Three Systems for Visual Motion Perception

Zhong-Lin Lu and George Sperling

New psychophysical methods enable the isolation and measurement of three systems of human motion perception. The first two are primarily monocular, sensitive, and fast. A third-order system computes motion from a *feature-salience map,* that is, a neural representation of visual space in which the locations of important visual features are marked. The third-order motion system is inherently binocular, insensitive, slow, but highly versatile—it computes motion from all ordinary and many exotic types of stimuli, and it is influenced by attention. This article describes how these systems were isolated and how the relations between them were defined.

For more than 100 years, visual motion perception has been a central problem in perceptual theory. On the one hand, motion appears to involve an early stage of pattern recognition (the "same" pattern must be located first here and then there); on the other hand, motion appears to invoke a unique perceptual experience quite different from pattern or shape perception. Almost from the beginning of the experimental study of motion perception, it has been evident that more than one kind of computation is involved, and there has been a plethora of dual-process motion theories. Although there clearly is a kernel of truth underlying most of these dichotomies and theories, there have been two pervasive problems: No one has been able to obtain a demonstrably pure measure of any proposed mechanism. Nor has there been a clear distinction between the algorithm by which motion is computed and the preprocessing of the visual image prior to the point of motion computation. In this review, we describe how combining a new paradigm (pedestal displays) with several critical subsidiary paradigms (interocular displays, stimulus superpositions with varying phases and directions, alternating-feature stimuli, and attentional manipulations) has enabled us to clarify these issues.[1]

Zhong-Lin Lu is Assistant Researcher in the Department of Cognitive Sciences and Institute for Mathematical Behavioral Sciences at the University of California, Irvine. **George Sperling** is UCI Distinguished Professor of Cognitive Sciences at the University of California, Irvine. Address correspondence to George Sperling, Department of Cognitive Sciences and Institute for Mathematical Behavioral Sciences, SST, University of California, Irvine, CA 92717; e-mail: sperling@uci.edu.

CONCEPTUAL FRAMEWORK

Functional Equivalence of Motion-Energy Detectors and Reichardt Models

Computational theories of motion perception date from Reichardt's model for insect vision,[2] which was adapted for human perception.[3] As illustrated in Figure 1, a Reichardt detector consists of two mirror-image subunits (e.g., "L" and "R") tuned to opposite directions of motion. Subunit R multiplies the signal from spatial location B with the delayed signal from an adjacent spatial location A. When the time for an object to travel from A to B in the external world is the same as the internal delay of the signal traveling from input A to the multiplier R of Figure 1, then the delayed signal from A will arrive coincidentally with the straight-through signal from B, and the resulting large output indicates motion from A to B. Similarly, subunit L multiplies the signal from spatial location A with the delayed signal from spatial location B to indicate B-to-A motion. The direction of movement is indicated by the sign of the difference between the subunits' outputs.

We present the Reichardt model because it is historically first and it is easiest to explain. There are several other ways to compute the same or very similar overall input–output relationships,[4] and they cannot be discriminated by psychophysical paradigms because psychophysics measures only functional input–output relations. This class of equivalent mechanisms has been designated standard motion analysis.[5] We prefer the term motion-energy detection because it is more descriptive. Our pedestal paradigm provides a sensitive test for isolating and measuring motion-energy detection, and we use it to determine which kinds of visual stimuli are and are not processed by a motion-energy computation.

First- and Second-Order Motion

A rigidly moving object is a drifting modulation of luminance.

Published by Cambridge University Press

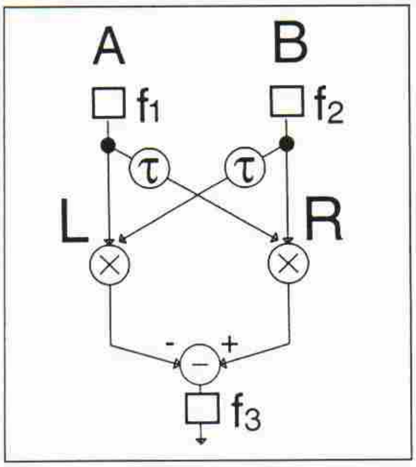

Fig. 1. Reichardt motion detector (simplified). A and B indicate adjacent locations of visual receptive fields, τ is a temporal delay, $\times$ indicates multiplication, and $-$ indicates subtraction; f_1 and f_2 are arbitrary spatiotemporal linear filters (receptive fields); f_3 is an arbitrary linear temporal-integrating filter. Leftward motion (B to A) is computed at L; rightward motion at R. A final output greater than zero indicates stimulus motion from A to B; an output less than zero indicates stimulus motion from B to A.

Motion-energy detection applied directly to drifting modulations of luminance is referred to as a first-order analysis because motion detection operates directly on the input or on a linearly filtered version of the input. (Linear filtering refers to the selective amplification of spatial or temporal frequencies in the input such as might occur, e.g., in a blurring or deblurring computation. In the visual system, such processing undoubtedly occurs prior to motion detection, but it would have no effect on any of the conclusions in our analysis.)

First-order analysis provides reasonable estimates of motion direction for an enormous range of stimuli. However, many investigators[6] have demonstrated clear motion perception in stimuli whose motion would be ambiguous for motion-energy detectors. For example, motion of the classes of *driftbalanced* and *microbalanced* stimuli cannot be extracted by motion-energy detectors. (A driftbalanced stimulus contains exactly the same expected motion energy to the left as to the right for every component spatial frequency in the stimulus, as well as for the stimulus as a whole. A microbalanced stimulus is one that remains driftbalanced even when it is viewed through an aperture of any arbitrary shape.) Such stimuli activate what are called second-order motion mechanisms because a stage of grossly nonlinear preprocessing (e.g., computing the absolute value of the difference of each point from the mean luminance) must occur prior to motion-energy analysis to expose the latent motion.[7]

Pedestal Immunity of Motion-Energy Detectors: Theory

Reichardt detectors have two mathematical properties that prove to be extremely useful for psychophysical experimentation. One is *pseudolinearity:* When a stimulus is composed of several component sine waves with different temporal frequencies, the detector's response to the sum is the sum of its responses to the individual components.[8] The second is that *static displays are ignored,* that is, the output to a stationary pattern is zero. From these properties, it follows that adding a stationary sine (the pedestal pattern) to any moving pattern would not change the output of a motion-energy detector in response to the moving stimulus. This property is the *pedestal immunity* of motion-energy detectors.

