Vision Res. Vol. 32, No. 6, pp. 1015–1031, 1992
Printed in Great Britain. All rights reserved

0042-6989/92 $5.00 + 0.00

The Perception of Globally Coherent Motion*

ENNIO MINGOLLA,† JAMES T. TODD,‡ J. FARLEY NORMAN‡

Received 10 April 1991; in revised form 19 November 1991

How do human observers perceive a coherent pattern of motion from a disparate set of local motion measures? Our research has examined how ambiguous motion signals along straight contours are spatially integrated to obtain a globally coherent perception of motion. Observers viewed displays containing a large number of apertures, with each aperture containing one or more contours whose orientations and velocities could be independently specified. The total pattern of the contour trajectories across the individual apertures was manipulated to produce globally coherent motions, such as rotations, expansions, or translations. For displays containing only straight contours extending to the circumferences of the apertures, observers' reports of global motion direction were biased whenever the sampling of contour orientations was asymmetric relative to the direction of motion. Performance was improved by the presence of identifiable features, such as line ends or crossings, whose trajectories could be tracked over time. The reports of our observers were consistent with a pooling process involving a vector average of measures of the component of velocity normal to contour orientation, rather than with the predictions of the intersection-of-constraints analysis in velocity space.

Aperture problem Velocity space Vector average Component motion Pattern motion

INTRODUCTION

The transition from vast arrays of locally ambiguous signals to globally coherent perception of objects is a fundamental problem in all early vision modalities. In the motion domain this early ambiguity is often referred to as the *aperture problem*, whereby only the normal component of velocity of any straight line or edge segment that extends beyond a cell's receptive field can in principle be measured by that cell (Fennema & Thompson, 1979; Adelson & Movshon, 1982). The question of *how* to combine information from many such local measures into a consistent object velocity signal has drawn considerable attention in the last decade, both within the computer vision community (reviewed in Uras *et al.*, 1988) and within the psychological and physiological investigation of biologial vision (Adelson & Movshon, 1982; Derrington & Suero, 1991; Ferrera & Wilson, 1987, 1990, 1991; McKee, Silverman & Nakayama, 1986; Movshon, Adelson, Gizzi & Newsome, 1985; Welch, 1989).

Many of the phenomenological consequences of the aperture problem were described by Wallach (1976). He observed that when both ends of a straight contour are occluded by a circular aperture, the contour is often perceived to move in a direction that is perpendicular to its own orientation—even when there is a visible textural grain on the contour or in the presence of intersections of contours that could potentially specify (some other) true direction of motion. These effects of contour orientation can sometimes be overcome, however, when the end points or terminators of several contours within a single aperture all move in the same direction. One way of achieving this effect is to present moving contours within an elongated rectangular aperture, which results in the well-known "barberpole illusion".

More recently Adelson and Movshon (1982) have argued that the perceptual ambiguities arising from the aperture problem can be overcome through an "intersection-of-constraints" computation in velocity space. According to their model, the correct direction of motion for patterns of contours with multiple orientations can be determined using a two-stage process. The first stage is performed by a network of directionally sensitive, orientationally tuned mechanisms. A local signal at the first stage is proportional to the normal component of the velocity of stimulus contrast (edge or grating) in a direction perpendicular to the orientational preference of a particular detector. Signals are next processed by mechanisms with different orientational preferences but overlapping spatial domains. When a stimulus containing components in two distinct orientations, such as a "plaid" pattern made of overlapping sinusoidal gratings of different orientations, drifts over a given region, two populations of detectors at the first

---

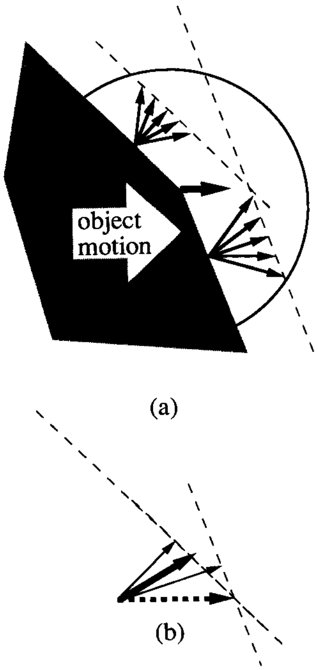
*Preliminary reports on the experiments described in this article were presented at the November 1990 annual meeting of the Psychonomic Society in New Orleans and at the May 1991 annual meeting of the Association for Research in Vision and Ophthalmology.

†To whom all correspondence should be addressed at: Center for Adaptive Systems, and Department of Cognitive and Neural Systems, Boston University, 111 Cummington Street, Boston, MA 02215, U.S.A.

‡Department of Psychology, Brandeis University, 415 South Street, Waltham, MA 02154, U.S.A.

stage respond. Assuming that both gratings are above detection threshold and of similar contrast and spatial frequency, the reported perception is generally of a single coherent motion in a direction which is not perpendicular to either grating. Rather, the reported motion follows a trajectory predicted by an intersection-of-constraints construction, as shown in Fig. 1. The resolution of component (grating) motions into pattern (plaid) motion is the second stage of the model.

Figure 1(b) also depicts two other possible solutions besides the intersection-of-constraints to the problem of resolving multiple or ambiguous motion directions within an aperture. The visual system could instead compute a vector average of the two directions normal to the orientations of contours, and several researchers have suggested such a mechanism may operate for some visual displays (Ferrera & Wilson, 1987; Watson & Ahumada, 1985; Williams & Phillips, 1987). Alternatively, where contours cross one another or meet at a corner, a perceptually useful feature may exist. The visual system may be able to track that feature and choose those motions of the oriented contours that are consistent with feature motion. Of the three solutions just described (intersection-of-constraints, vector averaging, and feature tracking) the first and third always yield the same result for translatory motion, and the second may be near to the other two depending on the configuration of contour orientations and speeds. For many psychophysical tasks it would be impossible to conclude from a subject's performance which of the solutions (or which combination) was being adopted. Thus, in an effort to investigate the psychological validity of these alternative solutions, we set out to create a set of displays that would produce different predicted outcomes for each one.



(a)

(b)

FIGURE 1. (a) An object with a corner moves horizontally behind a circular aperture. The thick arrow indicates the trajectory of the corner, while the thin ones indicate some of the motions consistent with local information at straight edge segments. (b) The two thin arrows indicate the component motions for the diagram in (a); the thick arrow represents the average of the two component vectors, and the dashed arrow represents the velocity space solution, which corresponds with the trajectory of the corner in (a).

A vexing confound for the velocity space model has been that the direction of motion of the luminance "blobs" formed by the overlap of the two component gratings always follows the resultant direction of the velocity space intersection-of-constraints solution. Considerable ingenuity has been employed in addressing this confound. Movshon *et al.* (1985) review several lines of evidence for rejecting the hypothesis that the visual system is merely tracking "features" formed by local maxima in luminance at the component intersections. In one experiment "one-dimensional" noise consisting of parallel stripes of varying widths was superimposed on plaid patterns, at times in orientations parallel to one of the component gratings and at times in the orientation perpendicular to the coherent plaid motion. They measured threshold elevation for detection of coherent motion and found that noise that was within about 20° of one of the component orientations was much more efficacious in masking the resulting percept than noise that was perpendicular to the direction of plaid motion, and concluded that ". . . the mechanisms responsible for the phenomenal coherence of moving plaids belong to a pathway which, at some point, passes through a stage of orientation selective spatial analysis". In another experiment they also measured the effects of adaptation on coherence, using a factorial design with which single gratings or plaids were used both as adapting and test stimuli. The strongest adaptation effects were found in the conditions where identical component orientations appeared in both adaptation and test stimuli, again implying the existence of a stage of early processing sensitive to component orientation. More recently Welch (1989), using a speed discrimination threshold task, found that the speed of component gratings, not of the resulting plaids, determines speed discriminations thresholds for the plaid pattern itself, lending further support to a two-stage model of motion resolution. Still more recently Derrington and Suero (1991) used a motion aftereffect paradigm to show that apparent speed of component gratings influences perceived direction of motion of plaids.

It has been known since the outset that not all signals of the first stage originating from a single spatial locus can be combined. Adelson and Movshon (1982) observed that plaids whose component gratings are of sufficiently different contrasts or sufficiently different spatial frequencies do not cohere, but are perceived as separate motions in the directions perpendicular to the orientation of each component. Movshon *et al.* (1985) also reported elevation in threshold for coherence as the angle between component gratings is increased. Ferrera and Wilson (1987, 1990) extended the investigation of the effects of component angles in perception of coherent motion. They analyzed various classes of component combinations, called Type I symmetric,

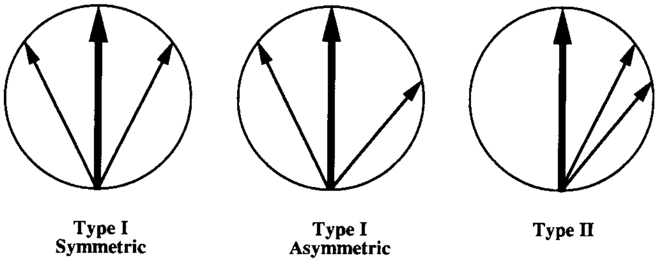


FIGURE 2. Velocity space diagrams for the classification of stimuli used by Ferrera and Wilson (1990). Thin arrows represent the component motions of sine wave gratings, while the thick arrow represents the resultant motion. Adapted from Ferrera and Wilson (1990).

Type I asymmetric, and Type II, as illustrated in Fig. 2. The psychophysical studies of Adelson and Movshon (1982) and Movshon *et al.* (1985) had employed only Type I patterns. Ferrera and Wilson (1987) found the phenomenal appearance of Type I patterns to be so different from that of Type II patterns that they referred to only the former as having the appearance of a rigid "plaid" in motion and referred to the latter as having the appearance of a fluid motion of luminance "blobs" sliding along the component gratings. Their 1987 study measured the ability of Type I and Type II patterns to mask the detectability of a test pattern. They found that the former produced a substantially increased masking effect over the effect of either component in isolation, over a wide range of angular separations in orientation between the two components. Type II patterns produced no such elevation, with the strength of masking being determined largely by relations of contrast, spatial frequency, and angular separation between test pattern and the component of the masking pattern which was closer in orientation to the test pattern. In a later study in which perceived direction of motion and thresholds for direction discrimination were measured, Ferrera and Wilson (1990) found that Type I patterns produced substantially lower discrimination thresholds than Type II patterns and that perceived motion directions of the former were essentially unbiased, while those of the latter were substantially biased toward the direction of the components. They concluded that "... neural mechanisms which compute two-dimensional image motion do not strictly implement the intersection-of-constraints construction proposed by Adelson and Movshon (1982)".

