

Color correspondence in apparent motion

MARC GREEN

Carnegie-Mellon University, Pittsburgh, Pennsylvania

To maintain figural identity during motion perception, the visual system must match images over space and time. Correct matching requires a metric for identifying "corresponding" images, those representing the same physical object. To test whether matching is based on achromatic (black/white) polarity and chromatic (red/green) color, observers viewed an ambiguous motion display and judged the path of apparent motion. Matching preserved black/white identity regardless of whether frames were viewed binocularly or dichoptically. Red/green identity was also preserved, but coherence of motion depended in part on the number of frames in the motion sequence and on the background luminance. These results suggest that correspondence is computed by a weighted metric containing terms for image features coded early in visual processing.

When we watch a film, television, or video game, perceived objects neither disintegrate nor melt into one another as they move across the screen. This may seem a mundane and trivial observation, but a complex computational operation underlies the maintenance of figural identity. Consider what the visual system must achieve: the identity and location of physical objects in space must be gained from the retinal image. With a moving scene, the problem becomes one of deciding which images, seen at different places and at different times, represent the same physical object. Some "correspondence" mechanism must compute a similarity metric to match images with different spatial and temporal coordinates. To understand figural identity, therefore, is to understand the meaning of similarity.

Figural identity has commonly been studied with stroboscopic motion. Wertheimer (1912) developed the basic paradigm by rapidly presenting two "frames," each containing a single image. An observer viewed the frames and then evaluated the smoothness of the apparent motion. Numerous experimenters (e.g., Kolers, 1972; Kolers & Pomerantz, 1971; Wertheimer, 1912) hoped to find the basis of similarity by varying the form and color of images across the two frames. Generally, however, motion smoothness was not affected by image similarity.

Ullman (1979) criticized early experiments on two grounds. First, the smoothness criterion may be inappropriate for investigating figural identity, or, as he named it, "the correspondence problem." Although the criticism is valid, studies using a competition paradigm (e.g., Burt & Sperling, 1981; Navon, 1976, 1983) also failed to find any effect of image similarity; when images in frame n had alternate motion paths, there was no tendency to move toward a more similar image in

frame $n + 1$. Second, Ullman criticized early experiments for using relatively complex forms, such as alphabetic characters and geometric shapes. After all, matching might not be based on global form. Instead, images might be represented by a number of primitive dimensions. Matching could occur by first computing similarity independently for each dimension and then summing across dimensions. Viewed in this way, the problem of figural identity becomes the problem of specifying how images are represented in early vision.

In recent experiments, I identified several primitives that dictate motion path. Given similar two-dimensional distances, motion path is biased toward neighbors with similar spatial frequency (Green, 1986a, 1986b), orientation (Green, 1986a, 1986b), and retinal disparity (Green & Odom, 1986). All of these image primitives are represented early in visual processing. Both psychophysical (Braddick, Campbell, & Atkinson, 1978; Regan, 1982) and neurophysiological experiments have demonstrated mechanisms selectively tuned to these dimensions.

There is also evidence that both achromatic (black/white) polarity and chromatic color polarity are coded in early vision. The present study demonstrates that these properties are also primitives in the correspondence process. Furthermore, I show why previous experiments failed to find that motion path could be biased by color similarity.

GENERAL METHOD

The display and procedures were similar to those described elsewhere (Green, 1986b). Briefly, observers viewed a series of frames (Figure 1) presented on a color monitor. Each frame contained two pairs of disks, A and B, located on the rim of an imaginary circle with a "+" in the center. In successive frames, the image positions rotated 45° clockwise (or counterclockwise). Each image in frame n then had two equidistant neighbors in frame $n + 1$. If there were no basis for correspondence matching, then motion would be ambiguous. If an image were selectively matched by one of its neighbors, observers would perceive a coherent circular motion (Green, 1986a, 1986b; Green & Odom, 1986).

Address correspondence to Marc Green, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213.