PEDESTAL IMMUNITY OF HUMAN OBSERVERS: EXPERIMENTS

The Motion Pedestal Test

We exploit the pseudolinearity of motion-energy detectors by creating compound stimuli (Fig. 2c) that are composed of two components: a stationary sine grating (the pedestal, Fig. 2a) and a linearly moving sine grating (the motion stimulus, Fig. 2b). The pedestal grating consists of stationary alternating light and dark bars for first-order stimuli (Fig. 2d) and of alternating high- and low-contrast texture bars for second-order stimuli (Fig. 2g). The linearly moving sine grating consists of moving light and dark bars for first-order stimuli (Fig. 2e) and moving high- and low-contrast texture bars for second-order stimuli (Fig. 2h). The peaks and valleys of the compound stimuli (Figs. 2f and 2i) wobble back and forth, moving first one way, then the other. Nevertheless, the output of a motion-energy detector is exactly the same for the compound pedestal-plus-motion stimulus as for the motion stimulus alone. (In actual practice, because the response of the early stages of visual processing prior to motion detection is a linear function of—i.e., faithfully represents—the dark-light difference only when the difference is about 5%,[9] the light bars must be no more than 2.5% lighter than the mean luminance, and the dark bars no more than 2.5% darker than the mean luminance.) The question is, how do human observers perceive the compound stimulus? Do they track the peaks (which implies a feature-tracking mechanism), or do they perceive the concealed linear motion of the test stimulus (as they would if their perception were mediated by motion-energy detectors)?

To answer this question, we used the following procedure. Subjects viewed a computer-generated display such as illustrated in Figures 2e and 2h, and reported the direction of apparent movement. In a series of trials, the modulation amplitude of the moving sine was varied. (Modulation amplitude is half the difference between positive and negative sine wave peaks.) We determined the threshold amplitude for 75% correct responses. A pedestal with twice this measured threshold amplitude was then added to the moving stimulus to produce a pedestaled stimulus (Figs. 2f and 2i). If the motion judgment were based on the output of a motion-energy detector, we would expect the subject's accuracy of left versus right judgments to be exactly the same with and without the pedestal. However, if the motion direction computation were based on stimulus features (peaks, valleys, light-dark boundaries, etc.), the pedestaled stimulus would appear to wobble, and it would be impossible for subjects to judge motion direction. We performed this basic experiment with four different types of motion stimuli.

Luminance Grating

The moving luminance grating, which consists of alternating dark and light bars, is the sort of first-order motion stimulus from which traditional motion psychophysics has evolved. Formally, it is a rigidly translating sine grating (Fig. 2e). Luminance stimuli (Figs. 2d, 2e, 2f) are used to determine the properties of the first-order motion system.

Texture-Contrast Grating

The moving texture-contrast grating (Fig. 2h) is a pure second-order stimulus—a binary noise (carrier) whose texture contrast is subjected to a drifting sinusoidal modulation. In this stimulus, all the alternating bars have the same overall or average luminance—there is no luminance modulation between bars. Bars are distinguished from one another by the difference in the microcontrast within each bar. One bar is composed of tiny black and white squares, whereas the other bar is composed of dark gray and light gray squares. Thus, one bar has high contrast between the dots of which it is composed, whereas the other bar has low contrast between the tiny dots of which it is composed. What differentiates the bars is the magnitude of contrast within each bar. The motion of a texture-contrast grating cannot be determined by motion-energy detectors that simply receive luminance inputs. A second-order process is required to expose the contrast grating's motion to motion-energy detection.[6] Essentially, such a process involves *fullwave rectification,* that is, computing the absolute value of each point's deviation from mean luminance so that, for example, extreme black and extreme white

