4. Observers were instructed to fixate on a small 9.2 min arc square at the center of the display screen. There were two rectangular apertures at equal distances above and below this fixation point, one of which contained a pattern of random checks undergoing coherent motion, while the other contained a similar pattern of random checks that was temporally uncorrelated. The observers' task on each trial was to identify which of the two apertures contained the moving pattern.

As in Expts 1 and 2, the displays were observed through a monocular viewing hood at a distance of 76 cm. The rectangular apertures all had a horizontal width of 24.5 deg and a vertical height of 9.2 min arc of visual angle. Each aperture contained a 320 × 2 grid of checks, each of which had a 50% probability of being black or white on any given frame of a motion or noise sequence (i.e. each check spanned a visual angle of 4.6 min arc). The vertical eccentricity from the fixation point to the closer edge of each aperture was held constant within each block of trials, but was varied across trials with two possible values of 1.2 or 8.6 deg.

On each trial, one of the two apertures was selected at

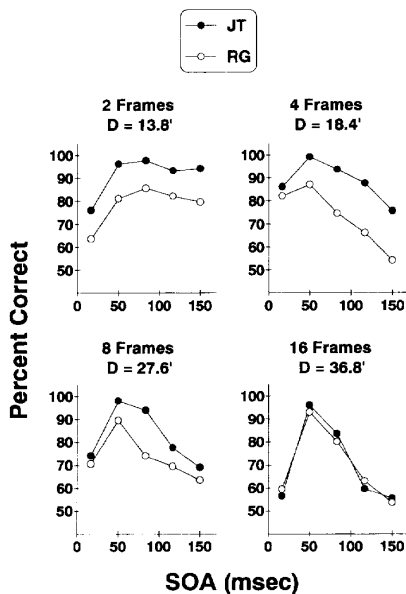


FIGURE 3. The percentage of correct responses as a function of SOA for each observer in Expt 2. The different psychometric functions were obtained for apparent motion sequences that varied in length and target displacement.

random to contain a pattern of coherent motion, whose direction was also selected at random to be either rightward or leftward. A motion or noise sequence could contain either 2 or 16 distinct frames, which varied across blocks. For the noise sequences these distinct frames were all uncorrelated with one another, while for the motion sequences the patterns were shifted horizontally in discrete jumps that varied adaptively from trial to trial based on the accuracy of the observer's responses. After each sequence was completed it was repeated again in a continuous cycle until the observer made an appropriate response. The SOA between each frame was held constant within a given block of trials, but was varied across blocks in a random order from possible values of 16.7, 50, 100, 200 and 400 msec.

The adaptive staircase procedure was adopted from Levitt (1970). Starting with an initial displacement of 4.6 min arc and an initial increment of 18.4 min arc, the displacements were increased by one increment after each pair of consecutive correct responses and decreased by one increment after each incorrect response. The increment was decreased by half after the first, third and seventh reversals in the direction of this staircase, and the block was terminated after 10 reversals. This procedure converges at a threshold level of 70.7% (i.e. when the probability of getting two consecutive correct responses is 0.5).

The two authors participated as observers. They each ran five staircases for each of the 20 possible display conditions (2 sequence lengths \times 2 eccentricities \times 5 SOAs), and the maximum displacement threshold (D_{max}) was estimated for each condition by the average of all 50 reversal points.

Results

The individual results for each observer are shown in Fig. 5. The most striking aspect of these data is the overall magnitude of the observers' maximum displacement thresholds. D_{max} was approx. 1 deg for the 2-frame displays, and it increased to almost 6 deg for the 16-frame displays under optimal conditions of eccen-

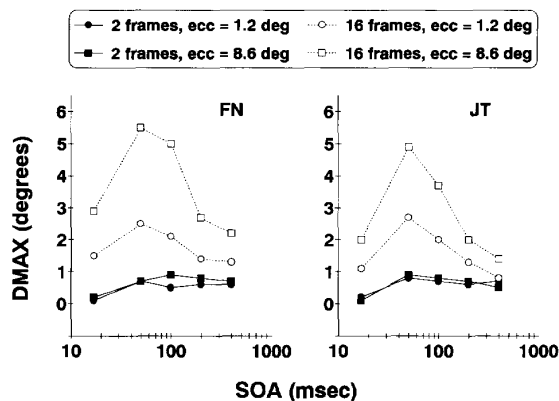


FIGURE 5. The maximum displacement thresholds for the two observers in Expt 3 as a function of SOA for apparent motion sequences that varied in length and retinal eccentricity.

tricity and SOA. With a frame-to-frame displacement of 5 deg and SOA of 50 msec the patterns moved at a rate of 100 deg/sec. At these threshold velocities the patterns appeared phenomenally to be primarily composed of uncorrelated noise, but over the course of several seconds there would be brief intervals in which there was a clear sense of some vaguely defined structure out in the periphery shooting past at a very high rate of speed.

The ability of observers to detect any motion at all under these conditions suggests that the human visual system must contain at least some motion detectors with spans of 5 deg or more. Note, however, that the observers' sensitivity to such large spatial displacements is critically dependent on several other display parameters. Our results indicate that these high velocity neurons all have delays of approx. 50 msec; that they are only sensitive to patterns moving in peripheral portions of the visual field; and that they require a considerable amount of temporal integration in order to trigger a phenomenal awareness of motion. All of these conclusions are also supported by the findings of van Doorn and Koenderink (1982b, 1984), Fredericksen *et al.* (1993, 1994a, b, c), and van de Grind *et al.* (1983, 1986, 1992, 1993) using very different procedures.

EXPERIMENT 4

The results of Expt 3 provide strong evidence that the effects of spatiotemporal integration on D_{max} increase with retinal eccentricity, but they do not reveal the precise functional pattern of this increase. Experiment 4 was designed therefore to provide a more detailed sample of how these integration effects change between retinal eccentricities of 1 and 9 deg.

Methods

The apparatus and procedure were identical to those used in Expt 3. The retinal eccentricities of the rectangular apertures containing the signal and noise patterns were varied across blocks from possible values of 1.2, 3.7, 6.1, and 8.6 deg. The sequence length was also varied across blocks between 2 and 16 distinct frames. The SOA for all displays was 50 msec. Both authors again participated as observers. They each ran five staircases for each of the eight possible display conditions (2 sequence lengths \times 4 eccentricities).

Results

The individual results for both observers are shown in Fig. 6. It is clear from these data that there is a strong interaction between retinal eccentricity and sequence length. For 2-frame displays presented in oscillation, the variations in retinal eccentricity had no effect whatsoever on the observers' maximum displacement thresholds. For 16-frame displays, in contrast, D_{max} increased linearly from 2.5 to 5.5 deg as the retinal eccentricity of the patterns was increased from 1.2 to 8.6 deg.

