5 A Systems Analysis of Visual Motion Perception

George Sperling and Zhong-Lin Lu

Sperling, G., & Lu, Z.L. (1998). A systems analysis of visual motion processing. In T. Watanabe (Ed.), High-level motion processing (pp. 153-183). Cambridge, MA: MIT Press.

Using new psychophysical methods, it recently has become possible to isolate and measure three systems of human motion perception. The first-order system responds to moving luminance patterns. The second-order system responds to moving modulations of feature types. The first- and second-order systems are primarily monocular, sensitive, and fast. A third-order system computes motion from a 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 chapter describes how these systems were isolated and how the relations between them were defined.[1] It also describes how these early motion computations fit into a larger framework of processing stages that precede and follow motion-direction processing. In order, these stages are: light adaptation, contrast-gain control, motion-direction computation, a salience field (for figure-ground resolution and attention gating), further perceptual and decision processes, and ultimately top-down cognitive control of attention.

For more than one hundred 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 that of 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 (Braddick, 1974; Pantle and Picciano, 1976; Fennema and Thompson, 1979; Marr and Ullman, 1981; van Santen and Sperling, 1984; Fleet and Jepson, 1985; Adelson and Bergen, 1985; Mather, Cavanagh, and Anstis, 1985; Chubb and Sperling, 1989a; Cavanagh and Mather, 1989b; Wilson, Ferrera, and Yo, 1992; Boulton and Baker, 1993). Although there clearly is a kernel of truth underlying most of these dichotomies and theories, there were two persistent problems. It had not been possible to obtain a demonstrably pure measure of any proposed mechanism, nor had 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. Primarily, in this review, we wish to demonstrate how combining a new paradigm (pedestal displays) with several critical subsidiary paradigms (interocular displays, stimulus superpositions with varying phases and directions, alternating-feature stimuli, masking, selective adaptation, and attentional manipulations) has led to a possible resolution of these issues (Lu and Sperling, 1995a, 1995b).

GENERAL ORGANIZATION OF THE VISUAL SYSTEM

Retinal Adaptation

Human vision is effective at ambient light levels that vary from dim starlight (about 10^{-3} cd/m^2) to intense sunlight (about 10^5 cd/m^2). On the other hand, most visual phenomena are relatively independent of ambient light levels (Hood and Finkelstein, 1986). Obviously, one of the first tasks of those modules of the visual system that are interested in computing properties of objects, independent of their illumination, is to derive for themselves input signals that are independent of illumination (Blackwell, 1946; Heinemann, 1955; Sperling and Sondhi, 1968; Whittle and Challands, 1969; Sperling, 1970; Heinemann, 1972; Shapley and Enroth-Cugell, 1984) (figure 5.1). This process is called *light adaptation*. Light adaptation is quite complex, occurring in many stages within the visual receptors, the rods and cones. Although experiments in motion perception seldom explore a wide range of luminance levels, it is nevertheless essential to have at least an approximate representation of light adaptation in a systems analysis of vision.

From a computational or systems point of view, there are three important aspects of light adaptation. First, the essence of light adaptation is dividing the input luminance at a point by the average luminance in a larger spatial and temporal neighborhood, and thereby extracting the point contrast. This aspect of light adaptation is represented as gain control (K_1) in figure 5.1, which gives an overview of early visual processing. Luminance is an image property. By itself, the luminance at a point is uninformative about an object's surface. On the other hand, point contrast represents the extent to which a point on the surface of an object reflects more or less light than its surround. Point contrast is more informative than luminance. But because of shadows, lighting variations, the unknown nature of the environment, and other factors, point contrast is not completely reliable either.

The second critical aspect of light adaptation is a change in the nature of visual receptive fields, which are represented in figure 5.1 as linear, separable, spatial and temporal filters. In the light, both the spatial and temporal filters have approximately balanced positive and negative regions; i.e., they are band-pass filters. In the dark they have only a single positive region; i.e., they are low-pass filters (figure 5.1). In the real visual system, the change from dark- to light-adapted receptive fields occurs gradually over a very wide range of luminances, and the processes of adaptation and of creating a center-surround receptive field are more separated than in this model. These manifold complexities of adaptation are only coarsely represented by a single stage of gain control, although this minimal model suffices for the present purpose.