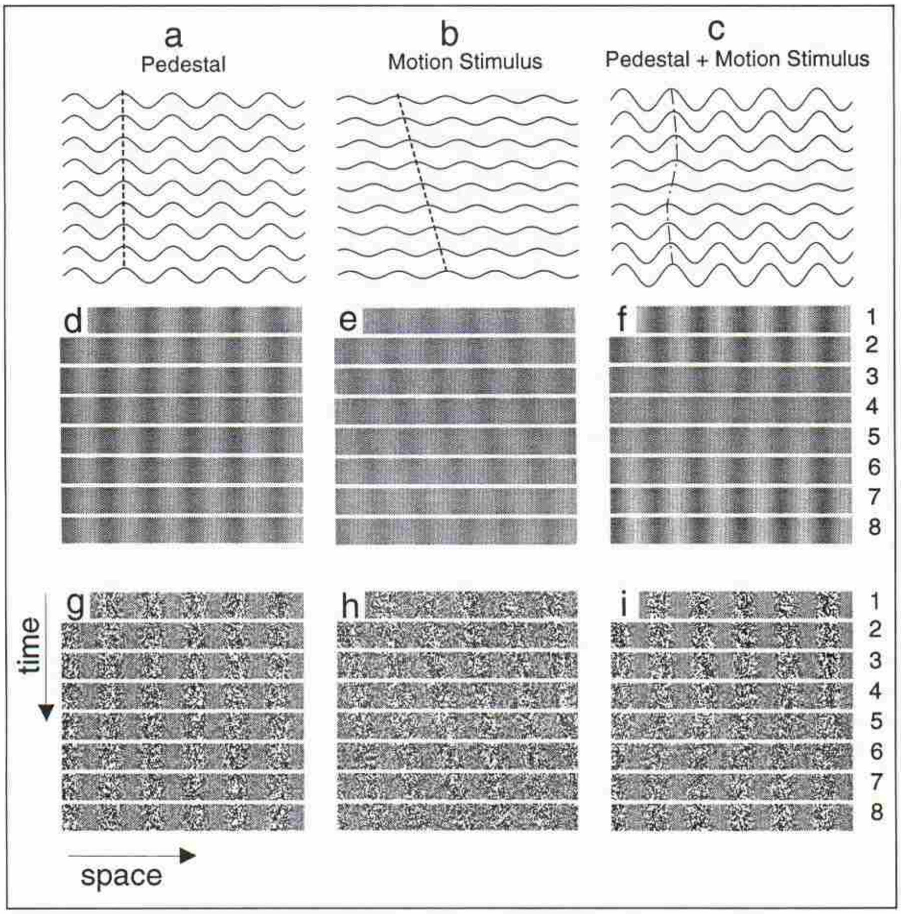

Fig. 2. The pedestal paradigm. (a) Schematic representation of eight frames of a stationary sine wave (the pedestal). The dashed vertical line indicates the (unchanging) location of the peaks. (d) Eight stimulus frames of the pedestal in a luminance-modulation stimulus. The actual frames were 3.1° × 1.6°; only a horizontal slice is shown. Luminance varies sinusoidally as a function of space. (g) Eight frames of the pedestal in a contrast-modulation stimulus. The average luminance is the same throughout the texture; contrast varies sinusoidally as a function of space. (b) Eight frames of a rightward-moving sine wave. The slanting line indicates the rightward movement of the peak. (e) Eight frames of a moving luminance-modulation sinusoid (first-order motion). From top to bottom, the sinusoid traverses one period. (h) Eight frames of a moving contrast-modulation sinusoid (second-order motion). (c) Pedestal plus motion stimulus, summation of the modulations of (a) and (b). The pedestal has twice the amplitude of the moving sine. The dashed line indicates the peak, which wobbles back and forth 1/6 of a period. (f) A pedestaled luminance-modulation motion stimulus, the sum of the modulations of (d) and (e). (i) A pedestaled contrast-modulation motion stimulus, the sum of the modulations of (g) and (h).

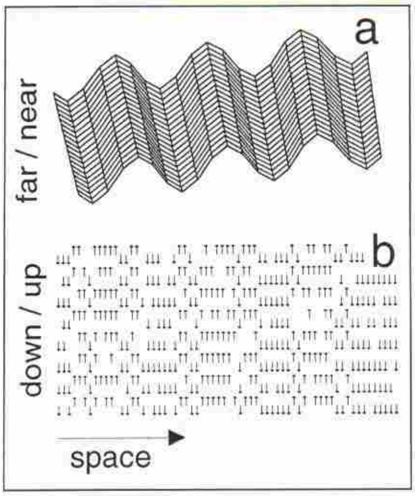

Fig. 3. Depth (a) and motion-defined-motion (b) gratings. In this representation of the appearance of a single frame of the depth grating, depth is sinusoidally modulated between "near" and "far." The actual depth grating was composed of randomly black or white pixels; the depth resulted from sinusoidally modulating stereoscopic disparity (see Fig. 4). In this representation of the motion-defined-motion stimulus, the arrows indicate the directions of dot motion between successive frames. The pattern of motion modulation (up vs. down) moves either to the left or to the right.

points would produce identical outputs.

Stereo-Depth Grating

The dynamic stereo-depth grating is created from stereo views of left- and right-half images composed of random dots. It appears in depth as a corrugated surface whose distance from the observer varies sinusoidally, as illustrated in Figure 3a. The grating (and its depth) exists only as a space-varying correlation between the dots in the left- and right-eye images. That is, the disparity between corresponding dots in the left and right monocular images defines the depth amplitude. Thereby, each new pair of left and right frames defines a corrugated grating (e.g., Fig. 4). In successive frame pairs, this grating moves consistently in one direction. Each monocular image alone is completely homogeneous without any hint of a grating, and successive monocular images are uncorrelated.

The pedestal added to the drifting pattern that defines the test is simply a second, static, corrugation. Because a depth grating has no consistent luminance or contrast modulation in space, it is invisible to both first-order and second-order motion systems.

Motion-Defined Motion

A motion-defined-motion grating (Fig. 3b) consists of random dots that move a small fixed distance between successive frames. The proportion of upward versus downward moving dots varies sinusoidally from left to right to define the modulation of the pedestal. A test stimulus is produced by drifting the up-down pattern horizontally in a consistent direction from frame to frame. In the pedestal-plus-test stimulus, the two modulations are added. Perceiving motion-defined motion requires computing the direction of motion of the dots and noting that the sine wave pattern of up-down dot motion drifts left or right with time. The ability to perceive this kind of motion-defined motion[10] seems to suggest a hierarchical organization of motion detectors. The movement of the motion modulation (i.e., the movement of the up-down motion pattern) is invisible to first- and second-order systems because there is no consistent modulation of luminance or contrast.

Results

Subjects perceive completely obvious apparent motion in all the motion-stimulus-alone conditions when the sine amplitude is sufficient. As Figure 5 shows, the temporal tuning functions (detectability as a function of number of bars that move past a point on the retina in 1 s) for all the motion types show typical lowpass filter characteristics (curves slope down to the right, indicating that high frequencies are attenuated and low frequencies "pass" without attenuation). The tuning functions can be divided into two groups: luminance grating and texture-contrast grating as one group (upper curves); depth grating and motion-defined motion as another group (lower set of curves). Within each group, the shapes of the temporal tuning functions are remarkably similar.

When subjects first view pedestaled luminance-modulation and texture-contrast-modulation stimuli, the wobble is dominant. However, with careful eye fixation and a little practice, they can learn to ignore the wobble and to perceive the linear motion. From this point on, remarkably, the presence of a pedestal with twice the amplitude of the moving stimulus has no ef-

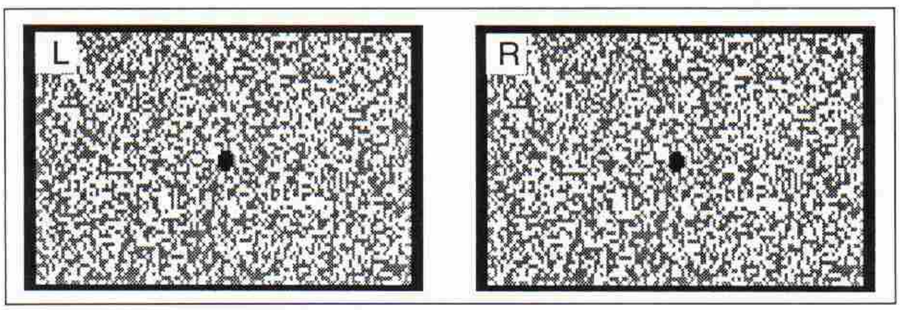

Fig. 4. Left (L) and right (R) stereograms illustrating the bottom third of one frame of a depth grating. Fixate the black dots to fuse the images and see the stereoscopic depth grating.