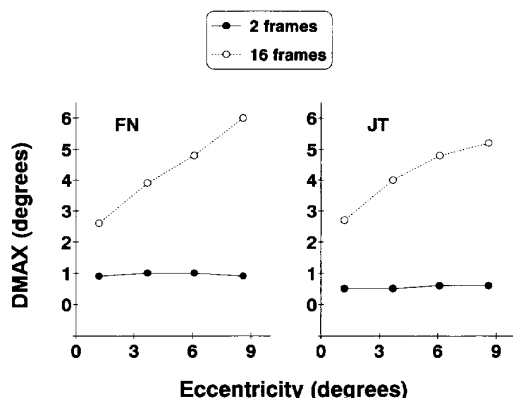


FIGURE 6. The maximum displacement thresholds for the two observers in Expt 4 as a function of retinal eccentricity for apparent motion sequences that varied in length.

EXPERIMENT 5

Experiment 5 was designed to provide a more detailed sample of how D_{max} increases with sequence length at both central and peripheral retinal eccentricities.

Methods

The apparatus and procedure were identical to those used in Expts 3 and 4. The number of distinct frames in the motion and noise sequences were varied across blocks from possible values of 2, 4, 8 and 16. The retinal eccentricity of the patterns was also varied across blocks between 1.2 and 8.6 deg. The SOA for all displays was 50 msec. Both authors again participated as observers. They each ran five staircases for each of the eight possible display conditions (4 sequence lengths \times 2 eccentricities).

Results

The individual results for both observers are shown in Fig. 7. It is interesting to note from these data that D_{max} increased continuously as the number of distinct frames

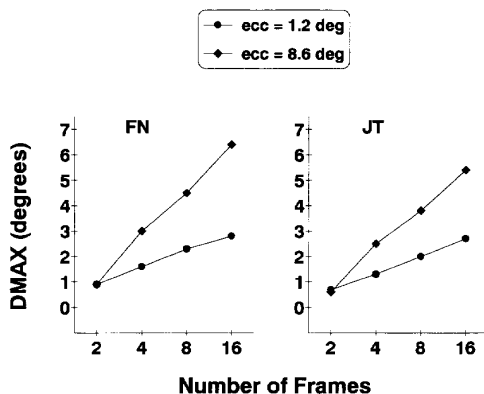


FIGURE 7. The maximum displacement thresholds for the two observers in Expt 5 as a function of sequence length for moving patterns at different retinal eccentricities.

in a motion sequence was increased from 2 to 16. Since all of these displays had an SOA of 50 msec, it follows from this result that local motion measures can be temporally integrated over relatively long durations of at least 750 msec. Note in addition that the functions shown in Fig. 7 do not appear to have reached asymptotic values for the 16-frame sequences, so it is quite possible that temporal integration can occur over even longer intervals. Indeed, van Doorn and Koenderink (1982b, 1984) and Fredericksen *et al.* (1994a, b) have shown that signal-to-noise thresholds for detecting motion continue to drop with display duration for periods of more than 1 sec.

EXPERIMENT 6

Experiment 6 was designed to investigate how the effects of spatiotemporal integration on D_{max} are influenced by changes in area of a moving pattern and the sizes of the individual checks of which it is composed.

Methods

The apparatus and procedure were identical to those used in Expt 3. As in that earlier experiment, the rectangular apertures all had a horizontal width of 24.5 deg. In this study, however, the vertical height of the apertures was varied across blocks between possible values of 18.4 min arc and 8.3 deg—i.e. their resulting areas were 7.5 and 203 deg². These different sized apertures were positioned as shown in Fig. 8 so that they each had a maximum vertical eccentricity of 9.2 deg. The number of distinct frames in each sequence was also varied across blocks between 2 and 16, and so were the sizes of the individual checks within each pattern, which had possible widths of 4.6 and 18.4 min arc. The SOA for all displays was 50 msec. Both authors again participated as observers. They each ran five staircases for each of the eight possible display conditions (2 sequence lengths \times 2 areas \times 2 check sizes).

Results

The individual results for both observers are shown in Fig. 9. The thresholds obtained for the 2-frame



FIGURE 8. Aperture area. The basic stimulus configuration employed in Expt 6 included pairs of apertures whose areas could vary from trial to trial, but whose maximum retinal eccentricities were all identical.

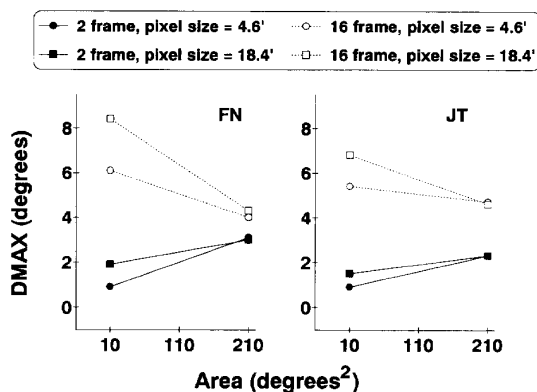


FIGURE 9. The maximum displacement thresholds for the two observers in Expt 6 as a function of stimulus area for moving patterns with different check sizes and sequence lengths.

conditions are similar to those reported previously by Baker and Braddick (1982, 1985a). They too found D_{\max} to be approx. 2 deg for large area patterns with maximum retinal eccentricities of 10 deg, and that these thresholds were reduced if the patterns did not cover a sufficient area. Note, however, that the 16-frame conditions produced a very different pattern of results. Reducing the area of those displays resulted in an increase in D_{\max} .

Although this may appear at first blush to be a surprising outcome, it can perhaps be explained by the visual system's inhomogeneous sensitivity to velocity in different regions of the visual field (e.g. see van de Grind *et al.*, 1983, 1986, 1993; Koenderink *et al.*, 1985). The

available evidence suggests that motion detectors with the largest possible spans all have receptive fields in more peripheral regions of the retina. Since observers in the present experiment could detect frame-to-frame displacements over 7 deg for small area 16-frame displays presented peripherally, it is reasonable to conclude that there must exist some bilocal correlators with spans of at least 7 deg that are sensitive to motion in those regions. Now suppose that the area of one of these high velocity patterns were increased to stimulate additional motion sensitive neurons with receptive fields in more central regions of the retina. If these neurons had sufficiently large spans to detect 7 deg displacements, then this increase in area would most likely improve performance. If not, however, then all that could be achieved by increasing area would be to increase the number of spurious correlations detected, which could only reduce an observer's sensitivity to the pattern's motion.