The third aspect of light adaptation is its effect on signal-to-noise ratios. Stages subsequent to adaptation see only the sum of signal and sensory noise (figure 5.1). Because the signal is divided by the average input, maintaining a constant signal-to-noise ratio requires that signals be proportional to average luminance. This is Weber's Law, and it arises as a simple consequence of gain control that is determined by the mean inputs (Sperling, 1989).

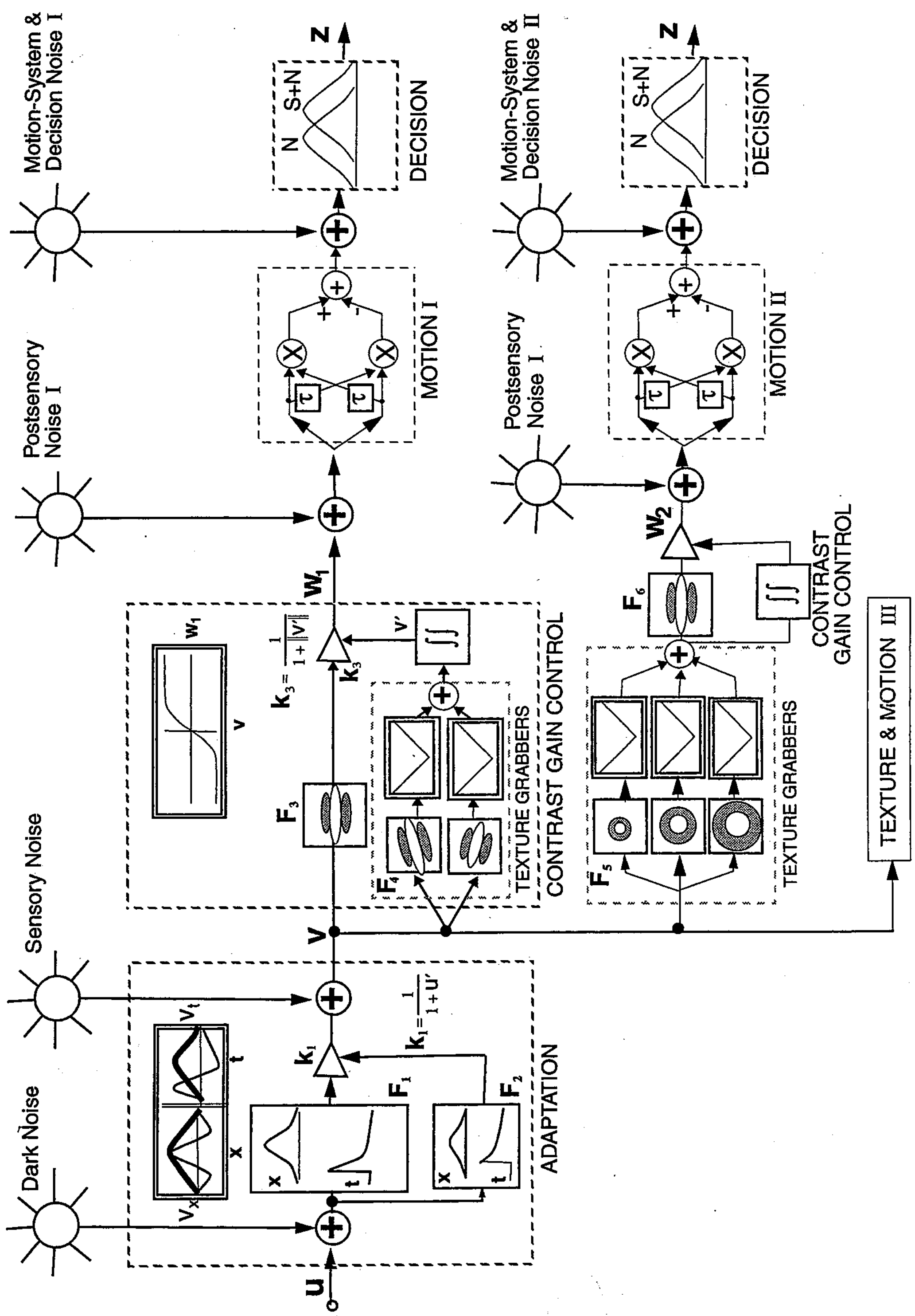
Light adaptation is both a physiological and a psychophysical phenomenon. Receptive fields are properties of neurons that are represented in psychophysics as channels (e.g., Blakemore and Campbell, 1969; Sachs, Nachmias, and Robson, 1971; Wilson and Bergen, 1979; Watson and Robson, 1981; Watson, Barlow, and Robson, 1983; Watt and Morgan, 1985). Channels (or receptive fields) and light adaptation are modeled computationally by linear filters and by gain-control mechanisms.

