



The Relative Efficacy of Cues for Two-dimensional Shape Perception

MARK NAWROT,*† ELIZABETH SHANNON,†‡ MATTHEW RIZZO*

Received 16 January 1995; in revised form 31 March 1995; in final form 19 June 1995

The visual system uses a variety of cues for form perception, including motion, color, binocular disparity, texture, and luminance. Physiological evidence suggests that these cues are processed by different neural mechanisms. Do the cues processed by some mechanisms convey any advantage for form perception when compared to cues processed by another? In response to this question we assessed the relative efficiency of several cues in conveying two-dimensional form from background noise. For the sake of comparison, every cue type used the same experimental design and stimulus set. Our results confirm that movement is one of the most efficient cues for shape perception. Also, a simple transient cue (an instantaneous flashing on or off) is equally useful. In comparison, local dot density (a type of texture cue) was the least efficient. The efficiencies of most other cues, such as color, stereopsis, and relative movement in noise, were conspicuously similar.

Form Shape Motion Color Neural pathway

INTRODUCTION

The perception of form is a fundamental aspect of vision. Differences in luminance, color, motion, depth, and texture are all possible cues for the delineation of boundaries between objects and background. For instance, a difference in hue and luminance makes it possible to see white clouds against a blue sky or fruit in the trees; motion makes it possible to detect a predator in the forest. Yet it is difficult to say which of these cues is most efficient in conveying two-dimensional form because generally they come into play in different settings. In the absence of a specialized psychophysical paradigm it is difficult to compare very different aspects of visual perception, such as motion perception and color perception.

Importantly, different cues for form perception are thought to depend on processes occurring in different sub-populations of visual neurons. For instance, some cortical neurons exhibit selectivity for oriented contours (Hubel & Wiesel, 1959), color (Zeki, 1973), retinal disparity (Barlow *et al.*, 1967), or direction of motion (Dubner & Zeki, 1971). Visual information is commu-

nicated from the retina to these cortical neurons through a set of pathways that appear somewhat specialized to convey a particular class of visual information. For example, the parvocellular or P-pathway, named for its connections to simian striate cortex (area V1) via parvocellular layers 3–6 of the lateral geniculate body, is characterized by color opponency and slow conducting axons that convey sustained signals (Schiller *et al.*, 1990; Livingstone & Hubel, 1987). This pathway may have stronger projections to secondary areas such as V4 and IT, located in the inferior occipital lobe and adjacent occipito-temporal regions (Felleman & Van Essen, 1991). These regions, along the so-called temporal cortical pathway, are presumed to play a role in the perception of color, luminance, and stereopsis. In contrast, the magnocellular or M-pathway is characterized by large, fast conducting axons conveying information about more transient visual signals. This pathway may have more predominant connections to more dorso-lateral areas in visual association cortex, including areas MT and MST among others. These regions, along the so-called parietal cortical pathway, subservise a role in motion perception.

In a psychophysical parallel to this physiological work, several researchers have used psychophysical methods to distinguish between the so-called transient and sustained channels in human vision (Tolhurst, 1975; Legge, 1978; Kulikowski & Tolhurst, 1973). The transient channel processes high temporal frequency information, which should include flicker and movement. The sustained

*Department of Neurology, 2156 Roy Carver Pavillion, University of Iowa College of Medicine, Iowa City, IA 52242, U.S.A.

†Department of Psychology, Cornell College, Mount Vernon, IA 52314, U.S.A.

‡Present address: Department of Psychology, North Dakota State University, Fargo, ND 58105, U.S.A.

channel processes low temporal frequency and moderate to high spatial frequency luminance information. By analogy, these “psychophysical” channels in humans might correspond to the M- and P-pathways identified in the monkey visual system. Homology issues aside, this conceptual link between the psychophysical and the physiological is an underlying assumption when the properties of physiological mechanisms are probed with psychophysical methods.

Guided by the selectivity of physiological mechanisms in vision, we evaluated the relative efficiencies of numerous potential cues for form perception. Relative motion is thought to be one of the most salient cues for segmenting a visual scene into different forms (e.g. Lamme *et al.*, 1993). On the other hand, Regan and Beverley (1984) found a slight advantage for luminance contrast over motion contrast for a shape perception task although there were other significant differences between the two stimuli used. To address the relative advantages of different cue types in form perception, it is essential both to isolate each particular cue from others and to use the same stimulus design for each of the cue. We adopted a uniform “signal-in-background noise” paradigm to study different cues and to determine their relative efficiencies for form perception.

METHODS

General stimulus design

To determine the relative usefulness of different visual cues for the perception of two-dimensional form, it is essential to use the same stimulus design for each different cue while still isolating each potential cue from all others. Additionally, relative efficacy is easily compared in the form of a “cue percentage” or signal threshold required for form perception with each different cue type. Random-dot stimuli are ideally suited for this purpose because: (i) each cue type can be represented singly within a subset of stimulus dots, (ii) remaining dots comprise a noisy background against which the particular cue is discriminated, and (iii) all alternative cues to form perception can be eliminated or controlled. For example, with motion, a subset of dots (cue dots) might move relative to background dots. Similarly, for color, a subset of dots might be a different hue from the background dots. With this design a simple “light-meter” mechanism could not be used for the form perception task. Instead, detection of these dots from the background requires a mechanism sensitive to each particular cue type. For this reason a blank or dot-free background region would be uninformative. Also, with random-dot stimuli, the proportion of the cue dots needed for accurate shape identification can be measured for each cue type. Therefore, because this design uses the same stimulus set, the same independent variable, and the same experimental design for every condition, the estimate of signal at threshold can be compared, even though the types of visual cues are very different.

The experimental task used here was a five-alternative forced-choice (5AFC) letter shape identification adapted from the procedure of Regan and Hong (1990). The letters were block capital versions of E, H, L, O, and T. These letters were chosen because they could be drawn with horizontal and vertical elements (namely no diagonal elements, as in N or Z). Moreover, these letters all shared many common elements with other letters in the set which eliminated the detection of a single local element as a possible strategy for correct letter identification.

Our operational definition of stimulus signal was the proportion of dots within a letter-shaped target region conveying a particular visual cue. All remaining dots within the target region were called noise dots and were identical in character to all other dots in the background region. In all cases dots conveying signal information were randomly distributed within the letter-shaped target region. For example, consider the case of 0% signal; the target region is identical to the background and the target shape cannot be identified from the visual cues. However, now consider the case of 100% signal; all dots in the letter-shaped target region convey signal information, making the letter shape identifiable in most cases.