One characteristic of the psychophysical methodologies reviewed thus far is the use of sinusoidal gratings at two orientations. Inevitably, the intersections of the two gratings produce zones of increased luminance that are at least in principle features that the visual system could track in order to recover a coherent motion direction. While the experiments of Adelson and Movshon (1982) and Movshon *et al.* (1985) strongly implicate an orientation sensitive early process, they do not preclude its coexistence with a feature tracking process. Moreover, the experiments of Ferrera and Wilson (1987, 1990, 1991) imply that some operation *other than* the intersection-of-constraints calculation is going on at least some of the time. They suggest that a solution found by vector averaging of component normal velocities may be generated in addition to, or instead of, an intersection-of-constraints solution (Ferrera & Wilson, 1990). Vector averaging may help to account for the bias in direction reports for their Type II displays, although as they point out, the difference between the two solutions is often quite small for Type I displays. In light of these observations, the goals of the present investigation were 3-fold: First, to control for the possible effects of motions of identifiable points, such as line crossings or terminators, whose trajectory, could be tracked over time; second, to compare the relative perceptual salience of a velocity space solution and what would be expected from a process of vector averaging of component velocities, as suggested by Ferrera and Wilson (1990); and third, to extend the study of the aperture problem to other categories of coherent motion besides pure translation, such as rotation in the picture plane or simple scaling transformations (i.e. expansion or contraction).

EXPERIMENT 1

Experiment 1 was designed to explore how the visual system combines velocity information from oriented contours in translatory motion, and specifically how, in Adelson and Movshon's (1982) terms, the transformation from component motion signals to pattern motion signals is accomplished. We sought to devise an experimental paradigm for which the intersection-of-constraints solution would yield a response different from a solution based on vector averaging of component motions. This was done by constructing displays whose prevailing motion was, by any account, horizontally rightward or leftward, but whose motion would also be seen as having an upward or a downward component. Subjects indicated upward or downward in a forced choice paradigm.

Methods

A Silicon Graphics Personal IRIS workstation, with 1280 × 1024 pixel resolution, was used for all experiments. It generated moving displays that were in certain respects like drifting gratings, but for which the explicit depictions of intersections was controlled. This was achieved through the use of numerous "windows" or apertures, each of which could independently contain a line segment of a certain orientation moving at a certain normal velocity. For our displays each aperture contained contours of one of two possible orientations, and the velocities of all contours of each orientation were identical, although the phases of the contours within each aperture and the choice of which aperture contained which orientation were randomized to mitigate against the formation of illusory collinear groupings that might form illusory intersection features (see Fig. 3). Our displays therefore contained an analog of the component velocity information of the overlapping sinusoidal gratings paradigm, but without features formed by intersections of contours of two orientations.

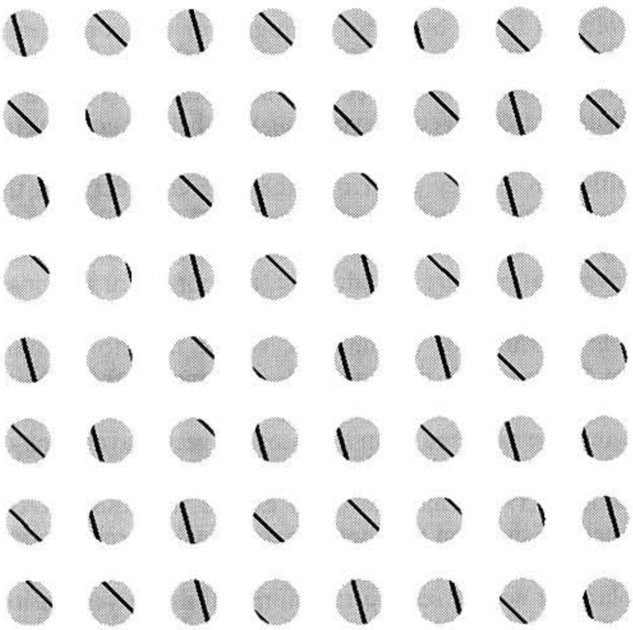


FIGURE 3. A depiction of a frame containing an upward bias of contour orientations for rightward motion. Within each aperture a line segment of one of two orientations moves with some normal velocity. The velocities of all segments of the same orientation are the same, though their phases are chosen independently. On the video screen used in the experiments, the background was blue, the contours gray, and apertures were alternately black and white. See the Methods section of Expt 1 for details on display construction.

In order to describe the key experimental manipulations, it is useful to portray the motions of individual contours in a somewhat nonstandard manner. Instead of starting the description by referring to the normal component of velocity, consider the construction of displays such as those of Fig. 3 as proceeding from the premise that there exists a *single* computed velocity, *henceforth called the* "true" *velocity, for the entire display*. The contours in each aperture move in accordance with that true velocity, such that the normal component of velocity to the contour varies as a cosine function of the contour orientation relative to the true direction of motion (see Fig. 4).

One important stimulus manipulation in these displays involved the presence or absence of identifiable feature points. In the *features absent* condition, each aperture contained a single moving line segment, at one of two possible orientations, with its endpoints occluded (see Fig. 3). To determine the correct direction of motion in this case, it would be necessary to integrate information over multiple apertures. In the *features present* condition, in contrast, each aperture contained two small rectangles aligned at one of the two possible orientations (see Fig. 5). For these displays the correct direction of motion could be determined within each aperture from the trajectory of any individual rectangle.

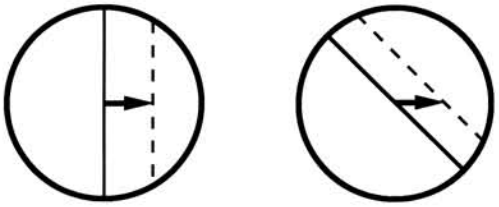


FIGURE 4. Two apertures, such as those shown in Fig. 3, contain oriented contours (solid) of different orientations. The arrows indicate the "true" direction of motion of the entire display, as described in the text. Dashed lines indicate the positions of the contours after a small time has elapsed. Displayed speed for each contour is a cosine function of the angle formed by the true velocity of the display and the orientation of the component contours.

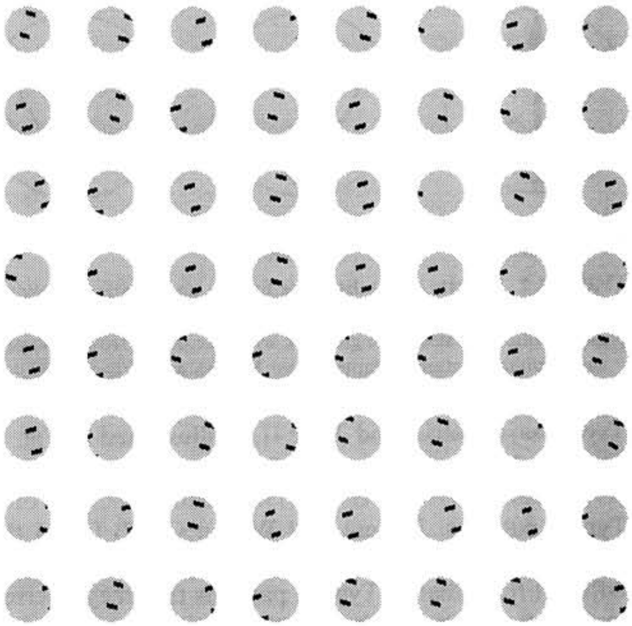


FIGURE 5. A depiction of a frame containing trackable features whose alignment is unbiased for rightward motion. The alignment of the short segments is randomly chosen within each aperture to be 15° to one side or the other of vertical.

The true directions of motion for these patterns were varied across trials from a set of 10 possible directions. For some of the displays the patterns moved along a horizontal trajectory—either to the right or to the left. For others the direction of motion deviated upward or downward from a horizontal direction by either 15 or 30°. The observers' task on each trial involved a forced choice judgment of whether the depicted pattern of motion appeared to be slanted upward or downward.

Another stimulus manipulation involved the particular combinations of contour (or feature) orientations presented in each display. In the *unbiased* condition, the contours in each aperture were all oriented at a 15° angle from the vertical, slanting either left-to-right or right-to-left. For these displays it was theoretically possible to achieve accurate performance using either a velocity space or a vector averaging solution. In the *upward biased* and *downward biased* conditions the contours in each aperture were all oriented to the same side of the vertical by either 15 or 45° (see Fig. 6). If observers adopted a vector averaging solution for these displays, their responses would be completely determined by contour orientations, regardless of the true directions of motion. If they adopted a velocity space solution, in contrast, then their responses would be completely determined by the true direction of motion, regardless of contour orientation. It would also be possible to adopt

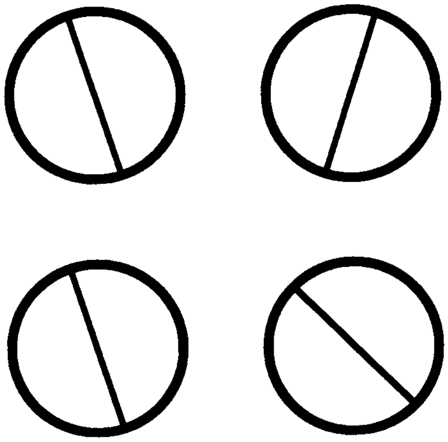


FIGURE 6. If the line segment orientations for displays such as Fig. 3 are randomly chosen as either 15° to one side or the other of the vertical, as on the top row of this figure, the display is said to be "unbiased" for the task of detecting horizontal or vertical components of display motion. Displays containing the orientations shown in the lower row of this figure (15 and 45° to the same side of vertical) are said to have an "upward bias", for motion which is left-to-right. Note that "unbiased" and "biased" are *not* synonymous with "Type I" and "Type II"; a comparison of stimuli according to the two forms of nomenclature is given in the text.

a compromise solution for which performance would be affected by both factors.