There are similar inhomogeneities in observers' sensitivity to spatial frequency, which could account for the effects of check size in this experiment (cf. Morgan, 1992). Neurons with receptive fields in the periphery tend to prefer larger features than do those whose receptive fields are closer to the fovea. It would be reasonable to expect therefore that increasing check size should improve performance with peripherally presented patterns, which as shown in Fig. 10 is exactly what occurred. Note, however, that the large area patterns showed no effects of check size, thus providing additional evidence that the detection of motion in these patterns is significantly influenced by the stimulation of more central regions of the retina.

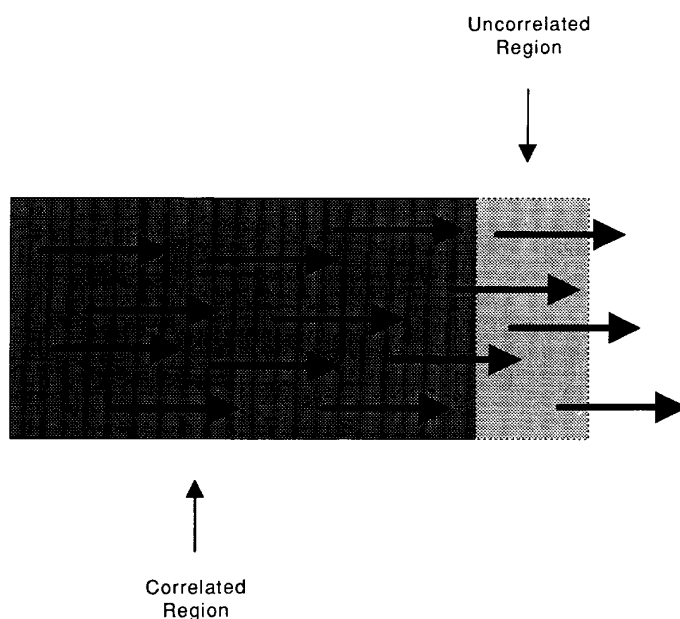


FIGURE 10. When a pattern is displaced behind a fixed aperture, some of the elements that were initially visible will be occluded by the aperture boundaries following the displacement. This produces a region with no correlated energy, whose width equals the size of the pattern displacement.

EXPERIMENT 7

At this point in our investigation we had confirmed many of the earlier findings of Baker and Braddick (1982, 1985a), van Doorn and Koenderink (1982b, 1984), Fredericksen *et al.* (1994a, b) and van de Grind *et al.* (1983, 1986, 1993), but we were still unable to reconcile the large effects of temporal integration in our experiments with the much more modest effects obtained by Snowden and Braddick (1989a). Although there were several differences in our respective methodologies, there was one factor in particular that we suspected might be especially important. In Expts 3–6 of the present series the patterns all moved within long thin apertures that were elongated in the direction of motion—a shape that was selected because it had been shown by van Doorn and Koenderink (1984) to be the optimal configuration for detecting motion in the presence of noise. Snowden and Braddick (1989a), in contrast, used patterns moving within square apertures.

There are several possible reasons why the shape of an aperture could affect how motion detectors are temporally integrated. Consider, for example, a model that employs higher order correlations to track the motion of a target over extended intervals of both space and time. To achieve the maximum amount of integration from such a mechanism, the spatial extent of the aperture in the direction of motion would be just as important as the number of distinct jumps of the pattern within it. Experiment 7 was designed to investigate this possibility by measuring D_{\max} for patterns moving within apertures of constant area that varied in shape.

Methods

The apparatus and procedure were identical to those used in Expts 3–6. The rectangular apertures all had a horizontal width of 24.5 deg, a vertical height of 9.2 min arc and a check size of 4.6 min arc. The retinal eccentricity of the apertures was varied across blocks between 1.2 and 8.6 deg, and the apparent motion sequences could contain either 2 or 16 distinct frames. The SOA for all displays was 50 msec.

It is important to recognize that when patterns are displaced in discrete jumps behind a fixed aperture there will always be some uncorrelated features due to accretion and deletion of texture at the aperture boundaries. Consider a pattern that is displaced by a distance D within a fixed aperture of height H and width W . As shown in Fig. 10, a portion of the pattern that is visible in the first frame will also be visible after the displacement—we refer to this as the correlated region. There will also be an uncorrelated region of width D in which points visible in the first frame are occluded by the aperture in the second. This uncorrelated region contains no useful information. Thus, the effective width of the stimulus for detecting motion is equal to the width of the correlated region, which is $W - D$ (see Lappin & Bell, 1976; van Doorn & Koenderink, 1984). For the displays used in the present experiment, the correlated region of every pattern had an identical area of 3.2 deg². The

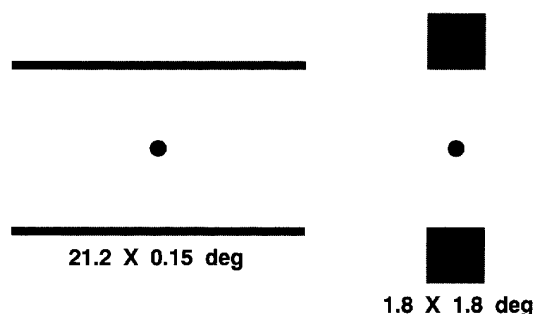


FIGURE 11. Aperture shape. The basic stimulus configuration employed in Expt 7 included pairs of apertures whose widths could vary from trial to trial, but whose correlated regions all had the same area.

horizontal width of this region was varied across blocks as shown in Fig. 11 with possible values of 1.7, 7.1, 14.1, and 21.2 deg. The width of the entire aperture equaled the width of the correlated region plus the magnitude of the pattern displacement. Although the correlated region remained constant within each block, the aperture width varied from trial to trial due to changes in pattern displacement throughout the adaptive staircase. Both authors again participated as observers. They each ran five staircases for each of the 16 possible display conditions (2 sequence lengths \times 2 retinal eccentricities \times 4 aperture widths).

Results

The individual results for both observers are shown in Fig. 12. It is clear from these data that the length of an aperture in the direction of motion is a critical factor for determining the effects of temporal integration on D_{\max} . Increasing the number of distinct frames within elongated apertures can increase D_{\max} by an order of magnitude for peripherally presented patterns. However, with approximately square apertures such as those used by Snowden and Braddick (1989a) the effects of temporal integration are much more modest.

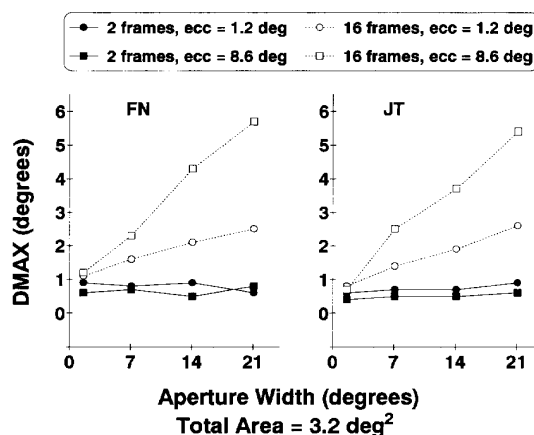


FIGURE 12. The maximum displacement thresholds for the two observers in Expt 7 as a function of aperture width for moving patterns with different retinal eccentricities and sequence lengths.