The stimulus letter-shaped target region was located within an 8.3×8.3 deg square background area containing 4000 randomly placed 2×2 min arc white dots on a black monitor face. Stimulus letters were constructed from strokes 4 deg in length and 1 deg in width. Each letter was constructed from 2, 3, or 4 of these strokes. Letter-shaped regions appeared either centered in the background, or displaced 1.3 deg diagonally towards one of the corners. The proportion of signal appearing within the letter-shaped target region was varied with the method of constant stimuli with the range of signal for each cue determined from pilot studies. While this range was typically between 10 and 60%, for several cue types this range was shifted so that the performance at the various levels would generate a well defined psychometric function spanning the threshold region. The range used for each cue type is given in each stimulus description below.

Stimulus presentations were 500 msec in duration, except for stereoscopic conditions, which had a stimulus presentation duration of 750 msec to allow additional time to ensure stereoscopic fusion. All testing was done in a dimly lit room (2.4 lx) at a viewing distance of 57 cm. The black and white stimuli were typically of high contrast (100%) but of comparatively low luminance (white = 19 cd/m^2).

Task

In all conditions the experiment used a simple 5AFC letter identification task. Observers fixated a small cross marking the position of the stimulus center before initiating each trial. Responses were given by one of five marked keys. Observers were instructed to respond as accurately as they could, and reaction time was not recorded. Observers were instructed not to change their

fixation during the stimulus presentation as eye movements were not necessary to perform the task, nor did they improve performance. Observers completed the task at their own pace, taking breaks between trials as required. Each observer completed 300 trials, two repetitions of each 150 item stimulus set, in each of the conditions outlined below.

Observers

Ten normal observers, aged between 21 and 37 yr (mean = 25 yr), participated in various portions of the experiment. All observers had normal acuity or were corrected-to-normal. All had normal stereoacuity to Titmus or TMO. All had normal color vision with the exception of one male observer who showed evidence of a deuteranomaly and was deliberately excluded from the color cue condition. A subset of these observers (usually seven or eight) participated in each condition. All observers performed very similarly within each of the conditions.

Overview of stimuli groups

For presentation here, we categorized the stimuli into two groups based on the properties of both cortical and subcortical processing pathways (see Merigan & Maunsell, 1993 for a recent review), functional deficits arising from damage to these pathways (Schiller & Logothetis, 1990; Schiller, 1993), and the properties of the transient and sustained psychophysical channels. Color, texture, and pattern perception are affected by damage to parvocellular layers of the LGN and by damage to the temporal cortical pathway. In contrast, motion and flicker perception are affected by damage to magnocellular layers of the LGN and by damage to the parietal cortical pathway. The selectivity of the perceptual losses, paired with the selectivity of the anatomical damage, suggests that these pathways have a predominant role in processing different characteristics of visual information. Of course, there are a number of additional subtleties in the processing specialization of each pathway. Additionally, there is growing evidence for cross-talk between these neural pathways at cortical levels (Lachica *et al.*, 1992; Ferrera *et al.*, 1992). However, it remains apparent that the P-pathway/sustained channel has a principle role for color and texture perception and the M-pathway/transient channel has a primary role for transient information and for motion perception (Merigan & Maunsell, 1993).

In the current study, motion, transient, and dynamic stereo stimuli are grouped together as being processed predominantly by the transient channel while color, texture, and static stereo stimuli are grouped together as being processed predominantly by the sustained channel. Although stereopsis has been alternately assigned to either the M-pathway (Livingstone & Hubel, 1987) or to the P-pathway (Schiller *et al.*, 1990), but to neither the temporal or parietal cortical pathway (Schiller, 1993), here we divided static and dynamic stereopsis between sustained and transient channels based simply on movement. For a general orientation, schematics of several

different stimulus types are shown in Fig. 1. Designs of the specific stimulus types are outlined below.

Transient channel stimuli

Motion stimuli. Random-dot cinematograms were generated by displacing dots between static, computer generated animation frames. The precise parameters of these displacements were held constant by first generating a distribution of displacements, and then randomly assigning displacements from these distributions to current dot positions, thereby generating new dot positions for the next cinematogram frame. For example, to create a noisy stimulus with random-dot movements, a flat distribution representing all directions of movement was first generated. New dot positions were generated by the random assignment of this distribution to current dot positions. Global stimulus parameters were maintained as a constant and the stimulus had zero net movement due to the flat distribution of directions. Additionally, by changing the movements of individual dots while maintaining global stimulus movement parameters, these stimuli avoided "dynamic occlusion cues" (Regan & Hamstra, 1992; Sáry *et al.*, 1994). These occlusion boundaries, a potential confounding cue to region boundaries (Sáry *et al.*, 1994), are the product of deletion observed when a set of dots follow a fixed path and are eliminated when they cross a fixed region boundary. Specific mechanisms that may subserve boundary formation from these cues have been addressed by several researchers (Andersen & Cortese, 1989; Bruno & Bertamini, 1990; Shipley & Kellman, 1993, 1994). However, in the current study, signal information was distributed randomly throughout the entire target region. The random assignment of displacements to particular dots minimized the occlusion cue, as did the assignment of dots to different distributions as a dot moved over the target/background boundary.

Two general classes of motion stimuli were used. First, stochastic motion stimuli were characterized by a random motion component. Second, uniform motion stimuli were characterized by dots all moving in the same direction at the same speed with no random component intermixed. In stochastic conditions two or more different parameters governed stimulus dot movement. Dots were assigned displacements from either the signal or noise distributions, randomly drawing an assignment from either group on successive frames. Additionally, dots were assigned further characteristics of their displacement [direction or distance (for speed)] from yet another set of signal and noise distributions. With this complex set of assignments individual dots appeared to moved stochastically, but global stimulus parameters were held constant. In stochastic conditions a "momentum conserving" wrap-around technique was used to preserve dot density and eliminate texture and occlusion cues at the target region boundaries. Dots did not cluster or dissipate, appear or disappear along the region boundaries. Instead, single, one-frame displacements that took a particular dot across a boundary also carried the dot in from the boundary on