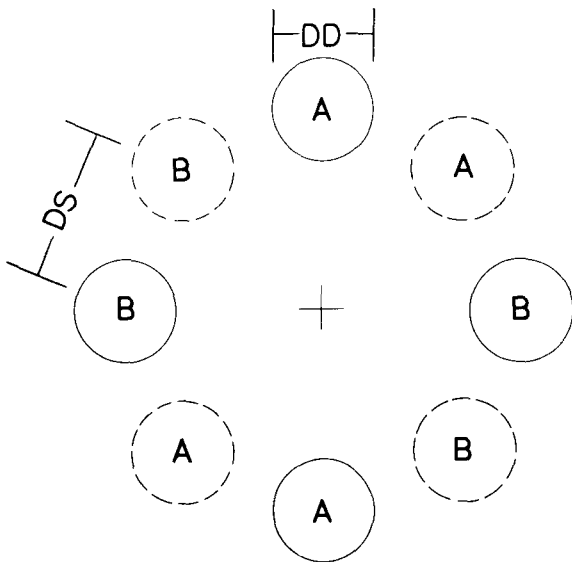


Figure 1. Schematic representation of the ambiguous motion display. Solid and broken lines indicate images displayed in successive frames. DD refers to the disk diameter and DS to the displacement size.

Observers were tested in a series of trials, each consisting, except where noted, of seven frames. Direction of rotation varied randomly from trial to trial. Following each frame sequence, observers made a forced-choice response to indicate clockwise or counter-clockwise motion. I collected data for 2 observers in each experiment, although the effects described below were informally confirmed using many others. Except for myself, all observers were experimentally naive.

EXPERIMENT 1: LUMINANCE POLARITY

Experiment 1 examined whether correspondence could be based on achromatic color (i.e., luminance polarity). The test images were white (97 cd/m²) and black (33 cd/m²) disks viewed on a gray background (65 cd/m²). The luminance was multiplied by a Gaussian weighting function to produce blurred edges, which create smoother motion (cf. von Grünau, 1978).

Display parameters (Figure 1) were set as follows: disk diameter (DD) was approximately 1.5°, frame duration was 84 msec, and interstimulus interval (ISI), during which only the gray field and fixation mark were visible, varied across trials. A second independent variable was disk separation (DS), the center-to-center displacement of a disk from one frame to the next. On half of the trials, one pair of disks (A) were increments and the other were decrements (B). The remaining trials served as a control condition, in which all disks were of the same polarity. Each data point is based on 60 judgments.

Results

On trials with different polarity pairs, observers readily perceived disks moving in a circular path. Figures 2 and

3 show the percentages of trials on which Observers K.C. and M.G. correctly judged motion path. Short displacements produced near-perfect performance at all ISIs. At larger displacements, however, accuracy decreased for brief ISIs. This is consistent with Korte's (1915) third law, which posits an increase in optimal ISI with greater displacements. In the control conditions, performance was at chance levels; motion either appeared ambiguous or reversed direction back and forth.

An additional experiment investigated whether the correspondence mechanism is monocular or binocular. Observers viewed the display haploscopically, so that the left and right eyes saw only alternate frames. Each eye there-

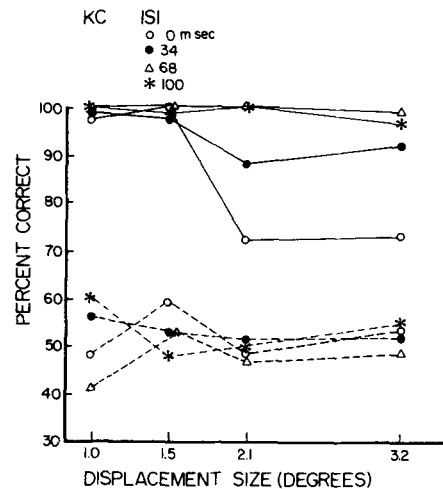


Figure 2. Percent correct judgments by Observer K.C. as a function of displacement size. Solid lines represent data obtained with white/black pairs, and broken lines represent data obtained with disk pairs of the same color.