One possible hypothesis to account for this result is that the mechanisms sensitive to the largest possible displacements involve higher order correlations of motions or features over multiple intervals of space and time (e.g. see McKee & Welch, 1985). An ideal detector for a target moving at high speed along a continuous trajectory would sample the correlation at as many points along that trajectory as possible. As the length of the visible trajectory is decreased, so would its ability to integrate information over time.

Although this might appear at first blush to be a plausible explanation for the results shown in Fig. 12, a closer examination reveals a puzzling discrepancy. For the 16-frame displays presented in the periphery with elongated apertures, the observers could reliably detect displacements of 6 deg per jump. Since the apertures in this experiment had a maximum horizontal extent of 24.5 deg, no individual feature would have been visible for more than four successive views, yet the results of Expt 5 have shown that D_{\max} continues to increase for 16 frames or more. This obviously cannot be due to the tracking of individual features.

Another possible form of temporal integration could involve cooperative interactions, in which neighboring neurons tuned to similar velocities have a mutually facilitative effect on each other that persists over some period of time. This type of interaction has been proposed previously by Williams and Sekuler (1984) and by Snowden and Braddick (1989a), but there is nothing in these proposals to suggest why the region of cooperativity should be so anisotropic with respect to the direction of motion.

Still another possible form of temporal integration suggested by van Doorn and Koenderink (1984) and Fredericksen *et al.* (1994a, b, c) is to summate or integrate the local correlations within some neighborhood over an extended period of time. To implement this strategy, it would first be necessary to separately summate the correlated pixels in a pattern for each possible target displacement to be detected. It is important to recognize that any temporally changing pattern will produce local correlations at all possible spatiotemporal displacements. Suppose, for example, that a pattern is composed of scintillating random noise, in which each individual pixel has a 50% probability of being black or white at each distinct moment in time. Although such a pattern would be globally uncorrelated, there would be numerous local correlations occurring by chance that could stimulate motion sensitive neurons within the human visual system. What would identify the pattern as dynamic noise is that all possible local correlations would be detected with equal frequency. These spurious local correlations would also occur for patterns undergoing coherent motion, but the neurons sensitive to the actual displacement of the pattern in that case would be responding in greater numbers. Thus, in order to detect the presence of coherent motion, a second stage process would be needed to compare the frequency of local correlations for displacements of different magnitude and in different directions (cf. van de Grind *et al.*, 1992).

As has been described previously by van Doorn and Koenderink (1982a, b, 1984)—see also Lappin and Bell (1976)—there are two primary factors that would affect the performance of this type of process: (1) the number of correlated features to be detected within the spatial and temporal region of integration; and (2) the proportion of uncorrelated noise. In the present experiment the number of correlated pixels at each frame transition was equated in all displays, but there were two sources of noise in these displays that varied across the different conditions.

One potential source of noise was the uncorrelated region of each aperture as described in Fig. 10. Because the area of this region is defined by the product of the aperture height H and the spatial displacement D , the level of noise in each display covaried with the aperture shape—i.e. for a given displacement the number of uncorrelated pixels decreased proportionally with the aperture height. The other potential source of noise occurred periodically at the frame transitions when each apparent motion sequence was repeated. For 16-frame displays, this would occur every 800 msec, but for the 2-frame displays it occurred every 100 msec. Thus, the amount of uncorrelated noise also covaried with the length of the apparent motion sequences.

It would appear from this analysis that the mechanisms of temporal integration proposed by van Doorn and Koenderink (1984) or Fredericksen *et al.* (1994a, b, c) can account naturally for all of the results of the present experiment. According to this view, D_{\max} should increase with any display manipulation that reduces the amount of uncorrelated noise, and vice versa. Experiment 8 was designed to explore this hypothesis in greater detail by examining how observer sensitivity to high velocity motion is influenced by several different types of noise.

EXPERIMENT 8

Methods

The apparatus and procedure were identical to those used in Expts 3–7, except that the displays were generated on a Silicon Graphics Personal Iris (4D/25 with Turbo graphics). The rectangular apertures all had a horizontal width of 24.5 deg, a vertical height of 9.2 min arc, and a check size of 4.6 min arc. The retinal eccentricity of the apertures was varied across blocks between 1.2 and 8.6 deg, and the SOA for all displays was 50 msec. Unlike our earlier experiments, the patterns were all constructed from two overlapping arrays of checks whose respective image intensities were additively superimposed. Thus, 25% of the composite checks were white, 25% were black and 50% were a medium gray.

For the aperture that contained coherent motion, there were five different noise conditions that varied across blocks. In the *Continuous, Unidirectional* condition, both overlapping layers moved together with the same frame-to-frame displacement. These patterns moved continuously for the duration of each trial, as if

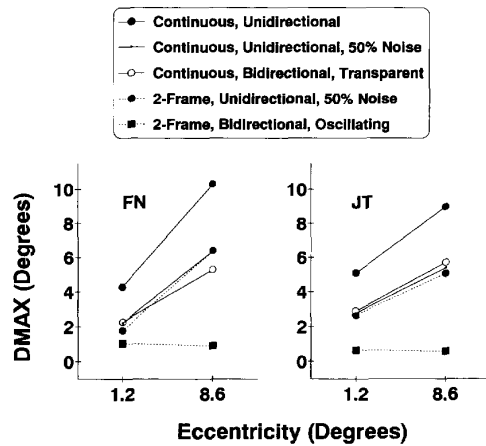


FIGURE 13. The maximum displacement thresholds for two observers as a function of retinal eccentricity for all of the different conditions employed in Expt 8.

an infinitely long strip of random noise were being moved behind the aperture. In the *Continuous, Unidirectional, 50% Noise* condition, one of the overlapping check patterns was moved continuously, while the other was temporally uncorrelated. In the *Continuous, Bidirectional, Transparent* condition, one of the overlapping check patterns was moved continuously to the right, while the other was moved to the left. In the *2-Frame, Unidirectional, 50% Noise* condition, both overlapping layers moved together for one transition, and were then replaced by a new pair of independent patterns for the next. This cycle repeated continuously with new random patterns presented on every other frame until the observer made an appropriate response. Finally, in the *2-Frame, Bidirectional, Oscillating* condition, the two overlapping layers moved together—first to the left and then to the right on alternating frame transitions. For the aperture that contained no coherent motion, both overlapping layers of checks changed over time as a nonrepeating sequence of temporally uncorrelated random noise, except in the *2-Frame, Bidirectional, Oscillating* condition. For those displays, as in the 2-frame sequences of Expts 3–7, the aperture without motion contained two uncorrelated frames of noise that were presented in continuous alternation.