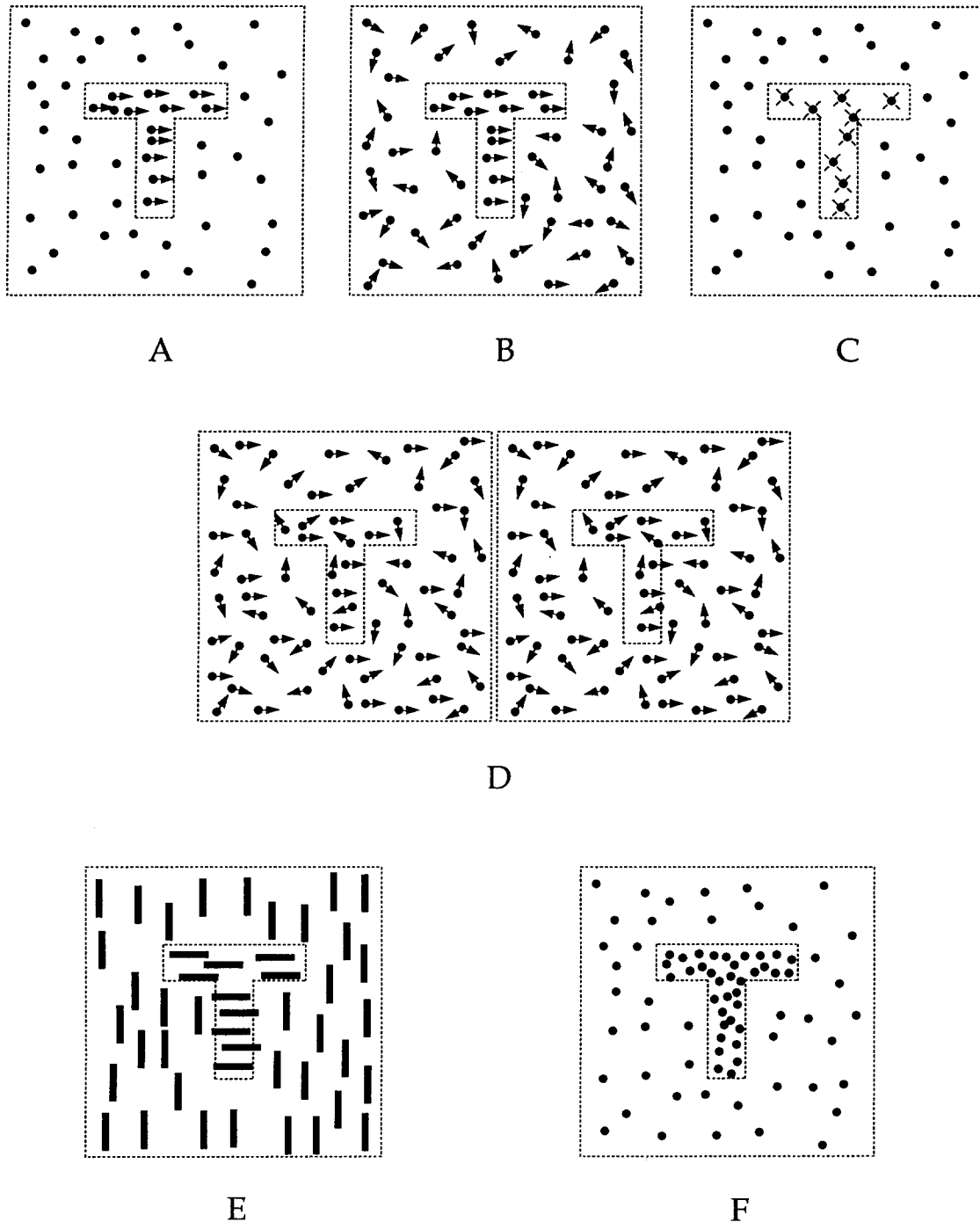


FIGURE 1. Schematics of several stimuli are shown: (A) movement against a stationary background, (B) movement against a noise background, (C) a transient change where dots flash on or off, (D) dynamic binocular stereopsis (this stereo pair is suitable for free-fusion), (E) texture orientation, and (F) local density. The Methods section should be consulted for specific details of stimulus construction.

the other side. That is, the amplitude of dot displacement within the defined target region was conserved.

Uniform motion conditions were created without background noise. Here, a single characteristic of dot movement between the background and target regions was varied. For example, one condition used stationary background dots and moving target signal dots. In the other conditions background dots all moved together at the same velocity, while within the target region a proportion of signal dots moved at a different speed, or in

a different direction (opposite or orthogonal). In uniform motion conditions the “wrap-around” technique from stochastic conditions was not employed. Instead, dots moved unrestrained across stimulus boundaries and assumed new movement properties once outside the target region. In these cases it is important to keep signal proportions low to prevent a confounding texture cue due to the differential dot movements in the background and target regions. For example, dots might be more sparse at the boundary of slow to fast movement while dots would

collect at the boundary from fast to slow movement. In stochastic motion conditions this possible texture cue was eliminated by the random motion components and the wrap-around technique.

In all motion conditions the cinematograms were filmed off-line, prior to the experiments, and saved as bitmap images. Each cinematogram was 33 frames in length. Dot displacements were 6 min arc between frames. With this displacement and a subsequent presentation rate of 30 msec per frame, dots moved at a speed of 3.3 deg/sec. The different motion stimuli are detailed below.

Stochastic motion conditions.

Motion in noise. In this condition the letter shape was defined by coherent movement against a noisy moving background. Here, dots in the background moved in a set of random directions. As mentioned above, this set of random directions had a flat distribution representing all possible directions, thereby producing only local random motion with no global stimulus flow. Signal dots in the target region moved in one cardinal direction. This particular direction was randomly selected for each trial. Again, within the target region, dots could change between signal and noise assignments in successive frames. The signal range was 10–60%.

Motion in opposite directions. In this condition the letter shape was defined by movement in the direction opposite the background movement. Here, dots in both the background and target region were partitioned into signal and noise. Again, noise dots moved in random directions while signal dots moved in one cardinal direction. However, this cardinal direction was opposite for signal dots in the background and signal dots within the target region. The particular directions of movement were randomly selected for each trial. Here, changes in the proportion of signal affected movement in both the background and target region. With 0% signal, all dots moved randomly. In this case, the target region was not visible. However, with 100% signal, dots in the background and target region moved in opposite directions. The range of signal used here was 10–60%.

Motion in orthogonal directions. In this condition the letter shape was defined by movement in a direction orthogonal to the background movement. The stimulus composition here was identical to that of motion in opposite directions (above), with the exception that signal dots in the target region moved in a direction orthogonal to the background. The range of signal was 10–60%.

Motion at different speeds. In this condition the letter shape was defined by movement at a faster speed than the background movement. Here, 75% of dots in both the background and in target region moved in the same cardinal direction. The remaining dots moved randomly. This characteristic of the stimulus was held constant and because it was the same in both the background and target region, it did not make the shape of the target region visible. However, within the target region, some variable

proportion of the signal dots were given larger displacements so that they moved at about 10 deg/sec, three times faster than background dots. This difference in dot movement speed made the target region visible against the background, without becoming an ineffective signal by traversing a small region with large displacements and then “wrapping around”. The range of signal (the proportion of dots moving at 10 deg/sec) was 10–60%.

Uniform motion conditions.