The stimulus manipulations can be summarized as follows: (1) identifiable feature points could be present or absent; (2) the contour orientations could be biased upward, biased downward or unbiased; (3) the prevailing direction of horizontal translatory motion could be rightward or leftward; (4) the true direction of motion could deviate from the horizontal by −30, −15, 0, 15, or 30°. Using all possible combinations of these factors, the resulting experimental design had 60 distinct conditions.

Each trial consisted of a 2.0 sec motion sequence composed of 60 distinct frames. Each frame displayed 64 apertures aligned in 8 rows by 8 columns (see Figs 3 and 5). The entire array of apertures subtended a horizontal visual angle of 10.6° at the viewing distance employed, 76 cm. The apertures were separated by gaps that were 0.8 times the aperture diameters, which subtended 47′. Alternate apertures were black or white, the background was blue, and all contours or features displayed within apertures were a medium gray. The alternation of black and white apertures was necessary to prevent the appearance of spurious motion of the apertures' interior background regions. (See Appendix.)

In the features absent condition each aperture contained a single straight contour with its endpoints occluded by the aperture boundary. The thickness of each contour was 10% of the aperture diameter. In the features present condition, each aperture contained a pair of small rectangles that were separated by a distance of half the aperture's diameter, with a length and width of 20 and 10% of the apertures' diameter, respectively. The orientations of all contours in the features absent condition and the alignment of the small rectangles in the features present condition were chosen in accordance with the bias condition on each trial. The relative phases of contours or features within the different apertures were determined randomly, and when they exited an aperture on one side, their computed positions were "wrapped around" to reenter the same aperture on the opposite side. The true velocity on each trial was 4.5°/sec. The normal components of velocity for the contours varied with their orientation, ranging from 0.9 to 3.0°/sec.

Two of the authors (JT and FN) and a graduate student volunteer participated in 5 experimental sessions. Each session contained 4 presentations of all 60 experimental conditions for a total of 240 trials. After each trial's 2 sec motion sequence, the observer pressed a left or right mouse button to indicate upward or downward motion. Observers viewed each trial through a monocular viewing hood at a distance of 76 cm.

Results and discussion

The results of Expt 1 are summarized in Fig. 7. Results for rightward and leftward motion were virtually identical, and have been pooled in the charts of Fig. 7. While the results were pooled over observers, each observer's data followed the same pattern. It is clear from these data that the presence or absence of identifiable feature points had a profound effect on the observers' perceptions. In the features present condition the observers performance was essentially perfect, in that they always responded upward when the true motion was upward, and they always responded downward when the true motion was downward. For the features absent condition, in contrast, the accuracy of the observers' judgments was barely above chance. Performance in that case appears to be primarily determined by the overall pattern of contour orientations presented in each display, irrespective of the true direction of motion.

Before further discussing the theoretical implications of Expt 1, we would emphasize that over the course of several pilot tests of the displays we chose parameters that yielded a perception of coherent motion for all conditions. The designation "coherent" appears to have ambiguous usage in the literature, with some apparently using it synonymously with "rigid". The appearance of some displays, namely those having the largest velocity differences between components, at times deviated from rigidity; these displays had a certain "wavy" appearance in the form of minor oscillatory deviations from a single prevailing direction. We still refer to these displays as coherent, however, in the sense that *in no case did we observe a scission* into two overlayed motion directions, one for each component.

It is important to recognize that the results obtained in the feature absent condition are not compatible with what would be expected based on a velocity space solution. If observers in this experiment had been able to make use of the intersection-of-constraints from relative velocities of different contour orientations, as suggested by Adelson and Movshon (1982), then surely they should have been capable of discriminating directions of motion that differ by as much as 60°

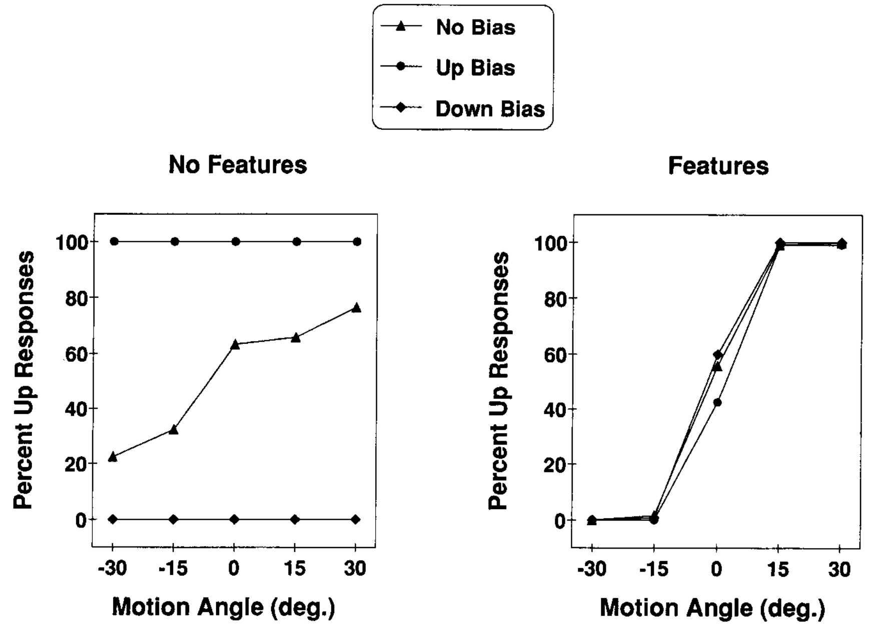


FIGURE 7. Results of Expt 1, pooled for all observers and for prevailing direction of motion. The percentage of up responses is in each condition plotted against displayed motion direction; true motion is downward for angles of −30 and −15° and upward for corresponding positive angles. Note that each data point in these plots represents 40 judgements for each of three observers.

The results are consistent, however, with what would be expected based on a vector average solution. To better appreciate why this is so, it is useful to consider the schematic diagram presented in Fig. 8. Figure 8(a) shows both the velocity space and vector average solutions for an upward biased display with a true direction of motion that is 30° downward. Whereas the intersection-of-constraints from the individual component velocities reveals the correct downward direction of motion, a vector average solution would indicate incorrectly that the direction of motion is upward, which is how this type of pattern is actually perceived. Figure 8(b) shows a similar set of solutions for an unbiased display with a true direction of motion that is 30° downward. Note in this case that the vector average and velocity space solutions both point downward, but that the downward orientation of the vector average solution is greatly attenuated relative to the true direction of motion. It would be reasonable to expect therefore that if observers' perceptions are determined by a process of vector averaging, then they ought to be relatively insensitive to the true direction of motion—even for displays that do not contain a contour orientation bias. Note in Fig. 7 that the results of the present experiment are consistent with these predictions.

Regarding the distinction between Type I and Type II patterns introduced by Ferrera and Wilson (1987), we note that in Expt 1 (and, to anticipate, in Expt 2) only 20% of our displays were Type I. Specifically, those displays were either unbiased with true horizontal motion, or biased displays whose true motion was 30° from the horizontal in the same direction as the bias. The

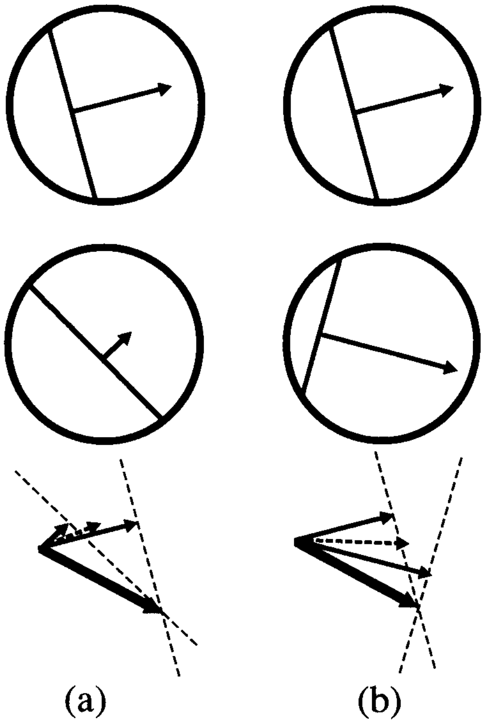


FIGURE 8. A comparison of vector average (thick dashed arrows) and intersection-of-constraints (thick solid arrows) predictions for the biased and unbiased displays of the feature absent condition of Expt 1. Note that the true veloicty is identical for both conditions, and corresponds with the intersection-of-constraints prediction. (a) For upward bias displays, the vector average prediction points upward, although the speed predicted by the vector average solution is greatly attenuated relative to that of the true direction of motion. (b) For unbiased displays with true downward motion, both the velocity space and vector average solutions point downward, although the vector average solution does so just barely.

rest of our displays were either Type II or "degenerate", in the sense that true motion was 15° from the horizontal, with one of the component patterns having the orientation normal to the direction of true motion. It is important to recognize in this context that for large angular separations between contours of two orientations, large differences in speed are needed to produce a Type II pattern, whereas small differences suffice for small angular separations. At limit, for the smallest differences in angle and speed, perception of motion is coherent and in the direction of the average of the component directions, even if not "veridical" relative to an experimenter's calculation of true pattern motion.

EXPERIMENT 2

Might the observers' inability to detect true motion directions for biased displays of Expt 1 be due to difficulty in integrating information from spatially separated apertures into a single coherent signal? One reply to this objection would be to remark that the phenomenal appearance of the displays is generally quite coherent. The displays appear to be moving in *some* direction that is consistent across apertures; that direction, however, is simply not the true one. We note that in pilot experiments we used a wide range of aperture spacings, including spacings considerably closer than those used in the reported experiment, and found no noticeable differences in performance.

The argument that motion information from contours of two different orientations needs to be spatially proximate in order for the intersection-of-constraints solution to be applied is seductive, but—we believe—misleading. A critical issue so often overlooked in the argument is the inevitable interaction of contours of two different orientations on oriented motion detectors. That is, the mere presence of image contrast in orientations other than a cell's preferred orientation can distort a cell's measurement of the time-varying distribution of oriented contrast in its receptive field. Moreover, reports of interactions among oriented contrast driven cells, such as cross-orientation inhibition (Morrone, Burr & Maffei, 1982; Snowden, 1989), suggest that the accomplishment of a "clean" measurement of motion energy in the direction normal to a contour's orientation would be no easy matter in a region containing corners or overlapping contours. Several researchers have analyzed the special circumstances induced by line ends and corners in the domain of static form perception (Grossberg & Mingolla, 1985a, b; Marshall, 1990a; Walters, 1987; Zucker, Dobbins & Iverson, 1989), and in the motion domain (Grossberg & Mingolla, 1990; Marshall, 1990b). It may be that measures of component velocity of the accuracy needed for the intersection-of-constraints construction are indeed feasible for our visual systems for pure periodic stimuli, such as sinusoidal gratings, despite the overlap of different orientations. In this case the consistent repetition of contours or the same orientation over wide areas may compensate for disruptions induced at regions of overlap. Periodic stimuli are hardly representative of most motion in everyday environments, however, where contours of two orientations meeting at corners abound.