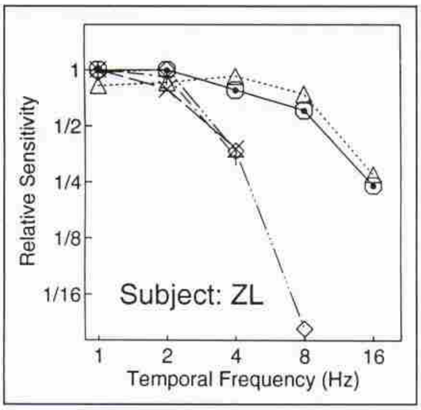

Fig. 5. Temporal tuning functions. The ordinate is the experimentally measured amplitude of the threshold modulation for correct discrimination of motion direction: sensitivity = $(\text{threshold})^{-1}$. The abscissa is the temporal frequency of a moving sinusoidal modulation. The axes are log scales. ○ indicates luminance-modulation motion for either pedestaled or nonpedestaled stimuli (thresholds are identical); △ indicates contrast-modulation motion for either pedestaled or nonpedestaled stimuli; + indicates simple (nonpedestaled) sinusoidal motion of depth stimuli; × indicates simple sinusoidal motion of motion-defined-motion stimuli; diamond indicates interocular sinusoidal motion of luminance motion stimuli. The curves have been vertically translated to expose their similarity in shape. Redrawn from Lu and Sperling, The functional architecture of human visual motion perception,[1] with permission from Elsevier Science Ltd.

fect on subjects' performances in the luminance- and contrast-modulation conditions. But pedestals reduce performance to chance-guessing levels with the depth and motion-defined-motion gratings. For pedestaled depth and motion-defined-motion stimuli, subjects report that they perceive only back-and-forth wobble motion, and cannot judge the direction of the (apparently invisible) linear motion component.

These results indicate clearly that there are two qualitatively different motion extraction systems. One class of systems utilizes a motion-energy algorithm. Figure 5 shows it has a much higher cutoff frequency (12 Hz) in its temporal sensitivity characteristics. (The cutoff is that frequency at which sensitivity has declined by a factor of 1/2.) Both luminance (first-order) and texture-contrast (second-order) stimuli use the motion-energy algorithm. Interestingly, the texture-contrast-motion system has the same temporal frequency characteristics as the luminance-motion system, despite frequent speculation that the second-order system is "slower" than the first-order system. The third-order system, which is indeed slower than the first- and second-order systems, can extract motion from depth and motion-defined-motion stimuli that are invisible to the first- and second-order systems.

FUNCTIONAL RELATIONSHIPS BETWEEN SYSTEMS

The pedestal experiments indicated the existence of three systems. Four subsequent procedures confirmed these systems and clarified the relationships between them.[1]

First- and Second-Order Motion Are Computed in Separate and Independent Channels

In one experiment, we superimposed (linearly added) a luminance and a texture-contrast stimulus, each with its own pedestal. The stimuli were of equal strength in terms of the number of just-noticeable differences above threshold. What happened? When the stimuli moved in opposite directions, there was no apparent motion: The two motion signals canceled exactly. When they moved in the same direction, however, there was enhanced apparent motion; its strength was given by probability summation of the response to the two component stimuli. (When two mechanisms attempt to detect the same motion stimulus, probability summation means that the response is correct if either mechanism succeeds.) There was no dependence of the motion strength on the relative phases of the two stimuli. If the two kinds of stimuli were combined prior to the motion computation, the sign (+ or −) of the combination would depend on the relative phase of the components. For example, stimuli of the same frequency and amplitude moving in the same direction but with a 180° phase difference (i.e., opposite sign) would perfectly cancel each other. The documented absence of any such phase dependence means that luminance- and contrast-motion strengths are first computed separately and independently; only then are the two motion strengths combined.

Motion-Energy Computations Are Primarily Monocular

Pedestaled luminance and texture-contrast stimuli were created with only four frames per cycle, successive frames being separated by 90°. In normal binocular or monocular viewing, motion in these stimuli was perceived as well as in stimuli with eight (or more) frames per cycle (e.g., Figs. 2e and 2h). However, when successive frames were directed alternately into left and right eyes,[11] subjects either could not perceive motion at all or had greatly reduced sensitivity.[1] Under such interocular presentation, the motion stimulus in each eye of an observer is ambiguous, and perception of coherent motion would be possible only if the motion-energy computations can combine information from both left and right eyes. These results suggest that with pedestaled luminance and texture-contrast stimuli, the direction of motion is computed primarily monocularly, but there may be a weak interocular component. We concluded that the motion-energy computations are primarily monocular.

Interocular Luminance Motion Is Computed in the Third-Order Motion System

Consider the display of a simple (not pedestaled) moving luminance sinusoid with successive frames separated by 90° (Fig. 2e). We found that converting such a stimulus from monocular to interocular presentation (Fig. 6) raises the contrast threshold (at low frequencies) by a factor of 12 (from 0.17% to 2.0%) and decreases the cutoff frequency from 12 to 3 Hz. The resulting tuning function is exactly like that of the depth and motion-defined-motion stimuli (Fig. 5). This result shows that the motion of an interocular luminance grating is perceived by the third-order system. The third-order system exhibits exactly the same cutoff frequency when it detects motion in interocular luminance stimuli as it does when it detects motion in stereo-depth and in motion-defined-motion stimuli.

The interocular procedure illustrated in Figure 6 with a luminance sinusoid can equally well be applied to the motion-defined-motion stimulus. Remarkably, the threshold for interocular presentation of motion-defined motion (without a pedestal) is almost the same as for monocular presentation. This result indicates that the motion-defined-motion computation is inherently binocular—it is indifferent to the eye of origin of any stimulus frame.