Motion in stationary noise. In this condition the letter shape was defined by movement against a stationary background. For each cinematogram frame, signal dots within the target region moved in one of the cardinal directions. The particular direction of movement was randomized between trials. The particular dots conveying signal information varied between frames. That is, individual dots within the letter-shaped region might make only one or two movements throughout the entire stimulus presentation when they were selected to convey the signal information. Otherwise, these dots were stationary like background dots. The range of signal was 1–10%.

Static signal against uniform motion. This condition is the stimulus reversed version of the previous condition. Here, the letter shape was defined by static dots while the background region was a uniformly moving field moving in one of the cardinal directions. Again, direction of movement was randomized between trials and the particular dots conveying signal information varied between frames. The range of signal was 1–10%.

Uniform motion in opposite directions. In this condition the letter shape was defined by signal movement in the direction opposite the background movement. As no noise motion was used, all dots in the background region moved in the same direction. The particular directions of movement were randomly selected for each trial. On every cinematogram frame, a small proportion of dots within the target region moved in the opposite direction. For instance, with 0% signal all dots moved together in the same direction and the target region is not visible. However, the target region was made visible with only a small proportion of opposite motion in the target region. Again, the particular dots conveying signal information varied between frames. This case is most similar to that used by Regan and Hong (1990). The range of signal used here was 1–10%.

Uniform motion in orthogonal directions. In this condition the letter shape was defined by movement in a direction orthogonal to the background movement. The stimulus composition here was identical to that of uniform motion in opposite directions (above), with the exception that signal dots in the target region moved in a direction orthogonal to the background. The range of signal was 1–10%.

Uniform motion at different speeds. In this condition the letter shape was defined by movement at a faster speed than the background movement. Here, all dots in both the background and the target region moved in

the *same* cardinal direction. Within the target region, some variable proportion of the signal dots moved at the standard speed of 3.3 deg/sec, while the background dots moved at 1 deg/sec. This difference in dot movement speed made the target region visible against the background. Again, the particular dots conveying signal information varied between frames. Also, low dot speeds were needed to prevent a confounding texture cue from the aggregation of dots at the transition between target and background, due to the difference in some dot speeds between the two regions. The range of signal (the proportion of dots moving at 3.3 deg/sec) was 1–10%.

Transient conditions (no motion).

Transient flashing on or off. In these two conditions the letter shape was distinguished from the background by a single, instantaneous accretion or deletion of dots within the target region. Here a proportion of the dots within the target region were added or deleted mid-way through the 500 msec presentation. This is an optimal stimulus for the activation of a transient channel because no onsets or offsets occur for 250 msec before or after the stimulus change (Phillips & Singer, 1974). In the deletion condition dot density is slightly decreased following the change. The accretion condition used the same stimuli, but in reverse order, starting below normal local dot density but restoring normal density following the stimulus change. Dot density was not a serious confound (see Results). The range of signal was 1–10%.

Stereoscopic conditions. The stereoscopic visual system can make use of both moving and stationary dots, with changing or static retinal disparities. Consequently, early stereo processing might proceed through either the transient or sustained channel. That is, the transient and sustained channels might both feed into a single binocular processing mechanism. Alternatively, binocular stereopsis might be a redundant visual processing scheme used in both the transient and sustained channels. To reflect this duality of stereoscopic processing, this study includes conditions for both dynamic and static stereopsis.

Stereoscopic stimuli were presented using a liquid crystal display (LCD) shutter system synchronized to the refresh rate of the monitor. In successive 15 msec intervals, alternate shutters were opened and the appropriate stimulus drawn to the screen. In this way, the left and right eyes viewed a slightly different version of the stimulus. The difference was the binocular retinal disparity used as a cue to target shape in these conditions. Control studies confirmed that these stimuli were devoid of any useful monocular form cues, including cross-talk between the stimuli presented to each eye due to phosphor persistence or incomplete occlusion by the LCD shutters. Also, because each eye had a 30 msec inter-stimulus interval between frame presentations, in dynamic conditions dot movement parameters were identical to those in the other movement conditions. In stereoscopic conditions the binocular disparity of signal dots was 6 min arc while the disparity of background dots

was zero. This magnitude of disparity equated it with the magnitude of dot movements, and this magnitude of disparity was within Panum's fusional area for random-dot stereograms (Fender & Julesz, 1967). While the dynamic condition is outlined immediately below, the static stereo condition is described later with the other sustained channel stimuli.

Dynamic binocular stereopsis. In this condition the letter shape was defined by the retinal disparity in the position of a moving random-dot stimulus. Similarly to the previous motion conditions, both background and target region dots moved at 3.3 deg/sec. Like the earlier speed condition, 75% of the dots in both regions moved in one of the cardinal directions while the remaining dots moved randomly. Again, from this motion alone the target region was indistinguishable from the background. However, here the left and right eyes viewed a slightly different version of the stimulus generating a retinal disparity for a proportion of dots within the target region. These particular dots were randomly chosen in each cinematogram frame, and could move in either the cardinal direction or in a random direction. Observers perceived a stationary letter standing in front of a moving background. The signal range was 10–60%.

Sustained channel stimuli

Color and textural conditions.

Color. In the color condition the letter shape was defined by a difference in isoluminant hue of static random dots. Prior to each set of trials, an equiluminant red and green pair was determined for each observer with the minimum motion technique (Anstis & Cavanagh, 1983). Using a gamma linearized color monitor, the brightness of green was adjusted to match the brightest red possible for that monitor.

Using these two isoluminant hue values for the subsequent experiment, the monitor face was set to an isoluminant yellow midway between the red and green. This replaced the black background used in previous conditions. Background and noise dots were assigned one of the two colors, while signal dots within the target region were assigned the other color. To perform the letter identification task, observers had to determine the shape of the region with dots of a different hue from the background dots. The signal range was 10–60%.

Texture. In this condition the letter shape was defined by a difference in local line orientation. To preserve overall luminance of the stimulus, the 4000 random dots of previous conditions were reorganized into 1000 randomly placed line elements of 8 min arc length. The orientation of these lines was either horizontal or vertical, generating an oriented texture pattern. Signal elements within the target region had a different orientation from the background. The signal range was 10–60%.

Density. In this condition the letter shape was distinguished from the background by increased dot density. Here, the local dot density of each part of the letter shape was determined and increased by some proportion. That is, the normal random-dot density was 0%. Here, it is

obvious that the letter shape would be indistinguishable from the background and the entire stimulus would appear more or less uniform. Doubling the local dot density (100%) within the target region generates a texture cue to the target shape. Additionally, the target region subsequently appears brighter than the background, making local luminance difference a potential confound, but also an integral part of the stimulus. The signal range was 50–150%.