Contrast-Gain Control

Light adaptation is followed by contrast-gain control (e.g., Sperling, 1989). Contrast-gain control is a phenomenon that is observed in neurons of the visual system, specifically in retinal ganglion cells (Shapley and Enroth-Cugell, 1984), in cells of the lateral geniculate nucleus (Derrington and Lennie, 1981; Kaplan and Shapley, 1982), and in primary visual cortex (Dean, 1981; Albrecht and Hamilton, 1982; Ohzawa, Sclar, and Freeman, 1982; Sclar, Maunsell, and Lennie, 1990; Bonds, 1991; Albrecht and Geisler, 1991; Heeger, 1992). Neurons in these locations do not modulate their responses with the input contrast beyond a certain level, so that contrast-gain control must have occurred earlier.

The importance of contrast-gain control is that, once a stimulus achieves a critical level of contrast, further increases in contrast do not affect the representation. Contrast-gain–controlled inputs allow subsequent processes to compute without the distraction of irrelevant contrast variations. Consider, for example, judgments of distance or velocity. Obviously, such a judgment should be independent of object contrast insofar as possible, i.e., once stimulus contrast is sufficient to make the object clearly visible.

For contrast greater than about 5 to 10 percent, many visual psychophysical tasks are contrast-independent (Pelli, 1981; Nakayama and Silverman, 1985; Jamar and Koenderink, 1985; McKee, Silverman, and Nakayama, 1986; Pavel et al., 1987; Parish and Sperling, 1991). On the other hand, much of what is known about the mechanisms of human motion perception has been



Dark Noise
Sensory Noise
Postsensory Noise I
Motion-System & Decision Noise I
U
ADAPTATION
F_1
F_2
k_1
$k_1 = \frac{1}{1+U'}$
V
CONTRAST GAIN CONTROL
TEXTURE GRABBERS
F_3
F_4
k_3
$k_3 = \frac{1}{1+\|V'\|}$
V'
W_1
MOTION I
DECISION
N
S+N
Z
TEXTURE GRABBERS
F_5
F_6
CONTRAST GAIN CONTROL
W_2
Postsensory Noise I
Motion-System & Decision Noise II
MOTION II
DECISION
N
S+N
Z
TEXTURE & MOTION III