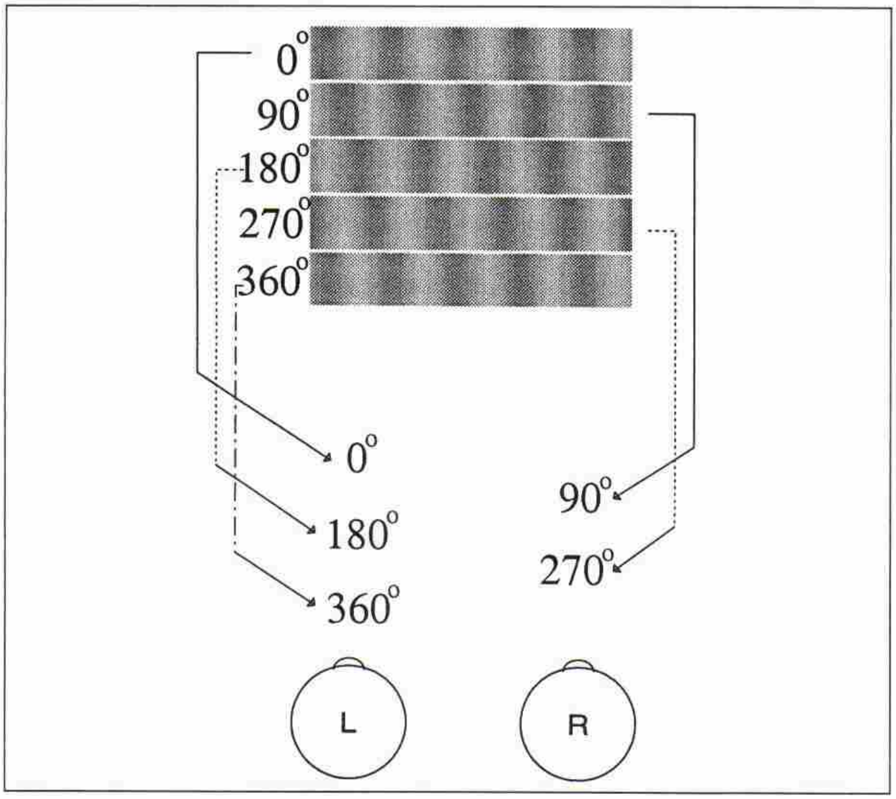

Fig. 6. Representation of an interocular stimulus presentation in which frames are alternately directed to the left (L) and right (R) eyes. Each successive stimulus has a spatial phase shift of 90°. Within an eye, the stimulus sequence, indicated on the bottom, is ambiguous as to direction of motion.

Classical Sinusoidal Stimuli

The conclusion from many experiments is that the motion of an apparently simple stimulus, such as a drifting, sinusoidal luminance grating, is computed by all three systems: First, the primarily monocular first-order system is extremely sensitive to luminance sine waves. Second, the primarily monocular second-order (texture-contrast) system detects the contrast extrema (the peaks and valleys) that occur at twice the frequency of the original sine wave. It is fast but not nearly as sensitive to these sine waves as the first-order system. Third, the binocular third-order system, which is both less sensitive and slower, detects sine waves when they reach sufficient amplitude. Thus, the drifting sine wave grating, which is regarded as a universal tool for visual psychophysics, turns out to be not a particularly useful analytic tool for discriminating between motion mechanisms.

Spatial Contrast Sensitivity Functions of the Systems

Motion-energy detection of luminance- and texture-contrast-modulation motion is primarily monocular. In interocular presentations, for which information from both eyes must be combined to solve the motion problem, the perceptual system relies on a third-order mechanism. We exploited these facts to compare the spatial contrast sensitivity functions of the first-order and third-order systems (the subject's detection threshold as a function of the number of bars per degree of visual angle, i.e., how densely the bars are packed) for comparable luminance sine wave stimuli. The first-order (luminance motion-energy) system was equally sensitive to spatial frequencies in the range from 0.6 to 4.8 cycles per degree of visual angle. The third-order system was 10 times less sensitive at 0.6 cycles per degree, and its sensitivity declined roughly in proportion to spatial frequency in this range. At 4.8 cycles per degree, the third-order system was 30 times less sensitive than first-order system. (The relative temporal frequency sensitivities of the systems were given in Fig. 5; the lower curves represent the third-order tuning function.) The third-order motion system achieves its ability to detect all the different kinds of motion stimuli at the cost of greatly reduced spatial and temporal sensitivity for those stimuli that the first- and second-order systems are especially adapted to detect.

ATTENTIONAL INFLUENCES ON MOTION PERCEPTION

Attention appears to play no role in the perception of first- and second-order motion.[12] We used a novel *alternating-feature* paradigm to demonstrate that in third-order motion, voluntary selective attention can determine not only the direction of perceived visual motion, but even whether motion is perceived at all.[1] Displays consisted of five frames. The odd frames (1, 3, 5) were composed of one class of features, and the even frames (2, 4) of a completely different class of features. Two examples are shown in Figure 7. In the first example (top row), the odd frames are composed of adjacent areas of white dots and black dots. All the areas of the odd frames have the same average luminance and the same magnitude of microcontrast. The difference in luminance between the tiny white dots in one bar and the dark gray background against which they are set is the same magnitude of microcontrast as the difference in luminance between the tiny black dots in the alternate area and the lighter gray background against which they are set. The direction of contrast between the dots and their background differs between areas, but the magnitude of microcontrast is the same. Because it has been shown that the second-order system detects only the magnitude of contrast and is not sensitive to the sign (+ or −) of contrast, these alternating areas do not provide a stimulus to that system. This indifference of the second-order system to the direction of contrast is the rectifying nonlinearity mentioned earlier as defining the difference between the first- and second-order systems. Thus, the alternating areas in Figures 7a, 7c, and 7e cannot stimulate either the first- or the second-order system.

The even frames in the first example in Figure 7 are composed of adjacent areas of high and of low contrast. These frames are invisible to the first-order system because they all have the same average overall luminance. They can be seen by the second-order system because the contrast differs between areas. However, they alternate from one small fraction of a second to the next with frames that do not stimulate the second-order system.

In the second example in Figure 7, odd frames are composed of adjacent areas of fine and of coarse textures that also differ in orientation. Even frames consist of a stereo-depth grating—adjacent areas of near and far (Fig. 4).

The alternating-feature paradigm is analogous to the interocular paradigm in that successive frames are displaced 90°, and within even frames or within odd frames, there is no motion signal. Unlike interocular displays, however, the alternating-feature displays have no motion between even and odd frames because both the first-order (luminance) and the second-order (contrast-modulation) systems are blind to the depth displays and to the black-and-white spot displays.

The only way for subjects to perceive consistent motion in alternating-feature displays is to connect the most salient stimulus features across frames, independently of what makes those features salient, a fundamentally different motion algorithm than either first-order or second-order motion extraction. The principles are as follows: The most significant features are marked in a feature-salience map (Fig. 7, row 3). Insofar as images are perceptually segmented into areas of figure and areas of ground, marked is equivalent to figure, and unmarked to ground. In contrast frames (Figs. 7b and 7d), the areas of high contrast are automatically marked, and areas of low contrast are not. Similarly, near areas in the depth grating are automatically marked, and far areas are not. The odd frames are attentionally neutral. Without attentional instructions, no consistent motion is seen. However, when the subject is instructed to attend to one or the other texture, these areas become marked in the salience map. Motion between marked areas is computed; direction depends on which type of texture element (fine or coarse stripes, black or white dots) is attended and marked (Fig. 7).