We note that the physiological and psychophysical estimates of orientation bandwidth are on the order of 60 deg full bandwidth at half-height (Greenlee & Magnussen, 1988; Webster & De Valois, 1985). This means that lines differing in orientation by 30° could be processed by the *same* mechanisms, leading naturally to a signal corresponding to that of a line whose orientation was the average of the two actual ones. Is it possible that the vector average solution might be peculiar to angles as small as the ones in our study, with the intersection of constraints still a viable alternative for larger angular separations? This appears implausible. A range of 60° represents two-thirds of the *possible* range of orientation separations, leaving little territory for alternative solutions. Moreover, the bandwidth estimates for orientation selectivity are generally carried out for "pure" stimuli, containing only one orientation. As indicated in the immediately preceding paragraph, we think that the response of mechanisms of the visual system in regions containing more than one contour orientation is unlikely to produce an accurate enough estimate of component velocities to be of use for the intersection-of-constraints computation.

A control for the objection that motion information could not be integrated across neighboring apertures in Expt 1 is easily generated, in any case, by displaying contours of both of two orientations, such as sampled in Expt 1, *within each aperture*. This has several desirable effects: First, intersections that form trackable features are generated, performance in the presence of these features can be compared with performance in the presence of features of a different kind in Expt 1. Second, since both orientations occur within each aperture, the information from the two normal velocity signals is *overlapping*, and would therefore be capable of being combined whatever the upper limit on spatial separation for combining signals. Moreover, these displays are closer in appearance to the classical combination of drifting sinusoidal gratings of differing orientations.

Methods

The three factors of contour orientation bias, direction of true motion, and prevailing translatory direction were just as in Expt 1. Unlike Expt 1, however, each aperture contained contours of both sampled orientations (see Fig. 9). Except for the manipulation of identifiable feature points, all display parameters were exactly as in Expt 1, resulting in 30 different stimulus conditions. The same three observers participated in 5 experimental sessions, each consisting of 4 trials of each of the 30 conditions.

Results and discussion

The data of Expt 2 are summarized in Fig. 10. Results for rightward and leftward motion were virtually identical, and have been pooled. While the results shown in Fig. 10 were pooled over observers, the results for

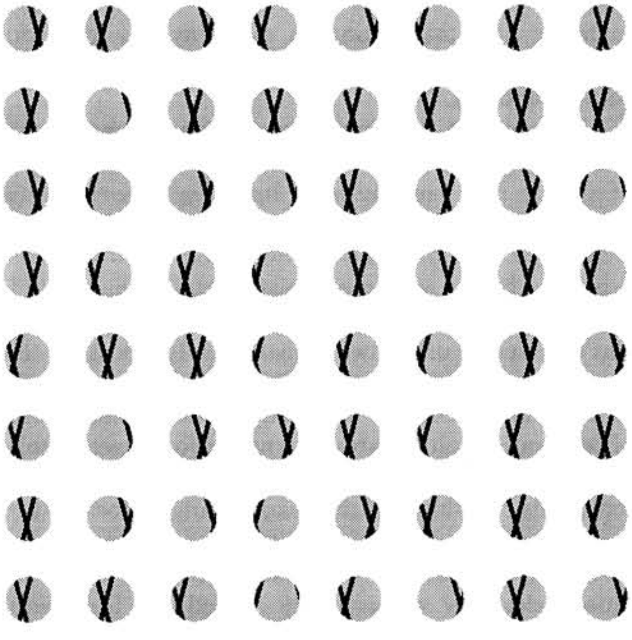


FIGURE 9. A depiction of a frame from a motion sequence of Expt 2. Each aperture contains contours of two orientations, 15° to either side of the vertical, resulting in an unbiased distribution of orientations. Each aperture explicitly depicts the intersection of its contours at some time within the motion frame sequence.

each of the observers followed the same pattern. Note that, as in Expt 1, the presence of orientational bias made the observers virtually incapable of detecting the true motion direction. For unbiased trials, however, observers are rather accurate for 30 and $-30°$ motion.

The results of Expt 2 confirm that the poor performance in Expt 1 cannot be attributed to the distance between or topological separation of apertures. That is, the arguments that information from the two components was too far apart to be combined through the intersection-of-constraints, or that the contours of the apertures themselves served to suggest in some way that motion signals should not be integrated across apertures, do not apply for Expt 2. For this experiment, contours overlap within *each* aperture, but performance still falls far short of the prediction of the intersection-of-constraints solution.

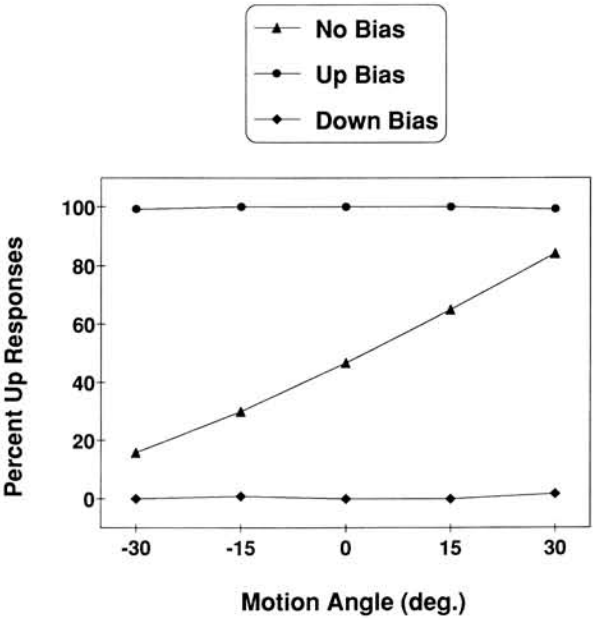


FIGURE 10. Results of Expt 2, pooled for all observers and for prevailing direction of motion. Percent up responses is plotted against displayed motion direction; true motion is downward for angles of -30 and $-15°$ and upward for corresponding positive angles.

Recalling the discussion of the phenomenal appearance of the displays of Expt 1, we note that for those displays of Expt 2 with the largest differences in component velocity differences, it was possible to see a "sliding" motion of one component direction over a different one. Subjects in this case based their judgments on the dominant perceived direction.

A naive account might hold that the presence of *any* trackable features within each aperture of the display ought to make the observers' task trivial. One can in principle recover a veridical velocity signal within each aperture, and since that signal is identical for each aperture in the experimental paradigm, there is no need to rely on intersection-of-constraints or any other heuristic for combining component directions to make the required experimental report. Clearly observers are incapable of acting in this manner. Any signal that may be generated from such features is being overwhelmed, at least in the case of biased displays, by signals sensitive to *orientation* of the contours. This latter tendency is not unreasonable, insofar as features such as contour crossings, while perfectly trackable in the optical domain, may not be of any particular utility. For example, they may not correspond to the same physical surface point over time, as is often the case with intersections involving occlusion contours.

The results of this experiment are consistent with those of Derrington and Suero (1991), regarding the influence of perceived component speed on perceived direction of moving plaid patterns and of Welch (1989), regarding the influence of component speed on discrimination thresholds. Our study provides yet another confirmation of the important contribution of an early, orientationally sensitive "component motion" process (Adelson & Movshon, 1982). There is some evidence to suggest, however, that the presence of identifiable points of intersection did have a small effect on the observers' performance in the unbiased condition. Note in Fig. 10 that the level of accuracy for this condition was considerably higher than for the comparable condition of Expt 1, in which no identifiable feature points were present (see Fig. 7). Note however, that a much stronger effect of identifiable feature points was obtained in the features present condition of Expt 1, where the alignment of features would have provided a weaker stimulus to the orientationally tuned mechanism involved in the detection of component motion.

It seems that the perceived direction of moving patterns may involve a trade-off between a vector average of the component contour motions and the trajectories of identifiable features. The evidence suggests that observers' perceptions are primarily determined by the component motions when the contours in a display are highly salient, but that the motions of identifiable feature points can also have a strong effect under appropriate conditions. It is especially interesting to note, while considering this hypothesis, that a similar form of compromise solution has recently been proposed by Ferrera and Wilson (1990). Instead of suggesting a trade-off between tracking of identifiable feature points and vector averaging of component directions, they suggest that a process that computes an intersection-of-constraints solution may compete with vector averaging of component directions. It is important to keep in mind, however, that the potential information provided by an intersection-of-constraints solution in their experiments was always confounded by the presence of identical information from the motion of identifiable features (i.e. the luminance "blobs" formed at the intersections of overlapping sinusoidal gratings). Since our data indicate that observers are incapable of perceiving a velocity space intersection-of-constraints solution for patterns of moving contours that do not contain identifiable feature points, and that the correct direction of motion can be determined from the motions of feature points in the absence of contours, we believe that a more plausible interpretation of the Ferrera and Wilson (1990) data is that observers employed a compromise solution between vector averaging of component contour directions and the tracking of individual feature points.