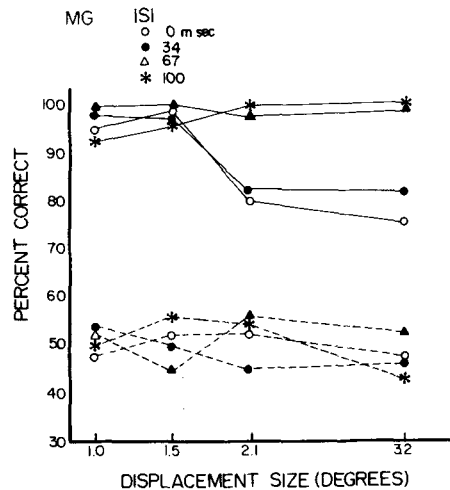


Figure 3. Percent correct judgments by Observer M.G. as a function of displacement size. Solid lines represent data obtained with white/black pairs, and broken lines represent data obtained with disk pairs of the same color.

fore saw 90° rather than 45° rotations of the disks. Since 90° rotations are ambiguous, correspondence could produce coherent motion only by integrating input from the two eyes. This experiment otherwise replicated the original with a step size of 1.5° and an ISI of 34 msec. Observers correctly reported direction of motion on all trials.

Discussion

In apparent motion, an image's path is biased toward a neighbor of the same polarity; increments above the background move to increments and decrements to decrements. This suggests that there are separate visual mechanisms selectively sensitive to luminance polarity. Psychophysical adaptation studies have already hinted at such a possibility. Hanly and MacKay (1979) reasoned that the sawtooth flicker of opposite phases would differentially stimulate increment and decrement detectors. For example, a slow onset and sharp offset should have a greater effect on decrement-sensing mechanisms. Observers were adapted to flicker of one polarity and were tested for sensitivity to flicker of the same or opposite polarity. Adaptation effects proved larger when adapting and test flicker were similar, suggesting that there are separate detectors for increments and decrements. The psychophysical evidence recently has been supported by physiological data (Schiller, Sandell, & Maunsell, 1986) showing that the visual system has separate neuronal pathways for increments and decrements.

Anstis and Mather (1983) also reported evidence that motion path is selected by luminance polarity. Their results showed, however, that path selection was biased only when displacement was small. In contrast, the present experiment found correspondence with displacements as large as 2.1°. Referring to the belief (Braddick, 1974) that there are two motion processing systems, Anstis and Mather concluded that polarity determined path only for the "short-range" motion system. The discrepancy could be explained by the use of much larger stimuli. As has been noted previously (Green & Blake, 1981), the displacement that produces short-range motion is not fixed, but increases with stimulus size.

As an alternate explanation, it might be conjectured that my displays stimulated the "long-range" system. It had been supposed (Anstis, 1978; Braddick, 1974; Green & Blake, 1981) that the long-range system takes binocular input, whereas the short-range system is monocular. Since observers saw dichoptic motion, the display must have stimulated the long-range system. However, Shadlen and Carney (1986) seriously weakened the monocular/binocular criterion by producing dichoptic short-range motion. Either there is only one mechanism, or the short-range system is also binocular. Regardless of which explanation is true, Shadlen and Carney failed to explain why their many-frame technique produced motion while their two-frame displays did not. The next experiment further demonstrates that two- and many-frame motion displays differ qualitatively.

EXPERIMENT 2: CHROMATIC COLOR

Experiment 2 tested for correspondence based on chromatic polarity. The design was similar to the one described above, except that the disks differed in hue; one pair was red and the other was green. The conditions for this experiment varied slightly from the first: (1) each disk had a diameter of 1.3°, (2) the disks' edges were sharp, (3) the independent variable was frame duration, and (4) the ISI was held constant at 0.

Since previous research (Green, 1986b) showed that luminance transients can disrupt correspondence matching, I took the precaution of using two different backgrounds. In some sessions, the disks were presented against a dark field. In others, observers viewed the disks against an isoluminant yellow background. Isoluminance points were determined for each subject prior to the experiment. Observers viewed red and green disks flickering at 15 Hz against a uniform 40-cd/m² yellow background and adjusted luminance to produce minimum flicker.

Results

Figure 4 shows the percentage correct as a function of frame duration. With an isoluminant background, both observers correctly judged the motion path on 90%–100% of the trials. Frame duration had little effect, producing only a slight drop in performance at brief exposures with larger displacements. Correspondence on a dark background was considerably weaker. Motion path was usually ambiguous with no clear direction. On the few trials on which observers perceived some overall direction, the motion was never as smooth or coherent as that seen on an isoluminant background. However, performance was above chance under most conditions.