In formal experiments, subjects maintain rigid eye fixation while sequences of five successive frames (Fig. 7) are presented at a frequency of 7.5 frames/s. An entire display is completed in 667 ms. From trial to trial, the even and odd features, the direction of motion, and other stimulus factors are varied randomly. When subjects view these displays prior to any attentional instructions, reports of motion direction are random. Once subjects have practiced attending to a particular feature, they perceive motion in the direction corresponding to the attended feature in 75% to 95% of trials. To switch attention to a previously unattended feature takes an hour or so of practice, but then direction reversal occurs. The same effect of attention occurs even when the stimuli are speeded up so that the duration of the entire display is only 300 or 333 ms. With extremely brief displays, the attentional effect is reduced to about 70%, but still far above chance.

In these displays, the same stimulus is perceived as nonmoving or moving ambiguously prior to attentional instructions, as moving in one direction when observers attend to one type of texture element, and as moving in the opposite direction when they attend to the other type of texture element. These findings indicate that not only inherent stimulus properties (such as high contrast or nearness) but also attention determines what features are salient. Observers are able to make consistent movement-direction judgments even in displays of five frames that occur within 300 ms (i.e., that are over in a flash). The ability to make such quick judgments indicates that direction computation is not based on a conscious tracking of salient features[13] but on an automatic motion computation.[1]

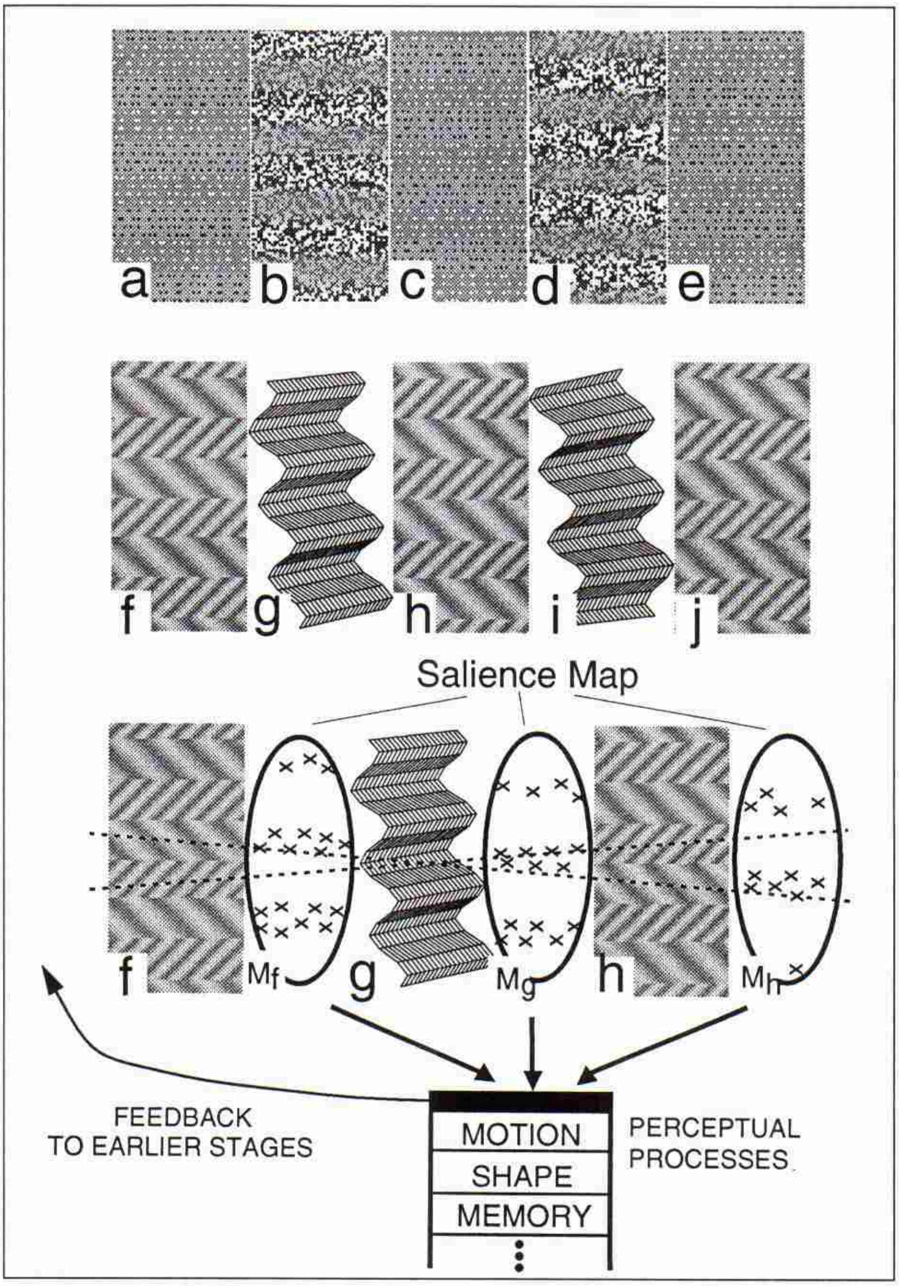

Fig. 7. Alternating-feature stimulus sequences for attention-generated motion, and their relation to feature-salience maps. The top row (a–e) shows an alternating-feature display with frames of black and white dots and frames of low- and high-contrast textures. A sequence of five consecutive frames is shown; each is displaced vertically by 90° from the previous one. The second row (f–j) shows a depth/texture alternating-feature display. The depth frames (g, i) are indicated schematically. The third row shows frames f, g, and h with their associated salience maps (M); the most salient features are marked with *X*s. In depth displays, the near peaks are automatically the most salient. No features in the texture displays are automatically salient. When the subject intentionally attends to the coarse grating, its features are marked in the salience map, and the direction of apparent motion is from upper left to lower right, as indicated by the dashed line. There is no support for upward motion (the other dashed line, from lower left to upper right); perceiving upward motion in this display would require attention to the fine stripes. The fourth row illustrates that third-order motion is computed directly from the salience map, which also provides guidance to other perceptual processes such as visual search and transfer to memory. Reprinted from Lu and Sperling, Attention-generated apparent motion,[1] with permission from *Nature*.