In a subsequent study Ferrera and Wilson (1991) measured the perceived speed of plaid displays in a discrimination paradigm in which subjects were asked to compare the speed of plaid test patterns with the speed of standard patterns composed of single patterns. This study is of particular importance in the investigation of the extent to which global detection of motion is based on feature tracking, vector averaging, or intersection-of-constraints, because the vector average velocity for two components is generally lower than the intersection-of-constraints velocity, as illustrated in Figs 1(b) and 8. (The extent of the difference varies widely with the particular component orientations and speeds chosen.) Ferrera and Wilson (1991) found that for all plaids, whether Type I or Type II, the perceived speed was slower than that predicted by the intersection-of-constraints, whenever the standard pattern was a grating of the same spatial frequency as the components of the plaid. When compared with a grating having the same spatial period as the nodes of a plaid, however, the plaid was seen as moving at the speed predicted by the intersection of constraints. The Ferrera and Wilson (1991) results afford difficulties in interpretation, since the measurement of perceived speeds is so dependent on the comparison used, but it is interesting to note that the deviation from the intersection-of-constraints prediction in the case of a comparison grating of the same spatial frequency as the components of the plaid is *toward* the value of the speed of the components. Moreover, the results for the cases where the comparison grating had the same spatial period as the nodes of the plaid suggest that *the nodes themselves* can generate some code for speed independent of either vector averaging or the intersection-of-constraints, and that the speed of the nodes was rendered more "accessible" to the subject by the presentation of a comparison grating of the same spatial period as that of the nodes. We note in this context that the results of Adelson and Movshon (1982) and Movshon *et al.* (1985), which are sometimes read as having "ruled out" tracking of nodes as a perceptual strategy for plaids, merely demonstrated that some two stage process involving the initial detection of signals from oriented components *was* at work. The latter results did not—and logically could not—rule out an additional contribution, of greater or lesser importance depending on the perceptual task, for feature tracking. Moreover these results are as consistent with vector averaging as with the intersection-of-constraints, insofar as those computations yielded the same direction for all the displays used in their studies.

Why would observers not take advantage of the potential information provided by a velocity space intersection-of-constraints solution when viewing displays such as those of Expts 1 and 2? Upon reflection, this result may not be perplexing. We noted earlier that the elegant simplicity of the geometric construction of an intersection-of-constraints solution may have lulled researchers into assuming that the required measurements might be *easy* for a visual system, whereas a consideration of interactions among oriented detectors that might perform the first stage measure revealed a paradox: either the detectors must be well separated in perceptual space in order to get a clean measure of component motions, in which case it is not clear when those component motions should be combined, or the two components must be measured by overlapping or proximal detectors, in which case accuracy is sacrificed because of contamination of oriented contrast signals at line crossings or corners.

There are additional reasons for doubting the perceptual plausibility of the intersection-of-constraints solution. Consider that not all visual motion is translatory, and the intersection-of-constraints solution is only applicable to rigid translations. The visual system can in general have no *a priori* criterion for determining to what extent a given contour's local motion signal is part of a larger pattern of translation—as opposed to, for example, a rotation or expansion about some point. On the other hand, the displays of Expts 1 and 2 are characterized by dozens of closely packed regions generating local motion signals. A key consideration for integrating so many distinct motion signals is likely to be noise suppression, in the sense of disregarding small local deviations from a consistent larger pattern, a task for which vector averaging is clearly suited. For any scene containing contours of a number of relatively homogeneously distributed orientations, the vector average of component motions will yield an accurate resulting signal for translatory motion. It is only for sparsely sampled orientations, particularly for biased samples such as used in our experiments, that vector averaging is likely to yield an erroneous result.

EXPERIMENT 3

If the hypothesis that the visual system performs an early pooling of component motion signals through vector averaging is correct, the next functional task would appear to be to interpret local patterns of velocities (optic flow). Considerable work has been done in computational analyses of schemes to recover categories or styles of motion, including rotation, expansion (contraction), and shear, as well as translation (Koenderink & van Doorn, 1977; Uras, Girosi, Verri & Torre, 1988; Waxman, 1984). Little is known, however, about the determination of motion patterns other than translation under conditions of restricted sampling of local contour orientations. The confound that local contour orientation generates in the optical signal with respect to an object's actual velocity can affect the measure of all these styles of motion, not just translation. We propose that vector averaging may be a general conditioning step for the analysis of several types of motion patterns, as opposed to the intersection-of-constraints solution, which at best would resolve only a limited class of contour patterns and only for translatory motion.

A central motivation for Expts 3 and 4 was the work of Koenderink and van Doorn (1977), who proposed that the visual system could decompose optic flow into components of curl, divergence, and shear through the use of simple detectors that are sensitive to certain combinations of local velocity signals, as indicated in Fig. 11. These detectors can be thought of as templates; a sufficiently good match of measured velocities with preferred velocities generates a signal that the detector's preferred style of motion exists in that region. The goodness of match might be described by the dot product of the local measured velocity vector with a unit vector in the detector's preferred direction at each location. Just as in considerations of the aperture problem for translatory motion, the local measured velocity may be merely the component of true velocity that is normal to some extended edge or contour moving through the region of sensitivity of one of the detector's lobes. Thus questions of how to combine different velocity signals in nearby locations, or of how orientational bias affects perception, such as were explored in Expts 1 and 2, have natural analogs in the domains of rotary motion or expansion and contraction (see Shiffrar & Pavel, 1991; Meyer & Dougherty, 1987). For example, configurations of moving edges such as found in a rotating spiral would be expected to produce an erroneous expansion response for detectors such as shown in Fig. 11.

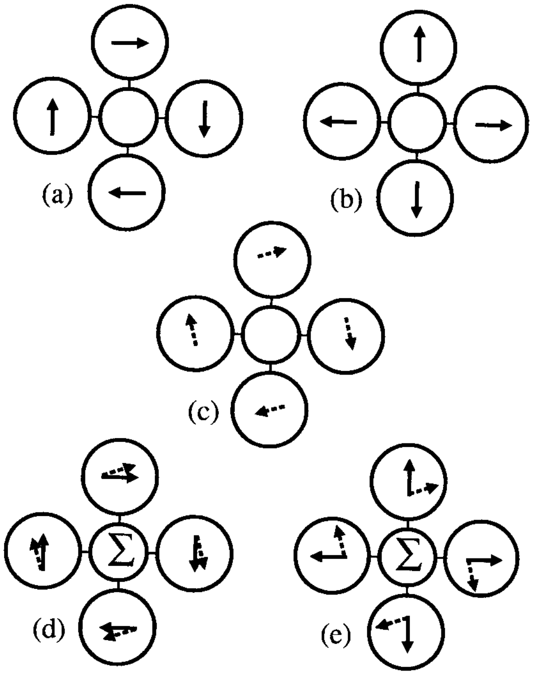


FIGURE 11. Two of the motion detectors proposed by Koenderink and van Doorn (1977). (a) This detector computes estimates of clockwise rotation about its center, while that shown in (b) computes estimates of expansion. Arrows indicate preferred motion direction in various spatial positions. Similar arrays of detectors would respond to contraction, counterclockwise rotation, translation and shear in various orientations. (c) Dashed lines indicate local measured velocity. (d) The detectors' responses are proportional to the sum of some function (e.g. dot product) of the angles formed between a unit vector in the detectors preferred directions (solid arrows) and the local measured velocity (dashed arrows). The rotation detector (d) would give a larger response than the expansion detector (e) for the input shown in (c).

Methods

Just as Expt 1 tested observers' ability to detect upward or downward components of motion in displays whose dominant motion was translation in a roughly horizontal direction, Expt 3 tested the corresponding ability for detecting components of divergence (expansion or contraction) about the center of displays whose dominant motion was rotary. We sought to create displays that would prove detectors such as shown in Fig. 11 in a manner that was as consistent as possible with the methods of Expts 1 and 2 (see Fig. 12).

To describe the experimental manipulations, it is useful to regard each aperture from a privileged Cartesian coordinate system, one of whose axes passes through the center of the entire display and the center of the aperture, and second of which bisects the aperture perpendicular to the first (see Fig. 13). The global motions displayed in Expt 3 can be understood as follows. If trajectories of contours within each aperture were everywhere tangent to the circle formed by a ring of aperture centers, a veridical rotation would result. A "rotation" display in Expt 3 approximated this pattern, with the proviso that motion of a contour within any aperture was at all times translatory. Thus, our rotation displays were "piecewise" composed of translations, but those translations formed a global approximation of rotary motion. Not all our displays simulated rotary motion, however. Just as we had displayed patterns of translation in Expt 1 that deviated by 15 or 30° from the horizontal, we similarly displayed patterns in which the motion within each aperture deviated by 15 or 30° from the tangent direction for that aperture, with the further restriction that those deviations were consistent for all apertures in a display. As a result, we generated displays

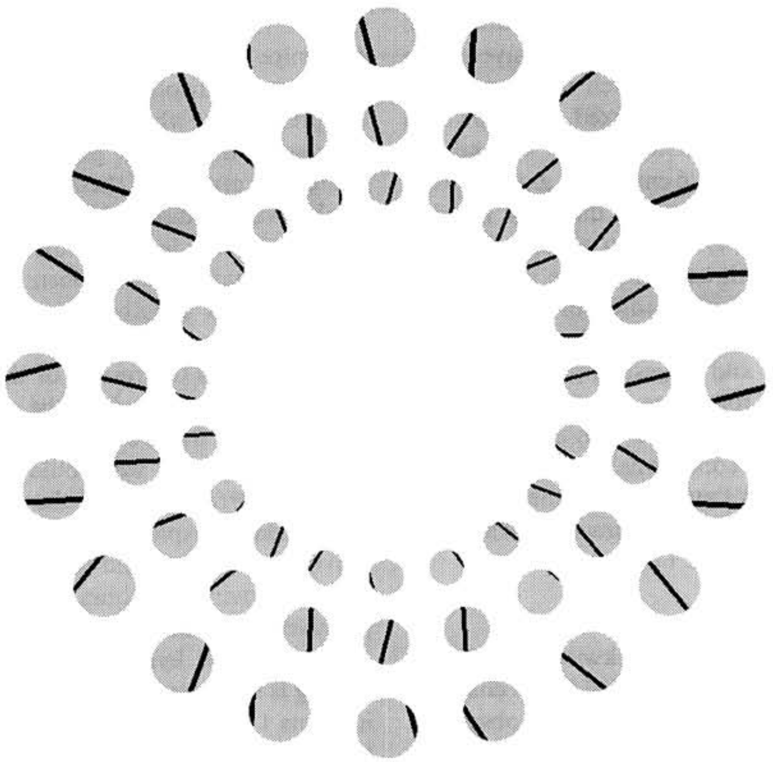


FIGURE 12. An example frame from an "unbiased, features absent" trial of Expt 3. See the Methods section of Expt 3 for details on display construction.

depicting two levels (15 and 30°) of "true" expansion and two corresponding levels of true contraction.