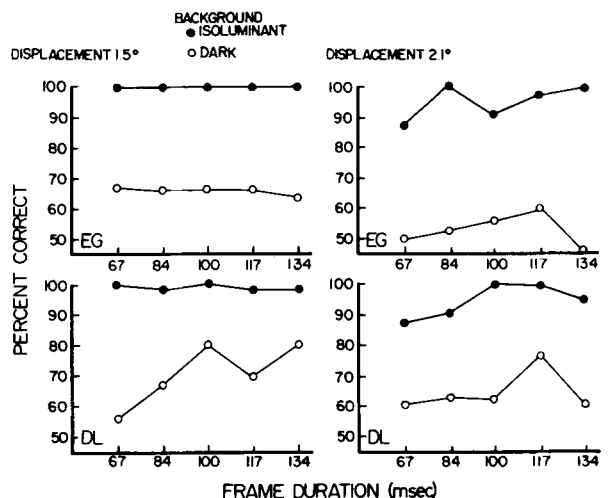


Figure 4. Percent correct judgments as a function of frame duration for red/green disk pairs. Solid dots represent data for judgments made with isoluminant yellow backgrounds, and open dots represent data for a dark background.

Precise isoluminant matches were not required for good correspondence. Although systematic measurements were not made, small variations in the intensity of the yellow, red, and green colors produced no impairment of coherent motion.

Most previous experimenters concluded that chromatic identity does not produce correspondence matching (e.g., Kolers & Green, 1984; Kolers & Pomerantz, 1971; but see Navon, 1983). Although part of my success lay in use of an isoluminant background, observers still performed better than chance with a dark background. There was, however, one other novel aspect to this experiment: most previous displays consisted of two frames, whereas observers here viewed a sequence of seven.

To test whether frame number might account for the remaining difference, the previous experiment was repeated with a frame duration of 100 msec. However, number of frames per sequence varied randomly across trials. Displacement was 1.5° , and 30 trials were run per data point.

Figure 5 shows that frame number had a pronounced effect on correspondence. With a dark background and only two frames, both observers performed near chance. Increasing frame number improved E.G.'s performance, but had no significant effect on D.L. Introduction of an isoluminant background again improved performance, but observers still showed improved performance when frame number increased from two to three.

Discussion

The main finding was that motion path can be based solely on chromatic identity. Given the choice of equidistant neighbors, images move toward the neighbor of the same hue. Path selection was robust, with little effect of either spatial displacement or frame duration.

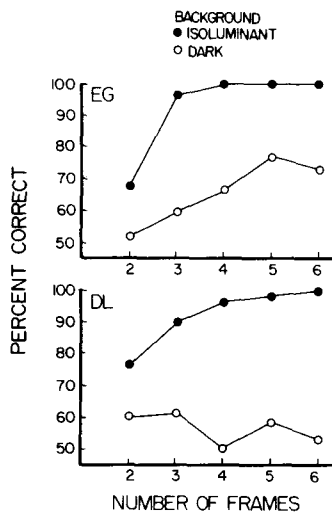


Figure 5. Percent correct judgments as a function of frame number for red/green disk pairs. Closed and open dots represent data for isoluminant and dark backgrounds, respectively.

The experiment also showed why previous experimenters failed to produce correspondence based on chromatic identity. When trying to determine whether an image attribute can determine path, the experimenter should hold all other features constant. As in a previous study (Green, 1986b), dark backgrounds seriously impaired coherent motion. To explain this result, I assume that the visual system contains separate chromatic and achromatic channels, and that red and green spots on an isoluminant background activate only the chromatic channels. When the background is dark, the spots further stimulate achromatic mechanisms. Since the achromatic channel is sensitive only to luminance change, it responds similarly to the red and green targets. The resulting achromatic signal is ambiguous for path selection. However, it is surprising that the chromatic information, which was still available, was insufficient to permit the continuation of perfect correspondence. Presumably, correspondence must be computed by a "preference metric" (Ullman, 1980), which sums similarity values in a number of dimensions such as color, spatial frequency (Green, 1986a, 1986b), orientation (Green 1986a, 1986b; Orlandy, 1940; Ullman, 1980), and disparity (Green & Odom, 1986).