Luminance. In the luminance condition the letter shape was defined by a difference in brightness of stationary random dots. Using a luminance linearized monitor, background and noise dots were assigned the luminance midway between brightest and black. Signal dots within the target region were given the highest luminance. Like most other conditions, the monitor face was black. In this condition the target would appear as a bright letter shape against a set of grey background dots and a black background. These were the largest luminance differences possible on this monitor while using the standard black background. The signal range was 10–60%.

Static binocular stereopsis. In this condition the letter shape was defined by the retinal disparity in the position of static random dots. This condition employed the same LCD shutter system as the dynamic stereopsis condition described earlier. Here all dots were stationary in both the background and target regions. However, a proportion of the dots within the target region were given 6 min arc of crossed retinal disparity. Background dots had zero retinal disparity. Observers perceived crossed disparity dots within the target region as forming a letter shape appearing nearer in depth than the background. The signal range was 5–30%.

RESULTS

Thresholds for detecting letter shape were determined by a probit method for each observer in each condition (Finney, 1971). Mean thresholds are shown in Fig. 2 with percent signal at threshold on the *Y*-axis, and the cue conditions along the *X*-axis. Within each group of conditions the individual cues are shown in ascending order of threshold value. Additionally, a threshold estimate was determined from the psychometric function derived from the average performance of the observers at each level of the independent variable. Due to the similarity in performance across observers, these “average function” thresholds were within 2–3% of the mean thresholds determined above and were often identical. With the current experimental design, this comparison of thresholds actually provides more information than would the comparison of d' values at some set signal proportion.

The thresholds are most easily considered as falling into one of three separate threshold ranges. First, motion against a static background and uniform motion at different speeds demonstrated superlative efficiency for delineating form. Both had thresholds below 1% signal (below hatched region in Fig. 2); observers reported that

these targets are much easier to see than any other types. Second, in the range 1–5% (hatched region in Fig. 2), uniform relative motion and simple transient cues demonstrated very high efficiency for simple form perception. Third, in a range between 15 and 30% (shaded region in Fig. 2), we find the thresholds for most other cues tested here. This includes the various forms of relative movement in noise, stereo, texture, color, and luminance. All of these cues have about the same efficacy for form perception. Finally, the density cue has a threshold well above 30%. This was the least efficient cue for form perception with the paradigm used here, but provides a simple luminance/texture based comparison.

DISCUSSION

This study was designed for direct comparison of thresholds from different visual cues in two-dimensional form perception. A low threshold indicates that the visual system is efficient in the use of that cue in the perception of two-dimensional shape, while a high threshold indicates that the visual system is less efficient in using that cue to perform a two-dimensional shape detection task. Because different types of visual information are processed differently in the brain, we wondered whether one type of cue would show an efficiency advantage for two-dimensional form perception and whether this might tell us about the roles of these cues in form perception. Overall, transient channel cues (in the absence of noise) appear to convey a more efficient source of information about two-dimensional form perception than do sustained channel cues. However, the addition of noise to transient channel stimuli appears to remove this efficiency advantage. While it is not clear how these efficiency differences between cues can be used to augment or constrain a general form perception mechanism such as Shipley and Kellman's (1994) spatiotemporal boundary formation theory, the results obtained here suggest several more specific conclusions regarding the use of various cues in simple form perception.

Motion

The most efficient source of two-dimensional form information is movement against a static or uniformly moving background. With a threshold well below 1% signal, movement against a static background was significantly better than most other types of relative movement in the conditions here (see Fig. 2). The same extraordinary efficiency was not seen for a static figure against a moving background. This lack of reciprocity suggests that movement has a special role in the delineation of figure that is not merely a function of contrast between moving and static dots. Movement is more efficient when it is a particular feature of a figure, rather than a feature of the background.

A small difference in dot speed is another very efficient source of form information. This is especially interesting as in the conditions tested here all dots moved quite uniformly in the same direction. In the conditions here,

even a difference in direction (opposite or orthogonal) is not as efficient a cue to two-dimensional form as a difference in speed. In fact, the “opposite” condition has a larger difference in velocity between figure and background, but was not as useful as a smaller difference in speed in the same direction. This result suggests that the interactions between different velocity tuned motion perception mechanisms have a greater functional significance for form perception than the interactions between mechanisms with different directional tuning. This sensitivity to form information derived from small differences in speed is possibly tied to the perception of depth and surfaces from motion parallax (Rogers & Graham, 1979; Ono & Steinbach, 1990).

The effect of a noisy moving background on form perception appears stable across the different types of relative movement. Background noise increased form detection thresholds by 15–20% compared to conditions

with uniform motion. Indeed, even with a noisy background the visual system appears to make slightly better use of differences in speed than differences in direction.

Simple transient signals

Surprisingly, another very efficient source of two-dimensional shape information is a simple transient signal, which was implemented as an instantaneous accretion or deletion of dots. Here, between 3 and 4% signal was required for shape perception. This threshold compares favorably with the threshold for movement against a static background if one considers the role of probability summation. That is, in the transient condition the observer had to detect a single instantaneous change in the stimulus. However, in the motion condition dot movement was generated by 32 successive changes in the stimulus. This gave observers 31 more opportunities to gather information about the shape of the target region. In