Also in analogy with displays of Expt 1, the orientations of contours within apertures were always samples from two possible orientations *relative to each aperture's coordinate system* (see Fig. 13). Once again those two orientations could be symmetrically aligned—this time relative to the radial axis, which for each aperture was aligned with the entire display's global center—or asymmetrically aligned. The former yielded unbiased displays, in the sense that there was no orientational bias favoring the perception of contraction or expansion, while the latter yielded biased displays. The effect of orientational bias depended in turn on whether the dominant motion of the display was clockwise or counterclockwise; any component of true contour motion outward along the radial axis would result in an "expansion bias" display, while any component in the opposite direction would result in a "contraction bias" display. As with Expt 1, which of two orientations appeared in a given aperture and the phase of contours within each aperture were chosen at random.

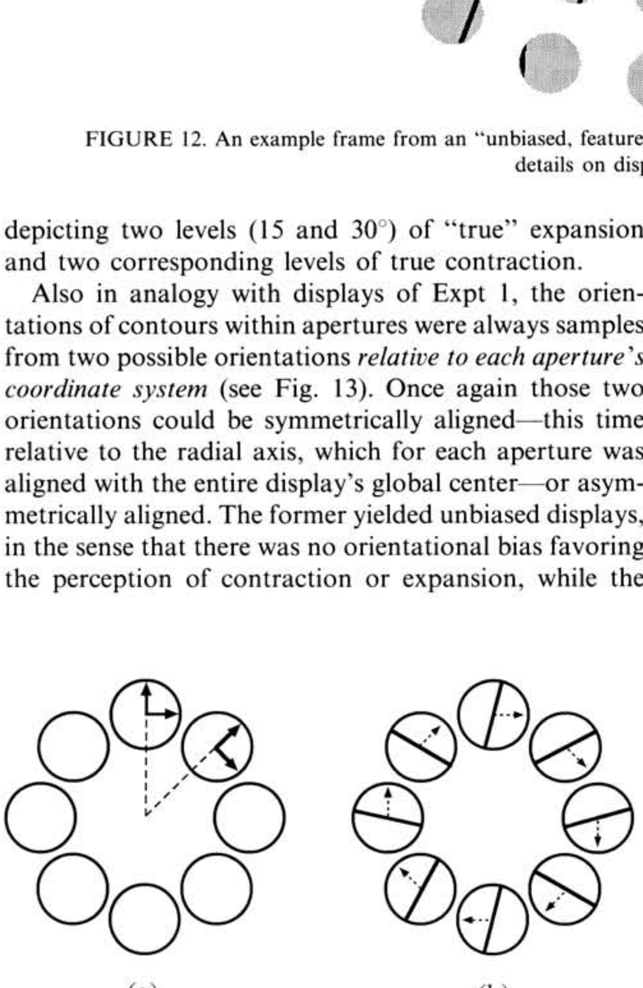


FIGURE 13. (a) Each aperture in a circular array can be assigned its own Cartesian coordinate system, with origin in the center of the aperture, one axis parallel to the radial direction from the center of the array to the center of the aperture, and the other perpendicular to the first. (b) A schematic of an array with unbiased orientations of contours, oriented 15° to one side of the other of each aperture's radial axis, undergoing pure rotation, as indicated by the dotted arrows. Note that each arrow is perpendicular to its aperture's radial axis, though it may not appear so due to static angle effects.

Four factors were orthogonally combined to yield 60 experimental conditions. (1) *Features*, as described in Fig. 6, could be *present or absent*. (2) *Bias* of contour orientation could favor perception of *expansion* or *contraction* or neither; the latter case is referred to as "unbiased". (3) *Style* of true motion could be either pure *rotation*, mild *expansion* or *contraction* (with each aperture's local true direction deviating by 15° from the tangent direction for an aperture), or strong expansion or contraction (with local true directions deviating by 30°). (4) Prevailing rotary direction could be either clockwise or counterclockwise.

Display motion parameters were the same as for Expt 1, except that the local velocities were computed to be consistent with the overall pattern of expansion, rotation, or contraction to be displayed for a given trial. For unbiased trials local orientations of contours were randomly chosen to be 15° to either side of the radial axis orientation for each aperture. For biased trials, the corresponding orientations were 15 or 45° to the same side of that axis. Velocities of the contours and features varied somewhat with orientation and eccentricity, but all fell in the range of 0.7–4.0°/sec.

Instead of the lattice arrangement of apertures used for Expts 1 and 2, for Expt 3 each frame displayed three concentric rings of apertures, with each ring containing 20 apertures whose position was fixed, as shown in Fig. 12. The radii of the rings subtended 3.3, 4.5 and 5.9° at the viewing distance of 76 cm. The gap between apertures was 0.8 times the aperture diameter for that ring. Aperture diameters were chosen so that 20 apertures and 20 gaps exactly fit each ring. The thickness of each contour was 10% of the radius of the aperture containing that contour, subtending 3.6′, 4.7′, and 5.9′ for the smallest, middle, and largest rings, respectively.

The same observers as for Expts 1 and 2 participated in 5 experimental sessions. Each session consisted of 4 trials of each of the 60 conditions. Procedures were identical to those for Expts 1 and 2, except that observers indicated perceived expansion or contraction in a forced choice by pressing appropriate buttons.

Results and discussion

Data for clockwise and counterclockwise motion were virtually identical, and have been pooled in the charts of Fig. 14. While the results were pooled over observers, each observers' results displayed the same pattern. Clearly, the observers were unable to reliably report whether the true motion contained components of expansion or rotation in the absence of features, regardless of bias condition, and could report true motion for all bias conditions when features were present. Expressed another way, in the "features absent" condition, most of the experimental variance was accounted for by the bias condition. Expansion bias displays were reported as expanding, and contraction bias displays as contracting, regardless of the true motion. The presence of features in this experiment rendered the true motion obvious and observers response was essentially perfect. Just as in Expt 1, the presence of features enabled observers to determine the true motion, and, also as in Expt 1, the perception of displays lacking features was controlled by the presence or absence of orientational bias.

From a theoretical point of view one of the most interesting aspects of these data is the apparent inability of observers to discriminate expansion from contraction in the unbiased, features absent condition. Suppose, for example, that the component motions of all contours were summed together in the manner described by Koenderink and van Doorn (1977) (see Fig. 11). For the biased displays the results of this summation would be completely determined by the contour orientation bias, regardless of the true pattern of motion, but for the unbiased displays we would expect to obtain more accurate performance. If the true motion contained a component of expansion, for example, then the outward moving contours would have a greater velocity than the inward moving contours, which ought to result in a positive divergence (expansion) signal.

One possible hypothesis that could account for this finding is suggested by the results for translatory motion in Expts 1 and 2. Suppose that the motions of neighboring contours were averaged together prior to performing the divergence summation. Since the local vector averages in the unbiased displays would deviate only slightly from the tangential direction in each local region (as shown in Fig. 8) the magnitude of the divergence signal would be greatly attenuated.

The remarkable similarity between the structure of the data in Expts 1 and 3 strongly suggests that similar mechanisms may be involved in the detection of translation, rotation, and divergence. Our working hypothesis is that the perceptual analysis of these globally coherent patterns of motion involves three distinct stages. The first stage is the detection of local motion signals, either

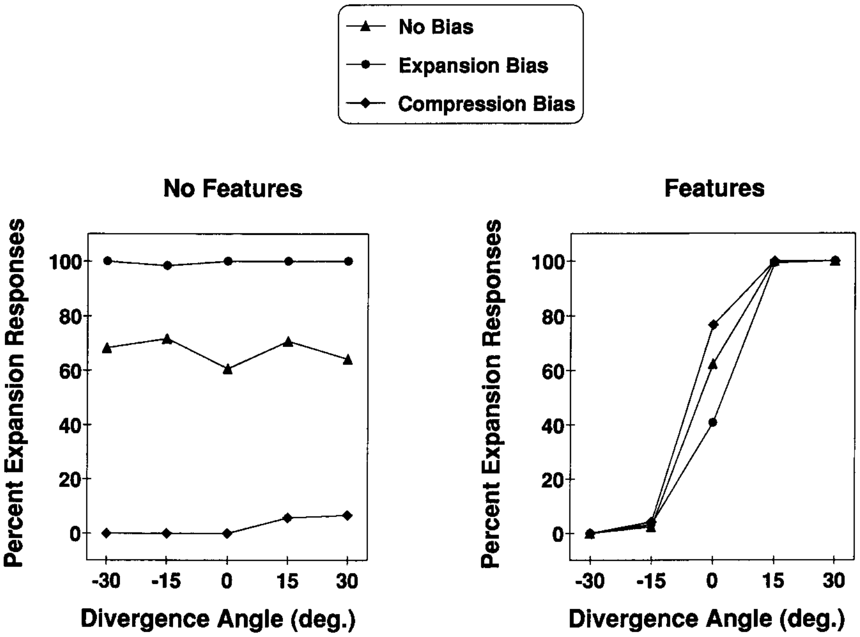


FIGURE 14. Results of Expt 3, pooled for all observers and for prevailing direction of motion. Percent expansion responses are plotted against displayed motion direction; true motion is expanding for positive angles and contracting for negative angles.

from oriented contours or from unoriented feature points. In the second stage the motion signals within some local spatiotemporal domain are averaged together, which in most natural circumstances would effectively suppress noise. Finally, in the third stage these local averages are filtered by an appropriate set of differential operators similar to those proposed by Koenderink and van Doorn (1977) to identify specific patterns of motion.

EXPERIMENT 4

We hypothesized that a common early mechanism, involving stages of local motion detection and vector averaging, subserves perception of rotation and expansion as well as translation. If so, creation of displays with features formed by intersection of contours should produce a similar pattern of results for displays involving rotation, expansion and contraction as were found in Expt 2 for translation displays.

Methods

The display parameters were identical to those used in Expt 3, except that there was no manipulation of the presence or absence of features; instead contours of two orientations were displayed within each aperture, resulting in the formation of intersection features. Thus with three different orientation biases and 10 different true directions of motion, there were 30 distinct experimental conditions. These were presented to the same three observers over 5 experimental sessions, each of which included 4 trials of every condition.