Figure 5.1 A model of early visual processing. The suns indicate stationary noise sources; rectangular boxes indicate linear filters; double rectangles enclose graphs of input-output relations. Triangles indicate gain-controlled amplifiers, the input (controlled signal) is the horizontal path, the gain-control path is vertical. Adaptation: The visual input **u** passes through a separable space-time filter F_1 that, in combination with the gain control filter F_2, creates a receptive field of a given spatial scale (i.e., a visual "channel"). The gain of the amplifier $\mathbf{k}_1$ is determined by the spatiotemporal surround of the through-signal u'. Amplifier gain is feedforward-controlled via filters F_2, which have greater spatial and temporal extents than those of controlled signal F_1. For very small inputs, the gain control signal u' is negligible, and the receptive field is entirely positive (low-pass) as indicated by the dark graphic plots in the double box within Adaptation. At high luminances, the receptive fields are band-pass, as indicated by the thin graphic plots. The orientation selectivity of first-order motion is represented by a spatial filter, F_3. Contrast-gain control of first-order motion is implemented as feedforward control by filters F_4, which have a somewhat greater spatiotemporal extent $\iint$ and cover a broader range of frequencies than the through-signal filter F_3. The double boxes with "V" inside indicate full-wave rectifiers whose outputs are approximately the absolute value of their inputs. Each combination of a filter and a rectifier is a texture grabber. The summed outputs of the texture grabbers determine gain control. In the second-order motion pathway, both the input and the gain control are created by summing the outputs of texture grabbers F_5. The motion-controlled signal has some additional spatial-frequency selectivity F_6, and the controlling signal is summed over the surrounding area $\iint$. The motion computation itself is identical in first-order (motion I) and second-order (motion II) systems, which are represented here as Reichardt detectors. The decision process is represented as a maximum likelihood decision between two alternatives (N, S + N). The third-order motion system is indicated only schematically by the rectangle (motion III). (Adapted with permission from Sperling, 1989, and Lu and Sperling, 1996b.)

derived from psychophysical experiments at extremely low-contrast levels (Robson, 1966; Kelly, 1982; Kelly, 1979; Koenderink and van Doorn, 1979; Burr and Ross, 1982; van Santen and Sperling, 1984; Lu and Sperling, 1995b). At very low contrasts (e.g., less than 2 percent), it can be shown that motion stimuli sneak through the stages of light adaptation and contrast-gain control without being distorted. Very-low-contrast signals are not transformed or distorted by contrast-gain control. Therefore, experiments with such stimuli are ideal for ferreting out the properties of the motion and subsequent decision mechanisms.

Daily life presents perceptual tasks that involve a full range of stimulus contrasts. In this chapter, we first describe motion experiments performed at very low-contrast levels which reveal the properties of motion systems. To deal with higher-contrast levels, see the section, Drastically Different Contrast-Gain Control, below, which, together with figure 5.1, specifies the gain-control mechanisms of first- and second-order motion in a general way. This is a first step in calculating the distortions imposed by high-contrast stimuli prior to a motion computation.

CONCEPTUAL FRAMEWORK

Functional Equivalence of Motion-Energy Detectors and Reichardt Models

Computational theories of motion perception date from Reichardt's model for insect vision (Reichardt, 1957, 1961) which was adapted for human perception by van Santen and Sperling (1984). As illustrated in figure 5.2, 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

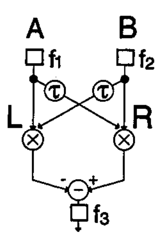


Figure 5.2 Reichardt motion detector (simplified). A and B indicate adjacent locations of visual receptive fields; τ is a temporal delay; $\times$ indicates multiplication; and $-$ indicates subtraction. The f_1 and f_2 are arbitrary spatiotemporal linear filters (receptive fields); f_3 is an arbitrary linear temporal-integrating filter. Leftward motion (B to A) is computed at L; rightward motion at R. Final outputs greater than zero indicate stimulus motion from A to B; outputs less than zero indicate stimulus motion from B to A.

multiplier R of figure 5.2, 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 subunit 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 (van Santen and Sperling, 1985; Adelson and Bergen, 1985), 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* (Chubb and Sperling, 1989b). We prefer the term *motion-energy detection* because it is more descriptive. The pedestal paradigm is a sensitive indicator for motion-energy detection. Concurrently, the pedestal paradigm provides a way of isolating and measuring first- and second-order motion systems, which are immune to pedestals, from the third-order motion system, which is not.