The highly disruptive effects of luminance transients suggest a heavy weighting in the preference metric. Transients stimulate mechanisms tuned to low spatial frequencies, and it seems clear (Green, 1986b; Hochberg & Brooks, 1978; Ramachandran & Anstis, 1987; Ramachandran, Ginsburg, & Anstis, 1983) that low spatial frequencies are highly weighted in the preference metric. There are three possible reasons for this. First, low spatial frequency information is processed faster than high-frequency detail (Breitmeyer, 1975; Green, 1981). Since correspondence is a very rapid process, matches may be made on a first come, first served basis.

Second, given similar bandwidth detectors, low spatial frequency images stimulate larger receptive fields. Path selection is biased toward the closest neighbor (Green & Odom, 1986; Ullman, 1979; Williams & Sekuler, 1984). Suppose that two-dimensional distance is computed, not simply by separation on the retina, but by retinal distance scaled by number of receptive fields. Figure 6 shows images from two successive frames. Given the same retinal distance, image components that stimulate large receptive fields will always have a "shorter" route and might therefore be more heavily weighted.

Third, low spatial frequencies may have an innate conspicuity. Attneave (1974) and Hochberg and Brooks (1978) suggested that the visual system contains separate "what" and "where" systems for marking object location and identity. The what-system is responsible for grabbing and directing attention to locations in the visual field, whereas the where-system performs pattern identification. Since the where-system is most sensitive to low spatial frequencies, luminance transients should be highly conspicuous. Moreover, the where-system is likely mediated

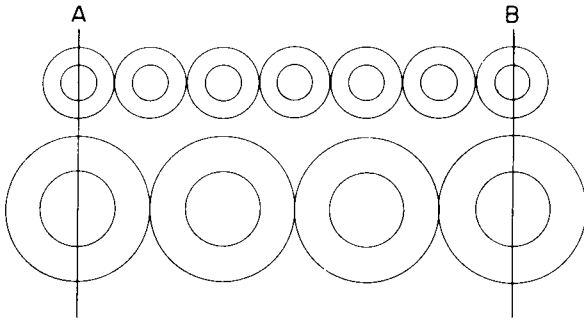


Figure 6. Schematic representation of the effect of receptive field size on "proximity." The distance covered by an object moving from A to B can be scaled by the number of receptive field widths covered.

by receptive fields with little form specificity (Green, 1984).

To compute correspondence, a reasonable procedure might require the following steps: (1) assign the image in frame n values on the dimensions mentioned above, (2) assign values to each potential match in the frame $n+1$, (3) compare the values for the frame n image to those of each potential match and create a preference metric, and (4) select the motion path to the match with highest preference. To perform this procedure, I would need to specify the dimensions that provide terms in the preference metric along with the relative weighting of each. Although I have made progress in identifying the metric's terms, little can be said about relative weights, except that low spatial frequencies are important (Green, 1986b).

The simple algorithm described above does not explain why motion coherence was much weaker with two than with three frames. The great difference between two- and many-frame displays was noted by Sperling (1976), who concluded that it was not merely due to probability summation (among frames) in the many-frame display. Navon (1983) also noted that continuous motion preserved identity better.

What causes the large difference between two- and many-frame displays? One explanation becomes apparent by thinking in the frequency rather than in the space/time domain. In this approach, correspondence is not a discrete process, but rather a continuous one. A visual mechanism, such as a "red detector," can be viewed as a filter. Rather than matching red tokens, it responds to the input as a stream and computes the spatiotemporal frequency spectrum. The spatiotemporal spectra of real and apparent motion differ (Watson & Ahumada, 1983) by the creation of extra spatiotemporal components in the "sampled" apparent motion. Smooth motion can be perceived if the spurious components are filtered away.

With two-frame motion, the resulting spatiotemporal spectrum is broadband, and some energy is likely to fall in quadrants signaling motion in opposite directions. Since chromatic mechanisms have weak temporal and direc-

tional tuning (Cavanagh, Tyler, & Favreau, 1977), they are unable to filter away the spurious components that obscure the true direction. In contrast, the extra components created by many-frame displays are narrowband and more easily separated by filtering.