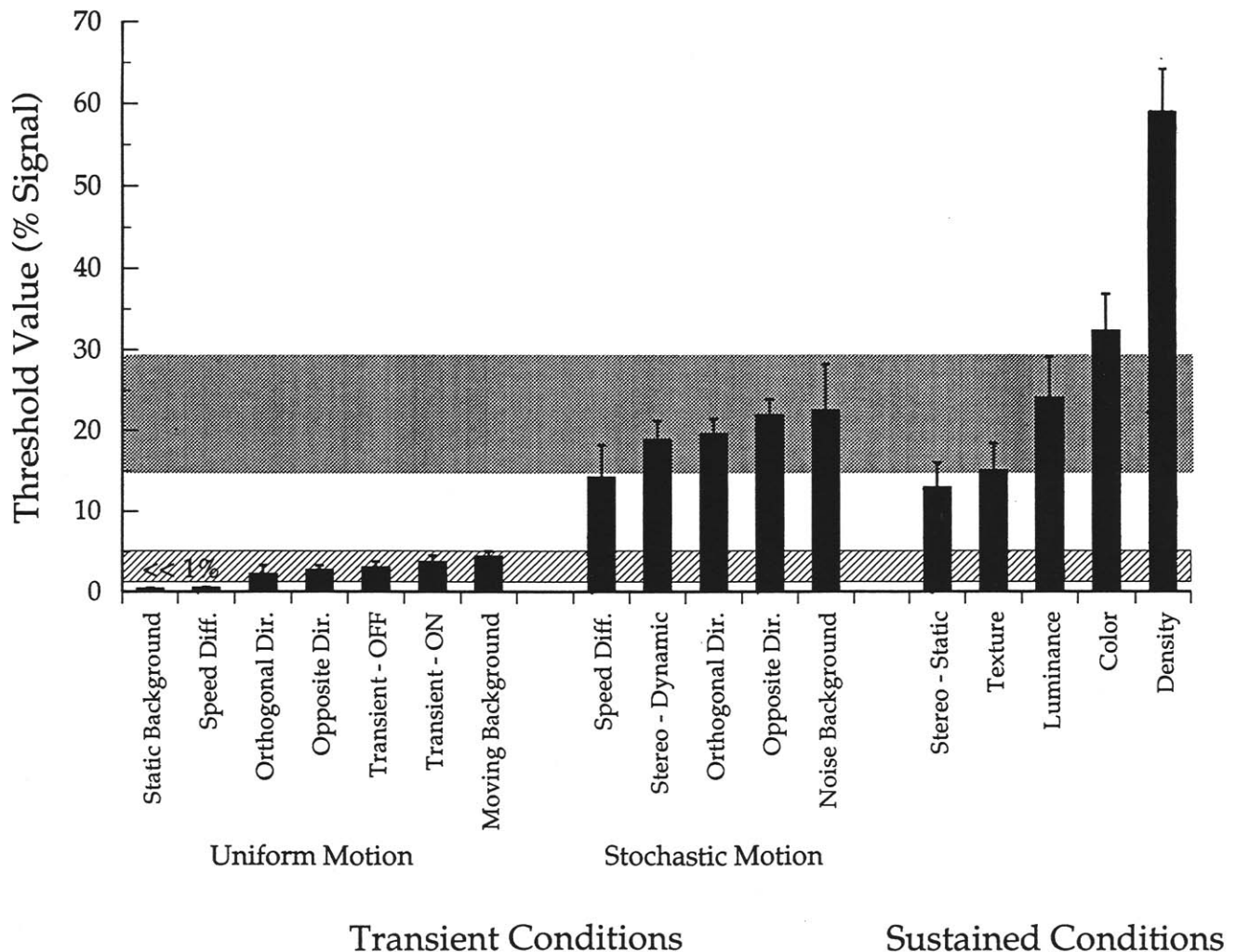


FIGURE 2. The mean thresholds for both transient channel and sustained channel conditions are shown in ascending order of mean threshold value. Error bars denote 1 SE. Most thresholds fall within one of three ranges of threshold values: (i) below 1% signal, which is below the hatched region; (ii) between 1% and 5%, shown by the hatched region; and (iii) between 15% and 30%, denoted by the shaded region. Simple transient signals such as flashes and uniform movement are the most efficient cues for form perception. Density is the least efficient. The mean threshold values for the remaining cues fall within a rather small range around 20% signal.

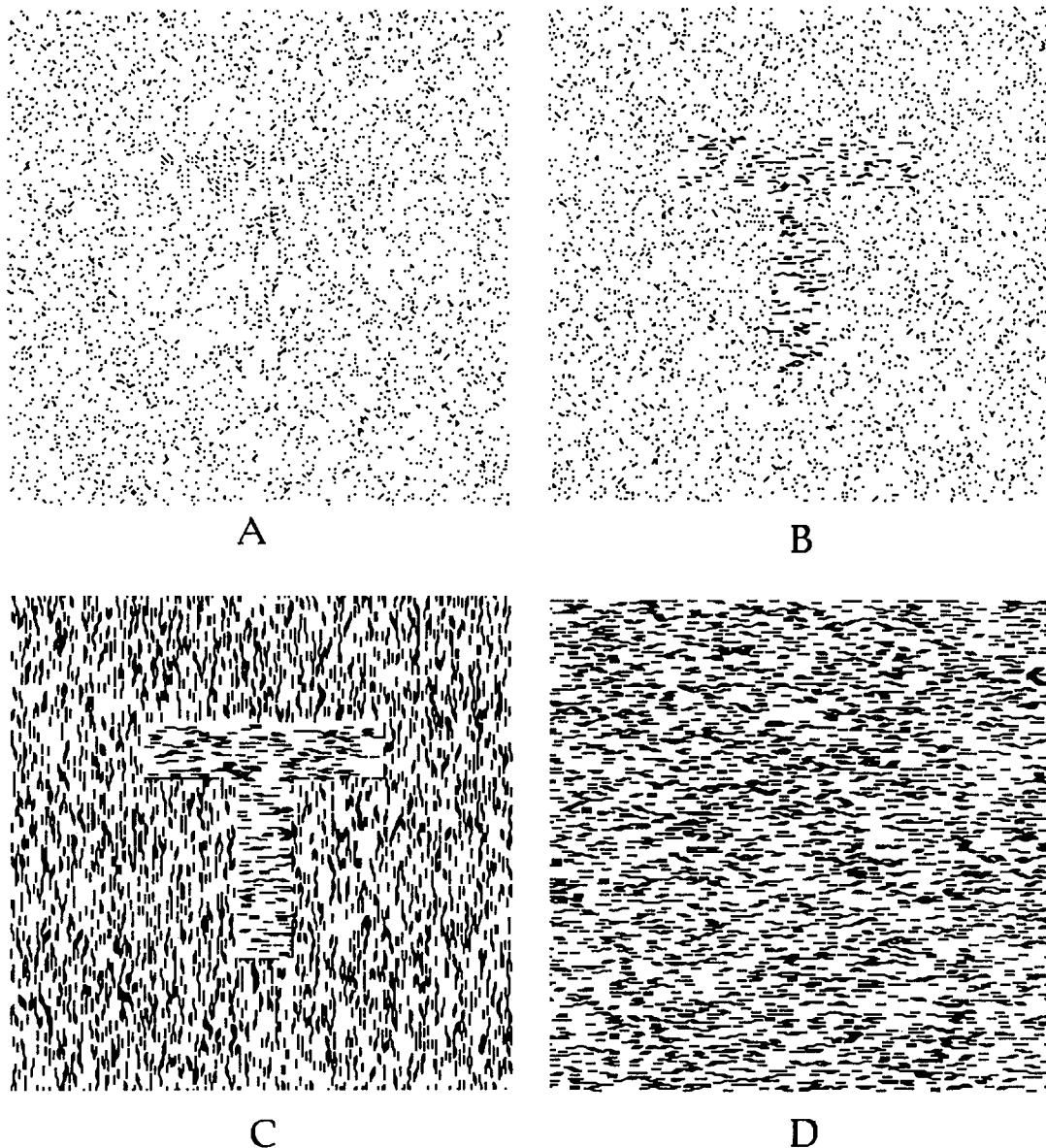


FIGURE 3. Several “temporally integrated” stimulus figures simulating the texture cues accrued across several frames are shown. All figures depict a “T” in the central region of the stimulus. (A) The integration of 30 frames from the motion against a static background condition. Signal proportion within the target region is 2%. While 2% signal is “suprathreshold” for motion cues, no figure is evident from the resulting static texture here. (B) The same stimulus figure with 5% signal and a 2 min arc dot displacement value. These higher parameter values are necessary to generate an appreciable texture cue through this integration. Also shown are stimulus figures created with integration of 5 frames, 100% signal, and 2 min arc dot displacements in the orthogonal (C) and opposite (D) direction of motion conditions. A textural cue for identifying the target shape is generated in the orthogonal case (C), but not in the opposite case (D). As explained in the text, such temporal integration does not occur in normal observers and these texture cues from visual integration or persistence are not used for the shape detection task in motion conditions.

as much as the local information and the visual processing mechanisms are similar for both motion and simple transients, the additional time to gather information appears to generate some benefit in shape detection performance for the motion condition.