First- and Second-Order Motion

A rigidly moving object is a drifting modulation of luminance. 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, for example, 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 (Ramachandran, Rao, and Vidyasagar, 1973; Sperling, 1976; Lelkens and Koenderink, 1984; Derrington and Badcock, 1985; Turano and Pantle, 1989; Cavanagh and Mather, 1989; Chubb and Sperling, 1988; Chubb and Sperling, 1989a; Victor and Conte, 1990; Chubb and Sperling, 1991; Smith, 1994) have demonstrated clear motion perception in stimuli whose motion would be ambiguous for motion-energy detectors. For example, motion of the classes of drift-balanced and microbalanced stimuli cannot be extracted by motion-energy detectors. (A drift-balanced 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 drift-balanced even when it is viewed through an aperture of any arbitrary shape [Chubb and Sperling, 1988; Chubb and Sperling, 1991].) 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 (Chubb and Sperling, 1989a; Sperling, 1989).

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. This property is called *pseudolinearity* because it holds only for sines and only when they have different temporal frequencies. 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 called the *pedestal immunity* of motion-energy detectors.[2]

PEDESTAL IMMUNITY OF HUMAN OBSERVERS: EXPERIMENTS

The Motion Pedestal Test

We exploit the pseudolinearity of motion-energy detectors by creating compound stimuli (figure 5.3c) that have two components: a stationary sine grating (the pedestal, figure 5.3a) and a linear moving sine grating (the motion stimulus, figure 5.3b). The pedestal grating consists of stationary alternating light and dark bars for first-order stimuli (figure 5.3d) and of alternating high- and low-contrast texture bars for second-order stimuli (figure 5.3g). The linearly moving sine grating consists of moving light and dark bars for first-order stimuli (figure 5.3e) and moving high- and low-contrast texture bars for second-order stimuli (figure 5.3h). The peaks and valleys of the compound stimuli (figure 5.3f and i) wobble back and forth, moving first one way, then the other. Nevertheless, the output of a motion-energy detector is similar for the compound pedestal-plus-motion stimulus and 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 less than about 5 percent [Nakayama and Silverman, 1985; Lu and Sperling, 1996b], the light bars must be no more than 2.5 percent lighter than the mean luminance, and the dark bars no more than 2.5 percent 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)?