Different Tasks, Same Mechanism

Consider a perceptual search task in which a subject searches an array for a target digit embedded among letters. Suppose half the characters are green and half are red. Informing the subject that the target is red increases search efficiency. Now consider a subject searching a rapid stream of many arrays, in which all-red arrays alternate with all-green arrays, each one falling on top of the previous one. In this case, when color indicates a temporal, not a spatial, location, there is no search advantage in knowing the target's color.[14] This and related experiments show that selective search of red items is accomplished not by early perceptual exclusion of green items, but by an attentional mechanism that directs pattern-matching processes to the location of red items. The critical aspect of these results is that, even in search for particular features, selective attention is mediated via locations of the attended features.

When subjects view a brief flash of an array that contains more characters than they can recall, a cue can be used to direct them to recall only a designated subset (a partial-report cue). Partial-report cues that designate to-be-reported characters by type (e.g., numbers vs. letters) or by feature (red vs. green) are much less effective than cues that designate locations (such as a particular row of the array). These results indicate that access to visual short-term memory is mediated via spatial location, not by properties of the material to be stored.[15]

All these phenomena are encompassed within a single theory of spatial attention. Initially, prior to a trial, the subject receives instructions to attend to specific features. The subject uses the instructions to set parameters that determine how information will be processed during the trial. This is a top-down process (high-level cognitive processes control low-level sensory processes). During the trial itself, the subject processes the to-be-attended features, a bottom-up process (sensory processes send information upward to perceptual and cognitive processes). At an intermediate level, the top-down and bottom-up interaction produces a salience map, a dynamic map of the locations of the most salient stimulus features. The salience map can be used directly to compute feature-salience motion, or the salience map can be used to guide bottom-up process-

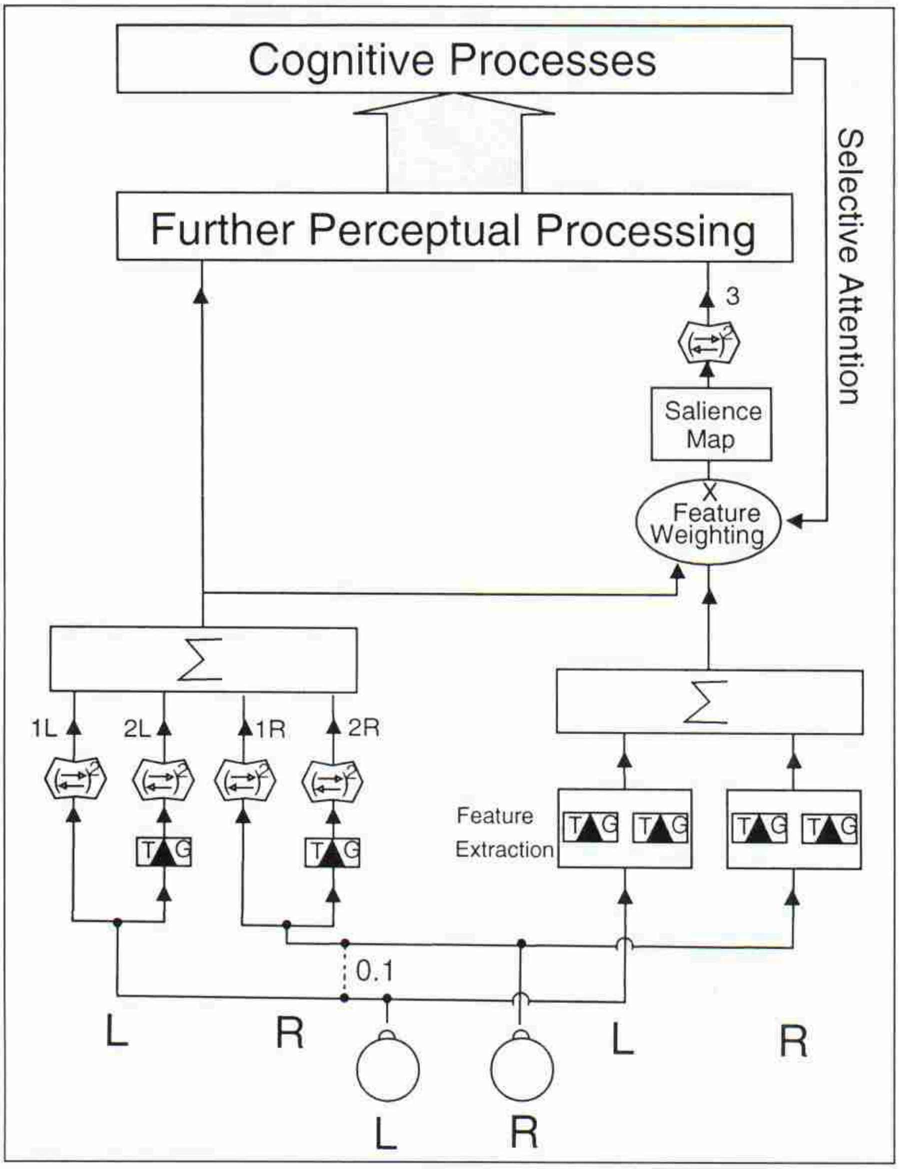

Fig. 8. Functional architecture of the visual motion system. The fast, primarily monocular mechanisms are represented on the left; third-order motion is represented on the right. L and R indicate left- and right-eye signals, respectively. The dotted line adjacent to 0.1 indicates that there is 10% crosstalk between the primarily monocular channels (i.e., 10% of the left channel's signal crosses over into the right channel, and vice versa). A motion-energy detector $(\rightleftarrows)^2$ acting on the input (or spatially filtered input) serves luminance-modulation (first-order) motion (1L and 1R). Second-order motion (2L, 2R) requires a texture grabber, TG (a spatial filter followed by fullwave rectification, i.e., absolute value of each point's difference from mean luminance) prior to a motion computation. Σ indicates (possibly complex) summation; $\times$ represents multiplication—the differential weighting of features determined by selective attention; the connection from motion-energy Σ to feature weighting conveys the motion features needed to solve motion-defined-motion stimuli by the third-order motion detector (3). Revised from Lu and Sperling, The functional architecture of human visual motion perception,[1] with permission from Elsevier Science Ltd.

ing (as in attention-guided search), or to control access to visual memory.