No confounding by texture cues

The efficiency of motion in two-dimensional shape perception is not the result of an intermediary texture process, a concern suggested by Regan and Hong (1990)

in their initial stimulus design. A texture cue could be generated at: (i) the visual processing level by a visual integration or persistence process, or (ii) the stimulus level due to inhomogeneities of local dot movements. For instance, consider the spatio-temporal “smear” that would result from dot movement if the display were integrated within the visual system over the presentation duration. This internal smear might generate a texture cue that could be used to perform the shape perception task, even if a motion processing mechanism were not

available. That is unlikely in view of the results here. The very small movement required to accurately perform the task does not allow for an appreciable texture cue. Figure 3(A) shows that no appreciable texture can be generated with even 2% signal over 30 frames in the motion static condition. For comparison, Fig. 3(B) shows the texture when integration occurs over 30 frames with 5% signal and a small dot movement so that a continuous line is generated by these movements.

Additionally, the comparison of performance between the opposite and orthogonal movement directions addresses the issue of internal smear. Again, as Regan and Hong (1990) were careful to point out, integration of dot movements in the orthogonal direction condition would generate a possible texture cue, while the opposite condition would not produce such a texture cue. Figure 3(C, D) demonstrates these two conditions for the motion with noise stimuli. This internal visual smear would produce a difference in performance between conditions of opposite and orthogonal movement. However, the mean threshold values for the opposite and orthogonal motion in noise conditions were only 2% apart, no larger than the standard error of measurement, and a *t*-test on the two sets of thresholds suggests the two distributions of threshold values are not significantly different ($P = 0.399$). The mean threshold values for the opposite and orthogonal movement in uniform motion conditions are similarly indistinguishable. Together, these results show that the visual system is using motion information to perform the task, not an indirect texture cue due to visual integration, persistence, or smear.

Also, the role of a confounding texture cue at the stimulus level was examined in an extensive set of control conditions in which only the last frame of each animation sequence was shown to observers. If a stimulus texture cue were generated by inhomogeneities of local dot movements, this cue should be maximal in the last frame of the animation sequence due to the "accumulation" of dot movements generating this possible cue. With the exception of the condition of uniform motion in orthogonal directions, performance was at chance for all conditions at all signal proportions used here. Observers could not even identify the target position within the background region. However, with uniform motion in orthogonal directions, a small texture cue was generated with some stimuli at the highest 10% signal level [for comparison a 100% signal is shown in Fig. 3(C)]. This texture cue was caused by dots forming a detectable line as they changed direction of movement at the transition between the target and background region. This texture cue was only effective for indicating the general position of letter strokes running parallel to background dot movement. This cue was ineffective for target identification, especially at the < 3% threshold level shown for this condition in Fig. 2.

Binocular stereopsis

Binocular disparity is markedly less useful for simple two-dimensional form perception than is relative motion.

For comparison, the local dot displacements that generated the motion signals and the retinal disparities were of the same 6 min arc magnitude. Even with this large disparity magnitude, both static and dynamic conditions of binocular stereopsis produced thresholds in the range of 15–20% signal. This consistency between stereo conditions is remarkable considering a single set of dots conveyed the disparity information in the static condition, but the set of dots conveying the disparity information changed every 30 msec in the dynamic condition. Additionally, the movement of the background dots in the dynamic condition had no effect; similar thresholds were found with random motion and with uniform motion. While significantly different thresholds in the two conditions would suggest two different mechanisms for static and dynamic stereopsis, the similar thresholds found here might actually suggest a unified underlying mechanism. That is, static and dynamic stereo could be processed by the same mechanism or they could be a redundant information processing strategy found in both the sustained and transient channels: the similar thresholds suggest the former. However, given the similarity of many different cues here, we can only conclude that binocular disparity information is a less efficient source of information about simple two-dimensional form than is uniform motion, regardless of the background in which the disparity information is presented.

CONCLUSIONS

Thresholds for a number of different cue conditions fell within a small range around 20% signal. These conditions include color, texture, and luminance as well as binocular disparity and relative motion with noise. As yet, we have no information on whether this similarity in thresholds is an indication of a single form processing mechanism. Such a mechanism might operate at a relatively high cognitive, attentional, or feature detector level, and require a certain amount of signal information for a form judgment in this paradigm, regardless of the type of information. This general form perception mechanism would probably exist in parallel with the mechanism that recovers form from relative motion, which requires much less signal. If this were the case, the similarity in thresholds of these non-motion cues might be the result of using the same experimental design for all different cues. Alternatively, these different cues could all be processed by different, specialized visual mechanisms that, for some yet undetermined reason, require a similar level of signal in noise that is higher than the signal required for relative motion.

Studies of selective deficits in brain-lesion patients may address this issue. If several different cues rely on the same mechanism, patients with damage to this mechanism should show a generalized deficit across conditions. However, specialized mechanisms are suggested when localized lesions produce cue or modality specific deficits. For instance, deficits for the perception

of shape-from-motion have been demonstrated in patients with mild mental retardation (Oross *et al.*, 1991), parieto-temporal lobe lesions, multiple sclerosis, and even in the good eye of patients with unilateral amblyopia (Regan *et al.*, 1991; 1992; Giaschi *et al.*, 1992a,b). However, the specificity of these deficits to different types of relative motion and the extent of these deficits to other types of cues to two-dimensional form must be determined. For instance, we have found a patient with a small surgical resection of an area in right hemisphere lateral occipito-temporal cortex who, at 2 weeks after surgery, performed normally in the motion against a static background condition but showed a tremendous deficit in the uniform motion at different speeds condition. These are the two conditions in which normal observers perform especially well with thresholds below 1% (see Fig. 2).