Both authors again participated as observers. They each ran five staircases for each of the 10 possible display conditions (2 retinal eccentricities \times 5 noise types).

Results

The individual results for both observers are shown in Fig. 13. One important aspect of these data that deserves to be highlighted is the extremely high displacement thresholds that were obtained in the *Continuous, Unidirectional* condition. For the peripherally presented patterns, both of the observers had a D_{\max} of approx. 10 deg, which corresponds to an upper velocity threshold of 200 deg/sec! Note that this is considerably higher than those obtained in any of our previous experiments.

The most likely reason for this difference is that the 16-frame sequences employed in those experiments were not sufficiently long to saturate the process of temporal integration (see van Doorn & Koenderink, 1984). If so, then observers' displacement thresholds for 16-frame displays would not be as large as what would otherwise be possible for arbitrarily long sequences such as those used in the present experiment.

D_{\max} was significantly reduced when the moving patterns were contaminated by uncorrelated noise, though even then, it was still possible for observers to detect up to 5 deg displacements of peripherally presented patterns. For those displays in which half of the checks were temporally uncorrelated, performance was unaffected by whether the moving checks had 2-frame lifetimes (in the *2-Frame, Unidirectional, 50% Noise* condition) or traveled in successive jumps over the entire length of the aperture (in the *Continuous, Unidirectional, 50% Noise* condition). This confirms the conclusion of Snowden and Braddick (1989a) and Fredericksen *et al.* (1994a, b) that temporal integration cannot be based on the tracking of individual features.

The relatively good performance in these 50% noise conditions raises an interesting question about why the 2-frame bidirectional oscillating displays were so difficult to detect. It cannot be due to the absence of successive jumps in a given direction, because that property was shared by the *2-Frame, Unidirectional, 50% Noise* condition, which produced much higher levels of performance. Nor can it be explained by inhibitory interactions of neurons tuned to motions in opposite directions, because that would be inconsistent with the relatively high levels of performance in the *Bidirectional, Transparent* condition.

There is one property of the *2-Frame, Bidirectional, Oscillating* displays that sets them apart from the other conditions employed in this experiment. For these oscillating displays, each individual jump in a given direction always depicted exactly the same pattern of random checks, whereas in all of the other conditions these patterns were continuously changing, either by random replacement, or by the accretion and deletion of texture at the aperture boundaries.

It is important to keep in mind that spatiotemporal integration in the detection of coherent motion cannot be based on the number of correlated checks in a pattern over a given region of space and time. What matters instead is the total number of bilocal correlators that are stimulated for each possible spatiotemporal displacement. There are two necessary conditions that need to be satisfied in order to stimulate a bilocal correlator: (1) a pattern must be displaced by an appropriate spatiotemporal extent; and (2) the spatial structure of the pattern must match the tuning characteristics of the correlator's front-end spatial filters (see Fig. 1).

Because of this latter condition, it is likely to be the case that there are several different neurons tuned to different spatial structures for each possible spatiotemporal displacement in each visual direction. Consider, for example, the neurons that are tuned to a

particular velocity in a particular region of the retina. Suppose that some of these neurons are selectively responsive to dark spots of a particular size against a lighter background. In order to ensure that any moving pattern can be detected, there would have to be other neurons tuned to light spots against a darker background and to spots of different sizes. Since it would not be possible to stimulate all of these neurons simultaneously, the amount of temporal integration would be expected to increase with the variety of different spatial structures that passes over each local region of the visual field.

The amount of structural variation in each local region is directly proportional to the number of distinct frames in an apparent motion sequence, and could therefore be responsible for the effects of sequence length in all of the experiments described thus far. Another relevant factor that covaried with sequence length is the amount of uncorrelated noise that occurred periodically at the frame transitions when an apparent motion sequence was repeated. Given the relatively high levels of performance for displays containing 50% noise in the present experiment, it is most unlikely that these intermittent noise bursts had much of an influence on the observer's displacement thresholds. Nevertheless, in an effort to disentangle these two different factors we performed one last experiment using moving patterns with translatory symmetry, so that there would be no burst of noise when a sequence was repeated.

EXPERIMENT 9

Methods

The apparatus and procedure were identical to those used in Expt 8. Each aperture contained a single 320×2 grid of checks, each of which had a 50% probability of being black or white on any given frame of a motion or noise sequence. The rectangular apertures all had a horizontal width of 24.5 deg, a vertical height of 9.2 min arc, and a check size of 4.6 min arc. The number of distinct frames in the motion and noise sequences was varied across blocks from possible values of 2, 4, 8, 16 and 32. The retinal eccentricity of the patterns was also varied across blocks between 1.2 and 8.6 deg, and the SOA for all displays was 50 msec.

All of the patterns were constructed with a translatory symmetry whose period was determined by the product of the number of distinct frames in a sequence and the spatial displacement D . This guaranteed that all of the checks in the moving patterns would be perfectly correlated over every frame transition, including the ones when the sequences were repeated. Both authors again participated as observers. They each ran five staircases for each of the 10 possible display conditions (5 sequence lengths \times 2 eccentricities).

Results

The individual results for both observers are shown in Fig. 14. Note in the figure that as the number of frames

per symmetry period increases from 2 to 32, there is a corresponding increase in D_{\max} from 0.5 to 7 deg for peripherally presented patterns. It is important to keep in mind while evaluating these data that the overall pattern of motion was identical in all conditions. Each individual check appeared on one side of the aperture and moved in a succession of discrete jumps until it disappeared on the opposite side. There were no uncorrelated checks at any of the frame transitions, except for accretion and deletion at the aperture boundaries (see Fig. 10), and there were no sudden reversals in direction. The only difference among the various conditions was the spatial structure of the moving patterns—or more specifically, their periods of translatory symmetry.

Let us now consider why the symmetry of a pattern should have such a large effect on observers' maximum displacement thresholds for the detection of coherent motion. Our basic hypothesis is that for each local region of visual space there are numerous bilocal correlators tuned to different velocities, and that for each of these velocities there are several different correlators tuned to different aspects of a pattern's spatial structure. As a pattern moves in successive displacements, a subset of the correlators in any given local region will be stimulated on each jump, depending upon the particular spatial structure that happens by chance to be positioned there at that moment. The number of neurons stimulated is likely to increase as new structural configurations pass through the region in subsequent jumps, but no further increases can occur once the pattern repeats itself due to translatory symmetry.