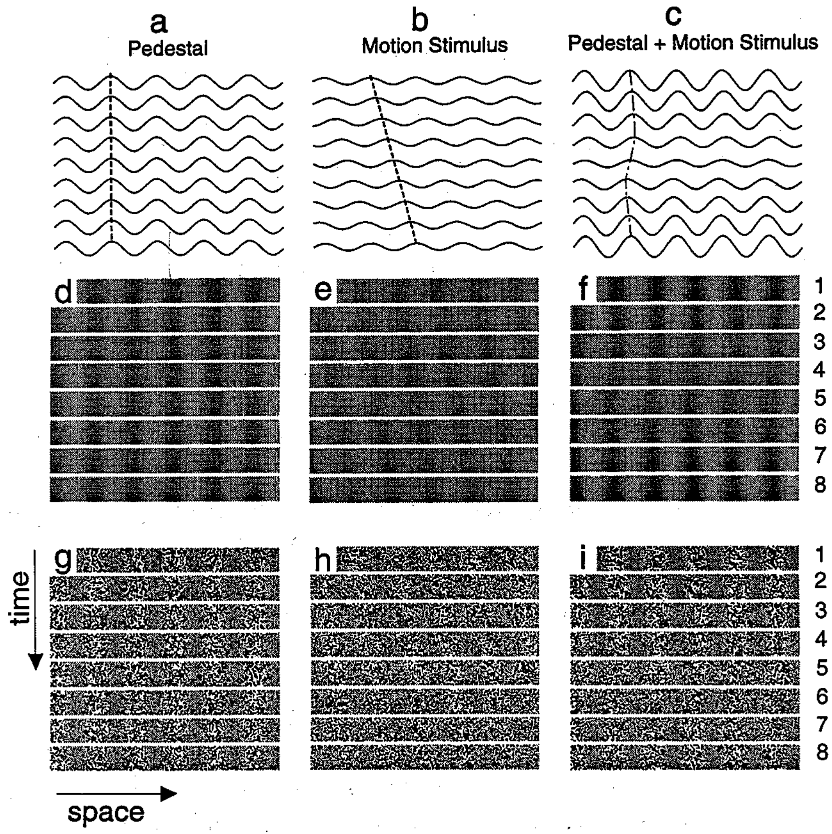


Figure 5.3 The pedestal paradigm. (a) Schematic representation of eight frames of a stationary sine wave (the pedestal). The dashed vertical line indicates the (unchanging) location of the peaks. (d) Eight stimulus frames of the pedestal in a luminance-modulation stimulus. The actual frames were 3.1° times 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, x. (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 one-sixth 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). (Reprinted with permission from Lu and Sperling, 1996a.)

To answer this question, we used the following procedure. Subjects viewed a computer-generated display, such as that illustrated in figure 5.2e and h, 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 percent correct responses. A pedestal with twice this measured threshold amplitude was then added to the moving stimulus to produce a pedestaled stimulus (figure 5.2f and i). If the motion judgment were based on the output of a motion-energy detector, we would expect the subject's accuracy of left vs. right judgments to be similar with and without the pedestal. On the other hand, 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 very difficult 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 (figure 5.3e). Luminance stimuli (figure 5.3d, e, f) are used to determine the properties of the first-order motion system.

Texture-Contrast Grating The moving texture-contrast grating (figure 5.3h) 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, while the other bar is composed of dark-gray and light-gray squares. One bar has high contrast between the dots of which it is composed, while the other bar has low contrast between its tiny dots. What differentiates the bars is the magnitude of contrast within each bar. What moves is the contrast modulation; the pixel pattern itself remains stationary. The motion of a texture-contrast modulator cannot be determined by motion-energy detectors that simply receive luminance inputs.

Exposing the contrast grating's motion to motion-energy detection has been shown to involve full-wave rectification (figure 5.4) (Chubb and Sperling, 1989a; Solomon and Sperling, 1994). Full-wave rectification means computing the absolute value (or any strictly monotonic function of the absolute value) of each point's deviation from mean luminance. Thereby, extreme black and extreme white points produce identical outputs. While full-wave rectification suffices to expose the latent motion in the stimuli of figure 5.3h, with other stimuli (Chubb and Sperling, 1988, 1989b, 1991), it can be shown that mere rectification is insufficient.

The essential component of second-order motion preprocessing is a *texture grabber* (figure 5.4). A texture grabber is a spatiotemporal filter that selects a particular kind or coarseness of texture (such as dots or oriented lines or boundaries) and reports the amount of such a texture at each location. Rectification is the process of converting the positive or negative output of a linear filter into a positive number, "the amount of." First- and second-order

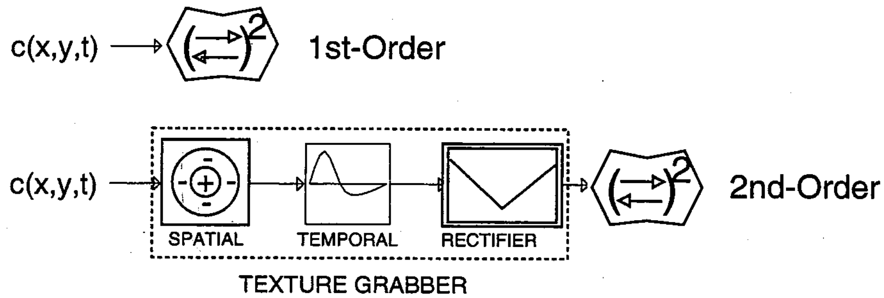


Figure 5.4 Processing of first- and second-order motion compared schematically. (a) In first-order motion, the motion detector (Reichardt or motion-energy) acts directly on point contrast, c(x, y, t), before any gross nonlinearities distort the input signal. (b) Second-order motion is computed on the summed output of texture grabbers (only one is illustrated). A texture grabber consists of three components: A spatial filter defines the feature to which the texture grabber is most sensitive. A biphasic temporal filter ensures that the texture grabber is most sensitive to changing vs. static stimuli. A rectifier converts its input (the positive and negative outputs of the filters) into an all-positive output that represents the amount of the feature present at the location of the texture grabber. (Based on Chubb and Sperling, 1988, and Werkhoven, Sperling, and Chubb, 1994.)