Finally, it is remarkable that uniform motion and simple transient cues have a much greater efficiency than the other cue types examined here. The high efficiency of the relative motion conditions suggest a specialized role for higher motion processing mechanisms, such as those isolated to lateral cortical area 18 of the right hemisphere in humans (Orban *et al.*, 1995). However, the high efficiency of simple transient cues suggests this might be a less specialized mechanism that detects simple transient signals, not motion *per se*. Perhaps this transient signal mechanism is found at an early stage in the transient channel, while motion processing occurs at a later stage along this same pathway. Of course, the common high efficiency that these cues share means that very little information is lost between early and late mechanisms if they are arranged in a hierarchical processing stream. As mentioned above, the study of selective deficits in brain-lesioned patients is one way to determine how these mechanisms are arranged in the cortex.

REFERENCES

- Andersen, G. J. & Cortese, J. M. (1989). 2-D contour perception resulting from kinetic occlusion. *Perception & Psychophysics*, *46*, 49–55.
- Anstis, S. M. & Cavanagh, P. (1983). A minimum motion technique for judging equiluminance. In Mollon, J. D. & Sharpe, L. T. (Eds), *Color vision: Physiology and psychophysics* (pp. 155–166). London: Academic Press.
- Barlow, H. B., Blakemore, C. & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *Journal of Physiology*, *193*, 327–342.
- Bruno, N. & Bertamini, M. (1990). Identifying contours from occlusion events. *Perception & Psychophysics*, *48*, 331–342.
- Dubner, R. & Zeki, S. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in monkey. *Brain Research*, *35*, 528–532.
- Felleman, D. J. & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.
- Fender, D. & Julesz, B. (1967). Extension of Panum's visual area in binocularly stabilized vision. *Journal of the Optical Society of America*, *57*, 819–830.
- Ferrera, V. P., Nealey, T. A. & Maunsell, J. H. R. (1992). Mixed parvocellular and magnocellular signals in visual area V4. *Nature*, *358*, 756–758.
- Finney, D. J. (1971). *Probit analysis*. Cambridge: Cambridge University Press.
- Giaschi, D., Regan, D., Kothe, A., Hong, X. H. & Sharpe, J. A. (1992a). Motion-defined letter detection and recognition in patients with multiple sclerosis. *Annals of Neurology*, *31*, 621–628.
- Giaschi, D. E., Regan, D., Kraft, S. P. & Hong, X. H. (1992b). Defective processing of motion-defined form in the fellow eye of patients with unilateral amblyopia. *Investigative Ophthalmology & Visual Science*, *33*, 2483–2489.
- Hubel, D. H. & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, *148*, 574–591.
- Kulikowski, J. J. & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology*, *232*, 149–162.
- Lachica, E. A., Beck, P. D. & Casagrande, V. A. (1992). Parallel pathways in macaque monkey striate cortex: anatomically defined columns in layer III. *Proceedings of the National Academy of Science U.S.A.*, *89*, 3566–3570.
- Lamme, V. A. F., van Dijk, B. W. & Spekreijse, H. (1993). Contours from motion processing occurs in primary visual cortex. *Nature*, *363*, 541–543.
- Legge, G. E. (1978). Sustained and transient mechanisms in human vision: Temporal and spatial properties. *Vision Research*, *18*, 69–81.
- Livingstone, M. S. & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *Journal of Neuroscience*, *7*, 3416–3468.
- Merigan, W. H. & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience*, *16*, 369–402.
- Ono, H. & Steinbach, M. J. (1990). Monocular stereopsis with and without head movement. *Perception & Psychophysics*, *48*, 179–187.
- Orban, G. A., Dupont, P., De Bruyn, B., Vogels, R., Vandenberghe, R. & Mortelmans, L. (1995). A motion area in human visual cortex. *Proceedings of the National Academy of Science U.S.A.*, *92*, 993–997.
- Oross, S., Fox, R. & Carlin, M. (1991). Deficits in kinetic form discrimination by mildly retarded adults. *Investigative Ophthalmology & Visual Science*, *32*, 1283.
- Phillips, W. A. & Singer, W. (1974). Function and interaction of on and off transients in vision. I. Psychophysics. *Experimental Brain Research*, *19*, 493–506.
- Regan, D. & Beverley, K. I. (1984). Figure-ground segregation by motion contrast and by luminance contrast. *Journal of the Optical Society of America A*, *1*, 433–442.
- Regan, D., Giaschi, D., Sharpe, J. A. & Hong, X. H. (1992). Visual processing of motion-defined form: Selective failure in patients with parietotemporal lesions. *Journal of Neuroscience*, *12*, 2198–2210.
- Regan, D. & Hamstra, S. (1992). Dissociation of orientation discrimination from form detection for motion-defined bars and luminance-defined bars: Effects of dot lifetime and presentation duration. *Vision Research*, *32*, 1655–1666.
- Regan, D. & Hong, X. H. (1990). Visual acuity for optotypes made visible by relative motion. *Optometry and Vision Science*, *67*, 49–55.
- Regan, D., Kothe, A. C. & Sharpe, J. A. (1991). Recognition of motion-defined shapes in patients with multiple sclerosis and optic neuritis. *Brain*, *114*, 1129–1155.
- Rogers, B. J. & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*, 125–134.
- Sáry, G., Vogels, R. & Orban, G. (1994). Orientation discrimination of motion-defined gratings. *Vision Research*, *34*, 1331–1334.
- Schiller, P. H. (1993). The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey. *Visual Neuroscience*, *10*, 717–746.
- Schiller, P. H. & Logothetis, N. K. (1990). The color-opponent and broad-band channels of the primate visual system. *Trends in Neuroscience*, *13*, 392–398.
- Schiller, P. H., Logothetis, N. K. & Charles, E. R. (1990). Role of the color-opponent and broad-band channels in vision. *Visual Neuroscience*, *5*, 321–346.

- Shipley, T. F. & Kellman, P. J. (1993). Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion? *Spatial Vision*, 7, 323–339.
- Shipley, T. F. & Kellman, P. J. (1994). Spatiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements. *Journal of Experimental Psychology*, 123, 3–20.
- Stoner, G. R. & Albright, T. D. (1993). Image segmentation cues in motion processing: Implications for modularity in vision. *Journal of Cognitive Neuroscience*, 5, 129–149.
- Tolhurst, D. J. (1975). Sustained and transient channels in human vision. *Vision Research*, 15, 1151–1155.
- Zeki, S. (1973). Color coding in rhesus monkey prestriate cortex. *Brain Research*, 53, 422–427.

Acknowledgements—Presentation of this method was previously made at the Non-Invasive Assessment of the Visual System Meeting of the Optical Society of America in 1992. We thank R. Short for help in data collection. This work was supported by the McDonnell–Pew Program in Cognitive Neuroscience, MRC Canada, and NIH PO NS19623.