One interesting aspect of these data that deserves to be highlighted is that the time scale of this temporal integration is quite large. D_{\max} continues to rise with symmetry periods up to 32 discrete jumps. Because all of the displays had a 50 msec SOA, it follows from this result that the integration period can be as large as 1.6 sec, which is also consistent with the findings of van Doorn and Koenderink (1984) and Fredericksen *et al.* (1994a, b).

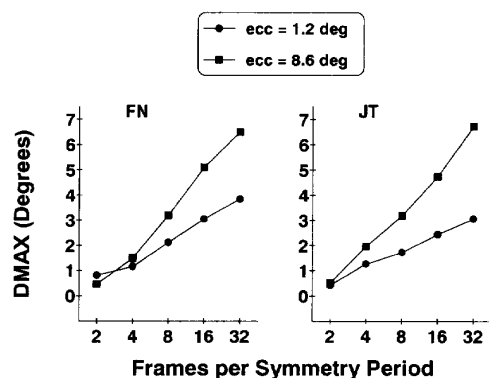


FIGURE 14. The maximum displacement thresholds for the two observers in Expt 9 as a function of the number of frames per symmetry period at two different retinal eccentricities.

DISCUSSION

The research described in the present article has examined the effects of many different stimulus variables on observers' maximum displacement thresholds for the perception of coherent motion. The results have shown that D_{\max} can vary over a surprisingly large range from 10 min arc to 10 deg depending upon the particular spatiotemporal properties of a moving pattern. When considered as a whole, there is a generally consistent pattern to these results, which suggests to us that the process of motion perception involves three distinct stages. We shall refer to these respectively as the *Detection Stage*, the *Integration Stage* and the *Decision Stage*. In the discussion that follows we will attempt to outline the basic operations of these stages, and how they can account for the effects of different stimulus manipulations on observers' maximum displacement thresholds.

The Detection Stage

The initial detection of motion energy must ultimately involve the stimulation of individual bilocal correlators. Based on the available physiological and psychophysical evidence, we can make several educated assumptions about how these correlators are constructed and how they are distributed within the visual field.

Assumption 1. The activity of bilocal correlators is governed by a simple delay and compare principle (see Fig. 1). There are several functionally equivalent strategies by which such a process could be neurophysiologically implemented. One possible mechanism, which is depicted schematically in Fig. 1, is designed to measure the spatiotemporal correlations among specific Fourier components of a moving pattern (Reichardt, 1957, 1961; van Santen & Sperling, 1984, 1985). It is also possible to conceive of this process as a spatiotemporal convolution (Adelson & Bergen, 1985), or to add a rectification stage in order to detect the motions of higher order, non-Fourier components (Chubb & Sperling, 1988). All that matters for the purposes of the present discussion, is that the mechanism be sensitive to specific aspects of image structure, within specific regions of the visual field, at different moments in time.

Assumption 2. For each local region of the visual field there are many different bilocal correlators tuned to different velocities and to different spatial characteristics of a moving pattern. This assumption is necessary because each individual motion detector as posited in Assumption 1, will only respond to a relatively narrow range of possible environmental events—that is to say, each detector is selective to a limited set of possible image structures, moving at a particular speed and direction, within a specific localized region of the retina. Thus, in order to accommodate the wide range of possible moving objects that one is likely to encounter in natural vision, an effective motion detection system would have to contain a large population of correlators with different tuning characteristics.

Assumption 3. In the human visual system, the minimum physiological delay for a bilocal correlator is

approx. 50 msec, and that is the delay parameter for most motion detectors with preferred velocities above 10 deg/sec. This assumption is based primarily on the elegant experiments of Koenderink *et al.* (1985) and van de Grind *et al.* (1986), who used several converging methods to estimate the spans and delays of bilocal correlators with receptive fields in different regions of the retina. Their findings were confirmed, moreover, by the results obtained in Expts 2 and 3 of the present series.

Assumption 4. The spatial distribution of motion detectors is inhomogeneous, such that bilocal correlators become more and more sensitive to high velocities and low spatial frequencies with increasing retinal eccentricity. This assumption follows naturally from the fact that the receptive field sizes of visually sensitive neurons are known to increase with retinal eccentricity—what is often referred to as cortical magnification. It is also supported by numerous psychophysical investigations (e.g. see Baker & Braddick, 1982, 1985a; Koenderink *et al.*, 1985; van de Grind *et al.*, 1983, 1986, 1993), including the results of the present experiments.

Assumption 5. The relative proportion of neurons tuned to different velocities is skewed, such that lower velocities are more densely sampled than are higher velocities. Although this assumption is somewhat more speculative than the others, it provides a plausible explanation for the observations of van Doorn and Koenderink (1982a, b, c, 1983, 1984) that increasing the velocity of a moving pattern produces a corresponding increase in observers' signal-to-noise thresholds.

The Integration Stage

It is important to keep in mind that an individual motion detector considered in isolation provides very little information, and is highly susceptible to spurious local correlations. Thus, in order to reliably distinguish motion from temporally uncorrelated random noise, it is necessary to pool information over a large population of detectors (see also the similar arguments of van de Grind *et al.*, 1992). From the results of the present experiments and the earlier findings of van Doorn and Koenderink (1982a, b, 1984) and Fredericksen *et al.* (1993, 1994a, b, c), it would appear that the spatiotemporal extent of this pooling is remarkably large. Observers' signal-to-noise and maximum displacement thresholds continue to improve as the size of a moving pattern is increased over many degrees of visual angle, or its duration is increased up to 2 sec. Our results suggest that this pooling is accomplished by combining the outputs of many different motion detectors. This could be achieved by a simple linear summation over space or time as suggested by Lappin and Bell (1976) and van Doorn and Koenderink (1984), although there is some evidence to suggest that the temporal pooling for relatively low velocities involves filtering the outputs of motion detectors through a leaky integrator (see Fredericksen *et al.*, 1993, 1994).

One other issue that needs to be considered with respect to pooling is the extent to which detectors tuned to different velocities are allowed to interact with one

another. For example, one possible method of integration would be to compute a vector average of all of the local velocities detected within a given neighborhood. A recent series of experiments by Watamaniuk *et al.* (1989) and Watamaniuk and Duchon (1992) has provided compelling evidence that observers are sensitive to the average speed or direction among a population of sparsely scattered dots moving at different velocities. Similar results have also been reported by Mingolla, Todd and Norman (1992) for patterns of moving lines within an array of apertures.

It is important to keep in mind, however, that a vector average cannot be the sole mechanism by which the outputs of bilocal correlators are integrated over space or time. An important limitation of a vector average is that it would preclude the perception of transparent motions in opposite directions, since the sum of their local correlations would effectively cancel one another. The fact that observers can perceive such motions with surprisingly little impairment, as was demonstrated in Expt 8 (see also van Doorn & Koenderink, 1982a, b; Lindsey & Todd, 1994), provides strong evidence that there may be multiple pooling processes performed in parallel, each of which is restricted to a relatively narrow range of velocities.