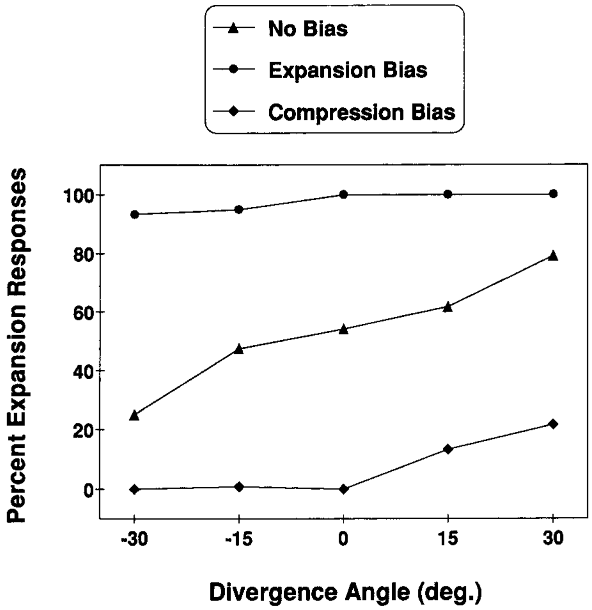


FIGURE 15.[1] Results of Expt 4, pooled for all observers and for prevailing direction of motion. Percent expansion responses is plotted against displayed motion direction; true motion is expanding for positive angles. As in Expt 3, observers' responses are primarily driven by the presence of orientational bias, despite the presence of a trackable intersection feature. Performance for unbiased displays is better than for the displays without features of Expt 3, but not as good as for displays with contour features (see Fig. 14).

Results and discussion

The combined data for all three observers are presented in Fig. 15. Note that the overall pattern of results was identical to that obtained in the earlier experiments, in that performance was primarily determined by the direction of contour orientation, irrespective of the true direction of motion. When there was an expansion bias, observers responded "expansion" on almost every trial; when there was a contraction bias, they rsponded "contraction" on almost every trial; and when there was no bias, the displays generally appeared as pure rotation.

These findings confirm the growing body of evidence from a number of laboratories that the motions of identifiable points of intersection can often be overwhelmed by the component motions of individual contours in determining an observer's perceived direction of motion. It should again be noted, however, that these component motions seem to be combined using a process of local vector averaging, as opposed to the intersection-of-constraints solution suggested by other investigators.

GENERAL DISCUSSION

The research described in the present article was designed to investigate several basic issues involved in perceptual analysis of globally coherent motion. The first of these issues concerned the perceptual significance of identifiable feature points. The significance of the motions of identifiable feature points is that their trajectories provide potential information about the true direction of overall pattern motion, whereas the motions of individual straight contours are inherently ambiguous because of the aperture problem. Our experiments employed several different conditions in which the relative perceptual salience of contours and features were systematically manipulated. In some displays the moving contours were presented in isolation, with no identifiable features at all. In others, pairs of contours were presented together with a single identifiable point of intersection. Finally, a third set of displays was employed, in which each aperture contained a pair of moving rectangles. The motions of the individual rectangles in this case would provide an appropriate stimulus for feature tracking, while the pairs of rectangles considered as a group would presumably stimulate orientationally tuned mechanisms.

With respect to perceptual salience in our own displays, we note that in the features present conditions of Expts 1 and 3, the alignment of features presumably produced a relatively weak signal from the orientationally tuned mechanisms and had little or no effect on the observers' perceptions. When pairs of overlapping contours were presented in Expts 2 and 4, on the other hand, the presence of identifiable points of intersection apparently produced a relatively weak signal from the feature tracking mechanisms, and the observers' perceptions were primarily determined by the motions of the component contours.

The perceptual salience of identifiable features in classical plaid patterns probably falls somewhere between the salience of the contour intersections (weak) and isolated features (strong) in our experiments. The intersections of contours in our experiment were purely geometric features, in that the luminance of each intersection was identical to that of each component contour. The intersections formed at the peaks of overlapping sinusoidal gratings, on the other hand, have twice the luminance of the peaks of the isolated component gratings. Because the intersection points in our displays had the same contrast with aperture interiors as the component contours, it is reasonable to conclude that the perceptual salience of the intersections would be diminished relative to that of intersections of sinusoidal gratings. We strongly suspect, therefore, that the presence of identifiable features is likely to have a stronger influence on perceived directions of moving plaid patterns than did intersections in our experiments (see Ferrera & Wilson, 1990).

The results of our experiments provide strong evidence that the perception of motion in human observers can sometimes by dominated by the motions of identifiable features, and at other times be dominated by the motions of the component contours, depending of which aspect of the stimulus pattern has a greater perceptual salience. Of course our use of the word "salience" is a circumlocution for a host of possible experimental conditions or task demands. For example Yo and Wilson (1990) recently reported that Type II plaids appear to move close to the vector sum direction when viewed in the periphery or briefly in the fovea. De Valois and De Valois (1990) also reported differences in global organization for peripheral and foveal viewing of the similar moving patterns composed of moving patches of luminance modulated in the form of Gabor functions. Also, Boulton and Baker (1990) have reported that ability to process motion signals coherently across frames varies with spatial frequency content of displays. These studies indicate that spatial separation between components or elements of a pattern can be an important variable in determining the salience afforded those elements in determining coherent motion. Our discussion of spatial separation in our displays was not intended to minimize the importance of this variable in general, but rather to argue that spatial separation between components in different apertures in Expts 1 and 3 could not be deemed a sufficient excuse for the discrepancy between the results and the predictions of the intersection-of-constraints solution.

Another important consideration with regard to the interpretation of our experiments is that some feature points cannot in principle provide useful information about the true direction of pattern motion. Consider, for example, the points in the present experiments where a moving contour was occluded by the aperture boundaries. A recent experiment by Shimojo, Silverman, and Nakayama (1989) provides a dramatic example of how the interpretation of such points can influence observers' perceptions. These authors showed that when a pattern of parallel contours is presented stereoscopically to appear behind an aperture, the contour terminators are perceptually analyzed as points of occlusion and have little or no effect on the perceived direction of motion. In effect the "barberpole illusion" is abolished. If however the same pattern is presented stereoscopically to appear in front of an aperture, then the points of termination are perceptually analyzed as identifiable feature points, and the completely dominate the perceived direction of motion. This finding provides strong evidence that the perceptual classification of feature points can have a critical influence on how they interact with smooth contours to determine the perceived direction of motion.

Our experiments also addressed the issue of how the movements of different contours are combined to determine the perceived direction of motion when there is no perceptually salient information available from identifiable features. One possible method of combination suggested by Adelson and Movshon (1982) and Movshon *et al.* (1985) involves an intersection-of-constraints solution in velocity space (see Fig. 1). Recent experiments by Ferrera and Wilson (1987, 1990, 1991) have provided compelling evidence, however, that a velocity space solution cannot provide a complete account of observers' perceptions of moving patterns. For certain types of displays, which they refer to as Type II motion (as in Fig. 2) observers seem to adopt a compromise between a velocity space solution based on intersection-of-constraints and a simple vector average of the individual component motions. Although this is one possible interpretation of their data, it should also be kept in mind that the displays employed by Ferrera and Wilson all contained identifiable features, where the peaks and troughs of component sinusoidal gratings intersected on another. Thus, another possible interpretation of their data is that observers adopted a compromise between a vector average of the component motions and the motions of the identifiable features, and that a velocity space solution played no role at all in determining the perceived direction of motion.

Our experiments were designed to resolve the confound of identical solutions found by intersection-of-constraints and by tracking features. We did so by designing displays in which there were no identifiable features, so that the only possible source of information about the true direction of motion would be the relative velocities of the individual component contours. The perceived directions of motion for those displays were perfectly consistent with what would be expected from a pure vector average of the individual component motions. That is, there was no evidence to indicate that a velocity space solution had any effect at all on the observers' judgment. Although a process of vector averaging may appear in this context to be a surprisingly inaccurate method of perceptually resolving the aperture problem, it does have some desirable properties. Vector averaging is an efficient technique for noise suppression, to eliminate spurious motion signals that frequently arise in natural vision due to chance spatiotemporal correlations between unrelated contours. Moreover, unless there is a systematic bias in the distribution of contour orientations, as in the present experiments, vector averaging can also provide a relatively accurate measure of the true direction of motion in any local region.

Understanding of the processes for detection of coherent motion within regions of the visual field remains a challenge to perceptual psychology. Considerable attention has been devoted to the early detection of motion signals, and a type of consensual model for an early detector, schematized in Nakayama's (1985) review and in Spillman and Werner's (1989) textbook holds that simple, contrast energy-driven filters can provide initial velocity signals, subject to the caveats of the aperture problem (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Watson & Ahumada, 1985). Others have proposed somewhat different forms for early filters (Grossberg & Rudd, 1989; Marr & Ullman, 1981), and schemes for combining the information from several detectors within local regions (Grossberg & Mingolla, 1990; Marshall, 1990b; Sereno, 1986, 1990). Whatever the ultimate deposition of the question of early filtering, clearly several fundamental tasks remain for the visual system to perform in order to appropriately segment, group, and identify styles of coherent motion in regions of the visual field. We suggest that vector averaging of component motion signals plays an important role in these subsequent processes.

In proposing the vector averaging hypothesis, we do not wish to suggest that all motion signals within a region must necessarily be averaged together. The phenomenon of motion transparency clearly indicates that such cannot be the case. Also, the visual system clearly ought not to average motion signals across object boundaries. Finally, the identification of styles of motion, such as translation, rotation or expansion, is probably simultaneously carried out at several spatial scales. As suggested by Johansson (1950), relative and common components of motions may be extracted by simultaneous interactions of detectors of different kinds of motions at different scales. We also note that just as the averaging of signal occurs over larger spatial region than the initial generation of motion signals, the evaluation of motion styles occurs at a still larger scale, comprising a pattern of many potentially different averaged motion signals. Also, while we have spoken of the processes as if executed sequentially, we cannot rule out the possibility of feedback among any of the stages. Indeed it seems natural that the existence of strong signals within some style of motion detector at the third stage may itself be taken as evidence for which signals ought to be pooled or which overridden at the second stage. Moreover, we have spoken of "regions" in a loose way in the present paragraph, and it may well be that the detectors at several stages attempt simultaneously to resolve the input motion patterns at several scales, and that the strength of activation of detectors responsible for smaller or larger regions at one stage is itself a contributing factor in processing at other stages. We are also cautioned by results such as those of Shimojo *et al.* (1989), showing that patterns of two-dimensional motion may be decomposed in strikingly different ways depending on how evidence for depth relations, including binocular disparity, helps determine which signals are to be grouped with which. Finally, we note that Ferrera and Wilson (1991) have recently extended their analysis of compound grating motion into the domain of speed perception. They report that observers estimates of pattern speed do not correspond precisely to an intersection-of-constraints prediction, but the relation of the vector averaging hypothesis to speed perception requires further study.