THE BIG PICTURE

The organization of the three motion systems is summarized in Figure 8. This schematic flowchart describes only motion-direction discrimination. Introspective observations indicate that when the outputs of the fast motion-energy detectors are viewed without the contribution of the feature-salience system (i.e., in stimuli with pedestals or at high temporal frequencies), the motion is perceived as "objectless" motion. The contribution of the feature-salience system seems to be necessary to bind the motion computed at a location with an object that is perceived to be in motion. The motion-energy systems provide raw motion data that require further perceptual processing (Fig. 8) for the extraction of velocity, three-dimensional structure from motion,[16] and other useful properties.

Much as the spectroscopic methods of atomic physics enabled physicists to unravel the structure of atoms, psychophysical methods enable vision scientists to map the mental processes involved in the computation of motion direction. At present, the perception of motion direction seems to involve three systems and five separate computations (primarily left- and primarily right-eye first-order motion, primarily left- and primarily right-eye second-order motion, binocular third-order motion) with the interrelations indicated in Figure 8. As in the case of atoms, we expect that, in the future, these processes will be further subdivided and additional ones will be discovered.

Acknowledgments—This research was supported by Air Force Office of Scientific Research, Life Sciences Directorate, Visual Information Processing Program.

Notes

1. For a more detailed account of the work described herein, and for relevant references, the reader should consult two recent summary articles: Z.-L. Lu and G. Sperling, The functional architecture of human visual motion perception, *Vision Research, 35,* 2697–2722 (1995); Z.-L. Lu and G. Sperling, Attention-generated apparent motion, *Nature, 377,* 237–239 (1995).

2. W. Reichardt, Autokorrelationsauswertung als funktionsprinzip des zentralnervensystems, *Zeitschrift Naturforschung, 12b,* 447–457 (1957); W. Reichardt, Autocorrelation, a principle for the evaluation of sensory information by the central nervous system, in *Sensory Communication,* W.A. Rosenblith, Ed. (MIT Press and John Wiley, New York, 1961).

3. J.P.H. van Santen and G. Sperling, Temporal covariance model of human motion perception, *Journal of the Optical Society of America A, 1,* 451–473 (1984).

4. A.B. Watson and A.J. Ahumada, A look at motion in the frequency domain, in *Motion: Perception and Representation,* J.K. Tsotsos, Ed. (Association for Computing Machinery, New York, 1983); E.H. Adelson and J.R. Bergen, Spatiotemporal energy models for the perception of apparent motion, *Journal of the Optical Society of America A, 2,* 284–299 (1985); J.P.H. van Santen and G. Sperling, Elaborated Reichardt detectors, *Journal of the Optical Society of America A, 2,* 300–321 (1985).

5. C. Chubb and G. Sperling, Second-order motion perception: Space-time separable mechanisms, in *Proceedings: Workshop on Visual Motion* (IEEE Computer Society Press, Irvine, CA, 1989).

6. C. Chubb and G. Sperling, Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception, *Journal of the Optical Society of America A, 5,* 1986–2006 (1988); V.S. Ramachandran, M.V. Rau, and T.R. Vidyasagar, Apparent movement with subjective contours, *Vision Research, 13,* 1399–1401 (1973); G. Sperling, Movement perception in computer-driven visual displays, *Behavior Research Methods and Instrumentation, 8,* 144–151 (1976); A.M.M. Lelken and J.J. Koenderink, Illusory motion in visual displays, *Vision Research, 24,* 293–300 (1984); A.M. Derrington and D.R. Badcock, Separate detectors for simple and complex grating patterns? *Vision Research, 25,* 1869–1878 (1985); K. Turano and A. Pantle, On the mechanism that encodes the movement of contrast variations—I: Velocity discrimination, *Vision Research, 29,* 207–221 (1989); J.D. Victor and M.M. Conte, Motion mechanisms have only limited access to form information, *Vision Research, 30,* 289–301 (1989); P. Cavanagh and G. Mather, Motion: The long and the short of it, *Spatial Vision, 4,* 103–129 (1989); C. Chubb and G. Sperling, Two motion perception mechanisms revealed by distance driven reversal of apparent motion, *Proceedings of the National Academy of Sciences, USA, 86,* 2985–2989 (1989); C. Chubb and G. Sperling, Texture quilts: Basic tools for studying motion-from-texture, *Journal of Mathematical Psychology, 35,* 411–442 (1991).

7. G. Sperling, Three stages and two systems of visual processing, *Spatial Vision, 4,* 183–207 (1989).

8. This property is pseudolinearity because it holds only for sines and only when they have different temporal frequencies. The original proof of these two properties (note 3) applies to displays of infinite duration and in continuous motion; it was later extended to sampled displays of finite duration in the two articles by Lu and Sperling (note 1).

9. K. Nakayama and G.H. Silverman, Detection and discrimination of sinusoidal grating displacements, *Journal of the Optical Society of America A, 2,* 267–274 (1985); Z.-L. Lu and G. Sperling, *Contrast saturation in first- and second-order motion perception,* Technical Report MBS 95-25 (Institute for Mathematical Behavioral Sciences, University of California, Irvine, 1995).

10. J.T. Petersik, K.I. Hicks, and A.J. Pantle, Apparent movement of successively generated subjective figures, *Perception, 7,* 371–383 (1978); J.M. Zanker, Theta motion: A paradoxical stimulus to explore higher order motion extraction, *Vision Research, 33,* 553–569 (1993).

11. M. Shadlen and T. Carney, Mechanism of human motion revealed by new cyclopean illusion, *Science, 232,* 95–97 (1986).

12. J.A. Solomon and G. Sperling, Full-wave and half-wave rectification in 2nd-order motion perception, *Vision Research, 34,* 2239–2257 (1994).

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

14. K.R. Cave and J.M. Wolfe, Modeling the role of parallel processing in visual search, *Cognitive Psychology, 22,* 225–271 (1990); S. Shih and G. Sperling, Is there feature-based stimulus selection in visual search? *Journal of Experimental Psychology: Human Perception and Performance* (in press).

15. G. Sperling, The information available in brief visual presentations, *Psychology Monograph,* 74(11, Whole No. 498) (1960); J.M. von Wright, On selection in visual immediate memory, *Acta Psychologica, 33,* 280–292 (1970).

16. D. Heeger, Model for the extraction of image flow, *Journal of the Optical Society of America A, 4,* 1455–1471 (1987); G. Sperling, M.S. Landy, B.A. Dosher, and M.E. Perkins, The kinetic depth effect and the identification of shape, *Journal of Experimental Psychology: Human Perception and Performance, 15,* 426–440 (1989).