SUMMARY

Since this is the last in a series of experiments, it is a good time to summarize what has been learned about correspondence. The experiments described here show that path selection is determined by achromatic and chromatic polarity. This adds new features to a list that already includes spatial frequency (Green, 1986a, 1986b), orientation (Green, 1986a, 1986b; Ullman, 1980), and disparity (Green & Odom, 1986). All of these properties have one thing in common: they are represented in the activity of mechanisms in the early stage of visual processing. The *early stage* refers, in the physiological sense, to neurons up to the primary visual cortex and, in the computational sense, to levels at which only data-driven (bottom-up) processing occurs.

This conclusion conflicts with other views of apparent motion. Starting with Exner (1875), many (e.g., Sigman & Rock, 1974; Toch & Ittleson, 1956) have suggested that apparent motion, as well as other forms of perception, involve problem solving, reasoning, and other high-level (top-down) cognitive processes. According to these authors, perception is the process of taking the retinal image and adding information from memory to infer the world events that produced it. Although some experiments have supported this view, others have not. For example, Kolers and Green (1984) found that observers did not interpret ambiguous motion displays by using reasoning or problem solving. Observers frequently preferred "impossible" motions, in which identity changed, to physically realizable ones, in which identity was preserved.

Most authors assume that visual processing requires a hierarchy of representations, starting with primitives such as spatial frequency, color, and so forth. The representation at one stage provides the input for an increasingly abstract representation at the next. The hierarchy has a dividing line, below which the representations are not available to consciousness. For example, Marr (1982) suggests that the retinal image is initially represented by the activity of Laplacean operators of different resolution. The output of these operators is combined to produce a new representation, the "primal sketch." We presumably have no conscious access to the Laplacean operators, but can use the primal sketch directly. According to this model, correspondence matching must occur at a preconscious level. First, it is spatial frequency specific (Green, 1986a, 1986b). Second, correspondence can be based on polarity. Since the primal sketch is only a cartoon, it has neither polarity nor spatial frequency information.

It seems likely that correspondence is computed at a low level, while construction of features available to awareness occurs at a higher level; that is, the early

representation—in terms of spatial frequency, orientation, disparity, color, and so forth—determines correspondence and provides the input to more abstract representations. When we view the apparent motion, correspondence matching occurs at a low level, and the objects we consciously perceive are dragged to a new location. There may be high-level processes responsible for filling in the path; however, these are likely to be mechanisms and sets of primitives separate from those responsible for correspondence.