processing differ only in that the photons of first-order processing are replaced with features (little texture elements) of second-order processes. A first-order grating consists of alternating columns of light and dark (columns with larger and smaller amounts of photons than the mean). A second-order grating consists of, for example, alternating columns of higher-contrast texture and lower-contrast texture (columns with larger and smaller amounts of texture than the mean) (Chubb and Sperling, 1988, 1989b).

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 5.5a. The grating (and its depth) exist 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., figure 5.6, L and R). 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 static second 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 (figure 5.5b) consists of random dots that move a small fixed distance between successive

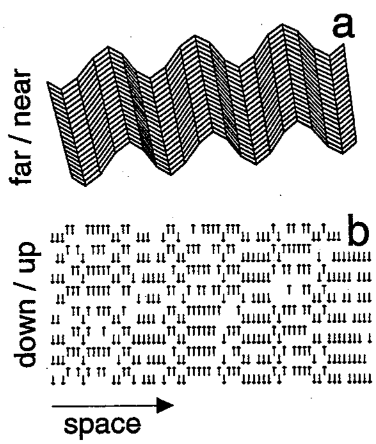


Figure 5.5 Depth and motion-defined motion gratings. (a) A single frame of the depth grating. The depth is sinusoidally modulated between near and far. The actual depth grating was composed of randomly black or white pixels; the depth resulted from sinusoidally modulating stereoscopic disparity (see figure 5.6). (b) A motion-defined motion stimulus. The arrows indicate the directions of dot motion between successive frames. Dots themselves have a lifetime of only two frames, and are reconstituted at a new random location when they have completed their jump. The pattern of motion modulation (up vs. down) moves either to the left or to the right. (Reprinted with permission from Lu and Sperling, 1996a.)

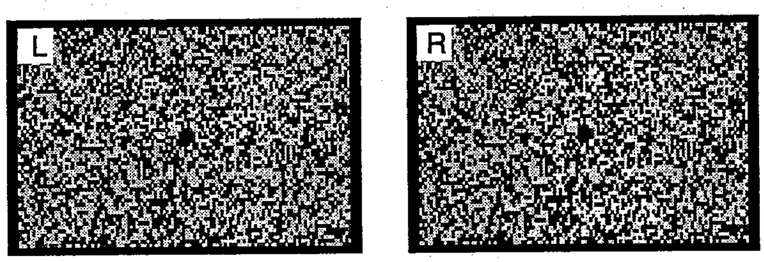


Figure 5.6 A stereogram illustrating a small section of one frame of the depth grating. Fixating the black dot of each panel with the corresponding eye should produce the perception of a horizontal depth grating. The depth in the illustration is vastly exaggerated relative to actual displays because of printing limitations. (Adapted with permission from Lu and Sperling, 1995a.)

frames. The proportion of upward vs. 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 (Petersik, Hicks, and Pantle, 1978; Cavanagh, Arguin, and von Grünau, 1989; Zanker, 1994) seems to suggest a hierarchical organization of motion detectors. The movement of the motion modulation (i.e., the horizontal movement of the up-down mo-