The Decision Stage

Unlike a vector average, a velocity specific pooling process would do nothing to eliminate spurious local correlations, and would therefore require a third stage of analysis to distinguish signal from noise. One possible method of performing this analysis would be to compare the relative amounts of activity among populations of motion detectors that are tuned to different velocities (cf. van Doorn & Koenderink, 1984; van de Grind *et al.*, 1992; Lappin & Bell, 1976). Consider, for example, a dynamic pattern of broad-band noise that is temporally uncorrelated. Although such patterns contain numerous local correlations that occur by chance, they can easily be identified by the fact that all populations of motion detectors would be stimulated with equal probability. These spurious correlations would also occur for an actual moving pattern, but the neurons tuned to its true velocity in that case would have a significantly higher probability of being stimulated. Thus, if one or more populations of motion detectors is firing with significantly greater frequency than others, it would be reasonable to conclude that those populations are responding to signal rather than noise. It is interesting to note in this regard that the same basic analysis could also be used to distinguish flickering patterns from motion, which could eliminate the need for opponent correlators at the front end (see Fig. 1).

Let us now examine how this three-stage analysis could account for the various stimulus factors that are known to affect D_{\max} . Our working hypothesis is that all these effects can be directly attributed to the number of spurious correlations detected relative to those resulting from an actual pattern of motion. According to this account, D_{\max} should increase as a result of any stimulus

manipulation that increases the number of signal correlations detected relative to those arising from noise, and vice versa. This prediction is based in part on Assumption 5 described above that the number of neurons available to be stimulated within the visual system diminishes with higher and higher velocities. Since the sensitivity of the Decision Stage would be proportional to the square root of the population size (see van Doorn & Koenderink, 1984; Lappin & Bell, 1976), it follows from this assumption that a greater proportion of the available detectors would have to be stimulated (relative to the level of background noise) in order to detect high velocities than would be necessary for low velocities.

The most direct of way of increasing the number of signal correlations detected is to increase the area of a moving pattern or the number of distinct frames in its displacement sequence (cf. van Doorn & Koenderink, 1984; Fredericksen *et al.*, 1993, 1994a, b, c). As more and more structure is moved over the visual field there is a corresponding increase in the number of motion detectors stimulated, which should improve sensitivity to a pattern's motion. Under optimal conditions, the variation of these parameters can have dramatic effects on performance, such that observers' signal-to-noise and maximum displacement thresholds can vary by as much as two orders of magnitude.

Performance can also be affected by stimulus manipulations that influence the number of spurious correlations detected. In Expt 8, for example, the displays that contained 50% uncorrelated noise produced much lower displacement thresholds than did those that contained 100% coherent motion. Similar effects can be obtained due to changes in aperture shape, because of the accretion and deletion of texture at aperture boundaries. The number of uncorrelated pixels in that case is linearly proportional to the height of an aperture perpendicular to the direction of motion, and the results of Expt 7 confirmed that increasing aperture height for a given area produces a corresponding decrease in D_{\max} .

It is important to keep in mind that the motion energy in a stimulus pattern can only be detected if it conforms to the individual tuning characteristics of an appropriate population of bilocal correlators. The evidence suggests that the highest velocity motion detectors all have temporal delays of approx. 50 msec, that they are primarily tuned to low spatial frequencies, and that they tend to have receptive fields in more peripheral regions of the retina. It follows therefore that D_{\max} should decrease as the structure of a moving display deviates more and more from these optimal conditions, and the results of the present experiments confirm that this is the case.

There are other predictable consequences of the inhomogeneous distribution of motion detectors that may appear at first blush to be counterintuitive. Consider, for example, a pair of displays that have the same maximum eccentricity but differ in area. Although increasing area should in general produce a corresponding increase in D_{\max} , that can only occur if the additional regions of the retina stimulated contain receptive fields of neurons that

are sensitive to the depicted pattern velocity. If not, then increasing area would only increase the number of spurious correlations detected, which would reduce D_{\max} (e.g. see Expt 6). A similar phenomenon would also be expected to occur from variations in the spectral bandwidth of a pattern. Although the number of motion detectors stimulated should in general increase with bandwidth, that can only improve performance if there are neurons tuned to the additional spatial frequencies that are also sensitive to the pattern's velocity. If not, then increasing bandwidth would only increase the number of spurious correlations detected, which should again result in a reduction of D_{\max} (e.g. see Chang & Julesz, 1983b; Cleary & Braddick, 1990a, b).

In all of the discussion presented thus far we have considered the detection of motion for complex noise patterns as an inherently global statistical process. Spatiotemporal integration is essential in this context in order to distinguish signal from noise due to spurious local correlations, but the fact that this integration can occur over many degrees of visual angle and for temporal durations on the order of 2 sec is remarkable nonetheless. While the pooling of information over such massive regions of space and time would be useful for detection, it would seem to cause problems for the process of figural segregation or the localization of a moving object to a precisely defined position. There is considerable evidence to suggest, however, that the visual system is somehow able to overcome these problems. For example, Regan (1986) has shown that vernier acuity for motion defined bars can be as low as 27 sec arc.

It is interesting to note that there are analogous tradeoffs in the analysis of luminance contours in the competing demands of spatially integrating information while maintaining well defined boundaries for figural segregation. Grossberg and Mingolla (1985, 1987) have recently proposed a possible mechanism for achieving both of these goals simultaneously. Their model employs two interacting systems: a boundary contour system that adaptively builds sharp edges around figurally segregated regions; and a feature contour system that integrates information about surface color within those regions. Given the success of this model on a wide variety of phenomena involving the perceptual organization of luminance contours, it seems reasonable to speculate that a similar type of mechanism might be equally useful for the processing of motion.