REFERENCES

- ANSTIS, S. M. (1978). Apparent movement. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. 8, pp. 656-673). New York: Springer.
- ANSTIS, S. M., & MATHER, G. (1983). Effects of luminance and contrast on direction of ambiguous apparent motion. *Perception*, **14**, 167-179.
- ATTNEAVE, F. (1974). Apparent motion and the what-where connection. *Psychologica*, **17**, 108-120.
- BRADDICK, O. (1974). A short-range mechanism in apparent motion. *Vision Research*, **14**, 519-527.
- BRADDICK, O., CAMPBELL, F. W., & ATKINSON, J. (1978). Channels in vision: Basic aspects. In R. Held, H. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. 8, pp. 3-37). New York: Springer.
- BREITMEYER, B. (1975). Simple reaction times as a measure of the temporal response properties of the transient and sustained channels. *Vision Research*, **15**, 1115-1159.
- BURT, P., & SPERLING, G. (1981). Time, distance and feature-trade-offs in visual apparent motion. *Psychological Review*, **88**, 171-195.
- CAVANAGH, P., TYLER, C. W., & FAVREAU, O. E. (1977). Perceived velocity of moving chromatic gratings. *Journal of the Optical Society of America*, **A1**, 893-899.
- EXNER, S. (1875). Über das Sehen von Bewegungen und die Theories das zusammengesetzten Auges. *Sitzungsberichte Akademie Wissenschaft (Wien)*, **75**, 156-190.
- GREEN, M. (1981). Crawford masking of sine-wave gratings. *Vision Research*, **21**, 801-805.
- GREEN, M. (1984). Masking by light and the sustained-transient dichotomy. *Perception & Psychophysics*, **35**, 519-535.
- GREEN, M. (1986a). Correspondence in apparent motion: Defining the heuristics. *Proceedings of Vision Interface '86* (pp. 337-342).
- GREEN, M. (1986b). What determines correspondence strength in apparent motion? *Vision Research*, **26**, 599-607.
- GREEN, M., & BLAKE, R. (1981). Phase effects in monoptic and dichoptic temporal summation: Flicker and motion detection. *Vision Research*, **21**, 362-372.
- GREEN, M., & ODOM, J. V. (1986). Correspondence matching in apparent motion: Evidence for three-dimensional spatial representation. *Science*, **233**, 1427-1429.
- HANLY, M., & MACKAY, D. M. (1979). Polarity sensitive perceptual adaptation to temporal sawtooth modulation of luminance. *Experimental Brain Research*, **35**, 37-40.
- HOCHBERG, J., & BROOKS, V. (1978). The perception of motion pictures. In E. Carterette & M. Friedman (Eds.), *Handbook of perception: Vol. 10. Perceptual ecology* (pp. 259-304). New York: Academic Press.
- KOLERS, P. A. (1972). *Aspects of motion perception*. Oxford: Pergamon Press.
- KOLERS, P. A., & GREEN, M. (1984). Color logic in apparent motion. *Perception*, **13**, 249-254.
- KOLERS, P. A., & POMERANTZ, J. (1971). Figural changes in apparent motion. *Journal of Experimental Psychology*, **87**, 99-108.
- KORTE, A. (1915). Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie*, **72**, 193-206.
- MARR, D. (1982). *Vision*. New York: Freeman.
- NAVON, D. (1976). Irrelevance of figural identity for resolving ambiguities in apparent motion. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 130-138.
- NAVON, D. (1983). Preservation and change of hue, brightness, and form in apparent motion. *Bulletin of the Psychonomic Society*, **21**, 131-134.
- ORLANSKY, J. (1940). The effect of similarity and difference on apparent visual motion. *Archives of Psychology*, Monograph No. 246.
- RAMACHANDRAN, V., & ANSTIS, S. A. (1987). Motion capture anisotropy. *Vision Research*, **27**, 755-762.
- RAMACHANDRAN, V., GINSBURG, A., & ANSTIS, S. (1983). Low spatial frequencies dominate apparent motion. *Perception*, **12**, 457-461.
- REGAN, D. (1982). Visual information channeling normal and disordered vision. *Psychological Review*, **89**, 407-444.
- SCHILLER, P. H., SANDELL, J., & MAUNSELL, J. H. R. (1986). Functions of the ON and OFF channels of the visual system. *Nature*, **322**, 824-825.
- SHADLEN, M., & CARNEY, T. (1986). Mechanisms of human motion perception revealed by a new Cyclopean illusion. *Science*, **232**, 95-97.
- SIGMAN, E., & ROCK, I. (1974). Stroboscopic motion based on perceptual intelligence. *Perception*, **3**, 9-28.
- SPERLING, G. (1976). Movement perception in computer-driven visual displays. *Behavior Research Methods & Instrumentation*, **8**, 144-151.
- TOCH, H. H., & ITTLESON, W. (1956). The role of past experience in apparent motion. *British Journal of Psychology*, **47**, 195-207.
- ULLMAN, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- ULLMAN, S. (1980). The effect of similarity between line segments on the correspondence strength of apparent motion. *Perception*, **9**, 617-626.
- VON GRÜNAU, M. (1978). Interaction between "sustained" and "transient" channels: Form inhibits motion in the human visual system. *Vision Research*, **18**, 197-201.
- WATSON, A. B., & AHUMADA, A. (1983). A look at motion in the frequency domain (*Tech. Mem. 84352*). NASA.
- WERTHEIMER, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, **61**, 161-265.
- WILLIAMS, D. W., & SEKULER, R. W. (1984). Coherent global motion percepts from local motions. *Vision Research*, **24**, 55-62.

(Manuscript received December 19, 1986;
revision accepted for publication July 13, 1988.)