To summarize, our results and those of several other researchers can be understood within the following framework. We believe that the measurement of visual motion is accomplished in three broad stages: (1) *Determination of local motion signals*, within what we term a "stage" there may be distinct processes, some of which recover the normal component of velocity from extended contours or edges and others of which determine the velocity of identifiable and trackable features, such as terminators or corners. (2) *Pooling of local velocity signals through vector averaging*, since velocity signals may be contaminated by a variety of causes, including shadows, surface highlights, or occlusions, a smoothing of directional information is performed before further analysis. A possible complication at this stage concerns the integration of direction information derived from tracking identifiable features with information from extended contours. (3) *Evaluation of global evidence for styles of motion* (*translation, rotation, expansion or contraction, and shear*) *within a region*, averaged signals from the third stage can then be processed by detectors such as those displayed in Fig. 11, in order to determine what styles of motion characterize a region.

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.

Adelson, T. & Movshon, J. A. (1982). Phenomenal occurrence of moving visual patterns. *Nature, 300*, 523–525.

Boulton, J. C. & Baker, C. L. (1991). Motion detection is dependent on spatial frequency not size. *Vision Research, 31*, 77–87.

Derrington, A. & Suero, M. (1991). Motion of complex patterns is computed from the perceived motions of their components. *Vision Research, 31*, 139–149.

De Valois, R. & De Valois, K. K. (1990). Stationary and moving gabor plaids. *Investigative Ophthalmology and Visual Science (Suppl.), 31*, 171.

Fennema, C. L. & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing, 9*, 301–315.

Ferrera, V. P. & Wilson, H. R. (1987). Direction specific masking and the analysis of motion in two dimensions. *Vision Research, 27*, 1783–1796.

Ferrera, V. P. & Wilson, H. R. (1990). Perceived direction of moving two-dimensional patterns. *Vision Research, 30*, 273–287.

Ferrera, V. P. & Wilson, H. R. (1991). Perceived speed of moving two-dimensional patterns. *Vision Research, 31*, 877–894.

Greenlee, M. W. & Magnussen, S. (1988). Interactions among spatial frequency and orientation channels adapted concurrently. *Vision Research, 28,* 1303–1310.

Grossberg, S. & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92,* 173–211.

Grossberg, S. & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics, 38,* 141–171.

Grossberg, S. & Mingolla, E. (1990). Neural dynamics of motion segmentation. *Proceedings of Graphics Interface/Vision Interface '90,* Halifax, Nova Scotia, Canada, May 1990 (pp. 112–119). Toronto: Canadian Information Processing Society.

Grossberg, S. & Rudd, M. (1989). A neural architecture for visual motion perception: Group and element apparent motion. *Neural Networks, 2,* 421–450.

Johansson, G. (1950). *Configurations in event perception.* Uppsala: Almkvist & Wiksell.

Koenderink, J. J. & van Doorn, A. J. (1977). How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. In Hauske, G. and Butendant, E. (Eds), *Kibernetic.* Oldenbourg: Munchen.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B, 211,* 151–180.

Marshall, J. A. (1990a). Adaptive neural methods for multiplexing oriented edges. In Casasent, D. P. (Ed.), *Intelligent robots and computer vision IX: Neural, biological, and 3-D methods.* Proceedings of the Society for Photo-optical Instrumentation Engineers (SPIE) Symposium on Advances in Intelligent Systems 1382, Boston, Mass., November 1990 (pp. 282–291).

Marshall, J. A. (1990b). Self-organizing neural networks for perception of visual motion. *Neural Networks, 3,* 45–74.

McKee, Z. P., Silverman, G. H. & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research, 26,* 609–619.

Meyer, G. (1987). Sawtooth pac people and the realization of illusory edges: Computational, cognitive, and utilitarian implications. Paper presented at the *28th Annual Meeting of the Psychonomic Society,* Seattle, Wash.

Morrone, C., Burr, D. C. & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proceedings of the Royal Society of London B, 216,* 335–354.

Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. (1985). The analysis of moving visual patterns. In Chagas, C., Gattass, R. & Gross, C. (Eds), *Pattern recognition mechanisms.* New York: Springer.

Nakayama, K. (1985). Biological image motion processing: A review. *Vision Research, 25,* 625–660.

van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2,* 300–321.

Sereno, M. I. (1986). Neural network model for the measurement of visual motion. *Journal of the Optical Society of America A, 3,* 72.

Sereno, M. I. (1990). Learning to see rotation and dilation with a Hebb rule. In Touretzky, D. (Ed.), *Advances in neural information processing systems.* San Mateo, Calif.: Morgan Kaufman.

Shiffrar, M. & Pavel, M. (1991). Percepts of rigid motion within and across apertures. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 749–761.

Shimojo, S., Silverman, K. & Nakayama, K. (1989). Occlusion and the solution to the aperture problem. *Vision Research, 29,* 619–626.

Snowden, R. J. (1989). Motions in orthogonal directions a mutually suppressive. *Journal of the Optical Society of America A, 6,* 1096–1101.

Spillman, L. & Werner, J. S. (1990). Visual perception: The neurophysiological foundations. San Diego, Calif.: Academic Press.

Uras, S., Girosi, G. F., Verri, A. & Torre, V. (1988). A computational approach to motion perception. *Biological Cybernetics, 60,* 79–87.

Walters, D. (1987). Rho-space: A neural network for the detection and representation of oriented edges. *Program of the Ninth Annual Conference of the Cognitive Science Society* (pp. 455–460). Hillsdale, N.J.: Erlbaum.

Watson, A. B. & Ahumada, A. J. (1985). Models of human visual-motion sensing. *Journal of the Optical Society of America A, 2,* 322–342.

Waxman, A. M. (1984). An image flow paradigm. In *The proceedings of the workshop on computer vision: Representation and control* (pp. 49–57). Annapolis, Md: IEEE.

Webster, M. A. & De Valois, R. L. (1985). Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *Journal of the Optical Society of America A, 2,* 1124–1132.

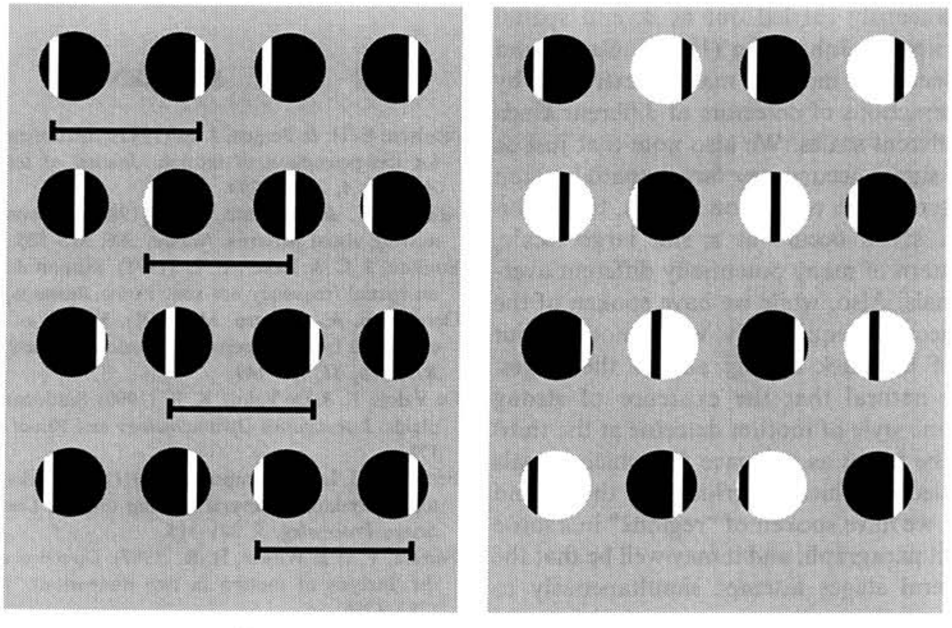

(a) (b)

FIGURE 16. A schematic representation of 5 frames of a motion sequence that displays an unwanted aliasing effect, and its removal through contrast alternation of apertures.

Welch, L. (1989). The perception of moving plaids reveals two motion-processing stages. *Nature*, *337*, 734–736.

Williams, D. & Phillips, G. (1987). Cooperative phenomena in the perception of motion direction. *Journal of the Optical Society of America A*, *4*, 878–885.

Yo, C. & Wilson, H. R. (1990). Perceived direction of 2-D patterns depends on duration, speed, contrast, and visual field quadrant. *Investigative Ophthalmology and Visual Science (Suppl.)*, *31*, 240.

Zucker, S. W., Dobbins, A. & Iverson, L. (1989). Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, *1*, 68–81.

Acknowledgements—Ennio Mingolla was supported in part by AFOSR 90-0175. The authors wish to thank two anonymous reviewers for helpful comments on an earlier draft of this manuscript. James Todd was supported in part by AFOSR 89-0016 and by BNS-8908426, jointly from NSF, ONR, and AFOSR.

APPENDIX

When beginning to design our displays we noticed that, if all aperatures were of the same luminance, a disturbing aliasing effect resulted. The aliasing occurred because of "wrapping around" of contour trajectories, whereby a contour that exited an aperture on one frame reappeared on the opposite side of the aperture on the next frame. The problem is schematically illustrated in Fig. 16(a), which contains four rows of spatially adjacent quadruplets of apertures. From top to bottom, each row represent successive frames or rightward motion, with all contours moving at the same rate. Note that because of the initial placement of the contours (randomized phase within the experiments) the contour in the second aperture from the left overshoots the boundary of its aperture and "wraps around" within its aperture in the second row. Similarly, in the transition from the third row to the fourth, the contour in the third aperture from the left wraps around. A resulting band of low luminance in the aperture backgrounds thereby runs diagonally downward and rightward from the leftmost part of the top row to the rightmost part of the bottom row, as indicated by the horizontal reference bars below each row of apertures. Such bands (or rather rightward moving "blobs" in the actual display) result regardless of initial placement of contours, and move at twice the rate of contour motion. Reversing the contrast of alternate apertures removes this disturbing effect, as shown in Fig. 16(b), without introducing any noticeable "reverse phi" disturbances.