REFERENCES

- Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2, 284–299.
- Baker, C. L. & Braddick, O. J. (1982a). The basis of area and dot number effects in random dot motion perception. *Vision Research*, 22, 1253–1259.
- Baker, C. L. & Braddick, O. J. (1982b). Does segregation of differently moving areas depend on relative or absolute displacement? *Vision Research*, 22, 851–856.
- Baker, C. L. & Braddick, O. J. (1985a). Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Research*, 25, 803–812.
- Baker, C. L. & Braddick, O. J. (1985b). Temporal properties of the short-range process in apparent motion. *Perception*, 14, 181–192.
- Barlow, H. B. & Levick, W. R. (1965). The mechanism of directionally sensitive units in rabbit's retina. *Journal of Physiology*, 178, 477–504.
- Braddick, O. J. (1974). A short range process in apparent motion. *Vision Research*, 14, 519–527.
- Braddick, O. J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London B*, 290, 137–151.
- Chang, J. J. & Julesz, B. (1983a). Displacement limits, directional anisotropy and direction versus form discrimination in random-dot cinematograms. *Vision Research*, 23, 639–646.
- Chang, J. J. & Julesz, B. (1983b). Displacement limits for spatial frequency filtered random-dot cinematograms in apparent motion. *Vision Research*, 23, 1379–1385.
- Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, 5, 1986–2006.
- Cleary, R. & Braddick, O. J. (1990a). Direction discrimination for band-pass filtered random dot kinematograms. *Vision Research*, 30, 303–316.
- Cleary, R. & Braddick, O. J. (1990b). Masking of low frequency information in short-range apparent motion. *Vision Research*, 30, 317–327.
- van Doorn, A. J. & Koenderink, J. J. (1982a). Spatial properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, 45, 189–195.
- van Doorn, A. J. & Koenderink, J. J. (1982b). Temporal properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, 45, 179–188.
- van Doorn, A. J. & Koenderink, J. J. (1982c). Visibility of movement gradients. *Biological Cybernetics*, 44, 167–175.
- van Doorn, A. J. & Koenderink, J. J. (1983). Detectability of velocity gradients in moving random dot patterns. *Vision Research*, 23, 799–804.
- van Doorn, A. J. & Koenderink, J. J. (1984). Spatiotemporal integration in the detection of coherent motion. *Vision Research*, 24, 47–53.
- van Doorn, A. J., Koenderink, J. J. & van de Grind, W. A. (1985). Perception of movement and correlation in stroboscopically presented noise patterns. *Perception*, 14, 209–224.
- Fredericksen, R. E., Verstraten, F. A. J. & van de Grind, W. A. (1993). Spatio-temporal characteristics of human motion perception. *Vision Research*, 33, 1193–1205.
- Fredericksen, R. E., Verstraten, F. A. J. & van de Grind, W. A. (1994a). Temporal integration of random dot apparent motion information in human central vision. *Vision Research*, 34, 461–476.
- Fredericksen, R. E., Verstraten, F. A. J. & van de Grind, W. A. (1994b). An analysis of the temporal integration mechanism in human motion perception. *Vision Research*, 34, 3153–3170.
- Fredericksen, R. E., Verstraten, F. A. J. & van de Grind, W. A. (1994c). Spatial summation and its interaction with the temporal integration mechanism in human motion perception. *Vision Research*, 34, 3171–3188.
- van de Grind, W. A., van Doorn, A. J. & Koenderink, J. J. (1983). Detection of coherent movement in peripherally viewed random-dot patterns. *Journal of the Optical Society of America*, 73, 1674–1683.
- van de Grind, W. A., van Doorn, A. J. & Koenderink, J. J. (1987). Influence of contrast on foveal and peripheral detection of coherent motion in moving random-dot patterns. *Journal of the Optical Society of America A*, 4, 1643–1652.
- van de Grind, W. A., Koenderink, J. J. & van Doorn, A. J. (1986). The distribution of human motion detector properties in the monocular visual field. *Vision Research*, 26, 797–810.
- van de Grind, W. A., Koenderink, J. J. & van Doorn, A. J. (1992). Viewing-distance invariance of movement detection. *Experimental Brain Research*, 91, 135–150.
- van de Grind, W. A., Koenderink, J. J., van Doorn, A. J., Milders, M. V. & Voerman, H. (1993). Inhomogeneity and anisotropies for

- motion detection in the monocular field of human observers. *Vision Research*, *33*, 1089–1107.
- Grossberg, S. & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries and emergent segmentations. *Perception & Psychophysics*, *38*, 141–171.
- Grossberg, S. & Mingolla, E. (1987). Neural dynamics of surface perception: Boundary webs, illuminants, and shape from shading. *Computer Vision, Graphics, and Image Processing*, *37*, 116–165.
- Koenderink, J. J., van Doorn, A. J. & van de Grind, W. A. (1985). Spatial and temporal parameters of motion detection in the peripheral visual field. *Journal of the Optical Society of America A*, *2*, 252–259.
- Lappin, J. S. & Bell, H. H. (1976). The detection of coherence in moving random-dot patterns. *Vision Research*, *16*, 161–168.
- Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.
- Lindsey, D. T. & Todd, J. T. (1994). Opponent-motion interactions in transparent motion. *Investigative Ophthalmology and Visual Science*, *35*, 1389.
- McKee, S. P. & Welch, L. (1985). Sequential recruitment in the discrimination of velocity. *Journal of the Optical Society of America A*, *2*, 243–251.
- Mingolla, E., Todd, J. T. & Norman, J. F. (1992). The perception of globally coherent motion. *Vision Research*, *32*, 1015–1031.
- Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, *355*, 344–346.
- Nakayama, K. & Silverman, G. H. (1984). Temporal and spatial characteristics of the upper displacement limit for motion in random dots. *Vision Research*, *24*, 293–299.
- Reichardt, W. (1957). Autokorrelationsauswertung als funktionssprinzip des Zentral-nervensystems. *Zeitschrift für Naturforschung*, *12b*, 447–457.
- Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Rosenblith, W. A. (Ed.), *Sensory communication* (pp. 303–317). New York: Wiley.
- Regan, D. (1986). Form from motion parallax and form from luminance contrast: Vernier discrimination. *Spatial Vision*, *1*, 305–308.
- van Santen, J. P. H. & Sperling, G. (1984). A temporal covariance model of motion perception. *Journal of the Optical Society of America A*, *1*, 451–473.
- van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, *2*, 300–321.
- Snowden, R. J. & Braddick, O. J. (1989a). Extension of displacement limits in multiple-exposure sequences of apparent motion. *Vision Research*, *29*, 1777–1788.
- Snowden, R. J. & Braddick, O. J. (1989b). The combination of motion signals over time. *Vision Research*, *29*, 1621–1630.
- Snowden, R. J. & Braddick, O. J. (1990). Differences in the processing of short-range apparent motion at small and large displacements. *Vision Research*, *30*, 1211–1222.
- Watamaniuk, S. N. J. & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*, 931–941.
- Watamaniuk, S. N. J. & Sekuler, R. (1992). Temporal and spatial integration in dynamic random dot stimuli. *Vision Research*, *32*, 2341–2347.
- Watamaniuk, S. N. J., Sekuler, R. & Williams, D. W. (1989). Direction perception in complex dynamic displays: The integration of direction information. *Vision Research*, *29*, 47–59.
- Watson, A. B. & Ahumada, A. J. (1983). A look at motion in the frequency domain. NASA Technical Memo TM-84352.
- Williams, D. & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research*, *24*, 55–62.

Acknowledgement—This research was supported in part by AFOSR grant (No. F49620-93-0116).