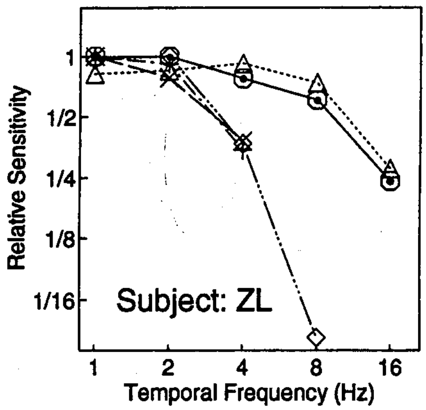


Figure 5.7 Tuning functions. The ordinate is the experimentally measured amplitude of the threshold modulation for correct discrimination of motion direction: sensitivity = $(\text{threshold})^{-1}$. The abscissa is the temporal frequency of a moving sinusoidal modulation. The axes are log scales. ○ 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-defined motion stimuli; ◇ indicates interocular sinusoidal luminance-motion stimuli. The curves have been vertically translated to expose their similarity in shape. (Reprinted with permission from Lu and Sperling, 1996a.)

tion 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 5.7 shows, the temporal tuning functions (detectability as a function of number of bars that move past a point on the retina in one second) for all the motion types show typical low-pass filter characteristics (curves slope down to the right), indicating that high temporal 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 effect on subjects' performances in the luminance and contrast modulation conditions. But pedestals reduce performance to chance-guessing levels with the depth and motion-defined 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 (first- and second-order motion systems) is immune to pedestals. Pending evidence to the contrary, we therefore assume that this class utilizes a motion-energy algorithm. Figure 5.7 shows that first- and second-order motion have a relatively high cutoff frequency (12 Hz). (The cutoff is that frequency at which sensitivity has declined by a factor of one-half). Interestingly, the second-order system has the same temporal frequency characteristics as the first-order system, despite frequent speculation that the second-order system is "slower" than the first-order system. The remaining types of motion are computed by what we have called the *third-order motion system.* The third-order motion system is slower than the first- and second-order systems, but it can extract motion from depth and motion-defined motion stimuli which are invisible to the first- and second-order systems.

A fundamental axiom of perception is that nothing can be perceived that is not computed. To perceive a wobble in the pedestaled stimuli of figure 5.3 requires a perceptual computation of wobble motion. Thus, perceiving both the linear and the wobble motion in both the first- and second-order stimuli of figure 5.3 immediately suggests a third-motion computation—presumably performed by the same versatile motion system that detects the motion of the depth and motion-defined motion stimuli.

FUNCTIONAL RELATIONSHIPS BETWEEN SYSTEMS

The pedestal experiments indicated the existence of three systems. Six subsequent procedures confirm these systems and clarify the relationships between them (Lu and Sperling, 1995b, 1996b, 1997).

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 (to avoid motion transparency). The stimuli were of equal strength in terms of the number of jnds (just noticeable differences) above threshold. What happened? When they moved in opposite directions, there was no apparent motion: The two motion signals canceled exactly. When they moved in the same direction without pedestals, there was enhanced apparent motion.

Same-direction motion strength equals or exceeds the prediction of 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, two stimuli of the same frequency and amplitude moving in the same direction but with a 180-degree phase difference (i.e., opposite sign) would perfectly cancel each other. If two stimuli combine before motion is computed (single-channel theory) there must be some phases that are better than either stimulus individually and others that are worse. The absence of any phase dependence (Lu and Sperling, 1995b) combined with the consistently much better (and never worse) detection of dual stimuli than individual stimuli means that first- and second-order motion strengths are computed by separate motion detectors, and only afterwards 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 degrees. In normal binocular or monocular viewing, motion in these stimuli were perceived as well as stimuli with eight (or more) frames per cycle (e.g., figure 5.3e and h). However, when successive frames were directed alternately into left and right eyes (Shadlen and Carney, 1986), subjects either could not perceive motion at all or had greatly reduced sensitivity. Under such interocular presentations, the motion stimulus in each eye of an observer is ambiguous, and perception of coherent motion would only be possible 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 degrees (figure 5.3e). We found that converting such a stimulus from monocular to interocular presentation (figure 5.8) raises the contrast threshold (at low frequencies) by a factor of twelve (from 0.17 percent to 2.0 percent) 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 (figure 5.7). 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

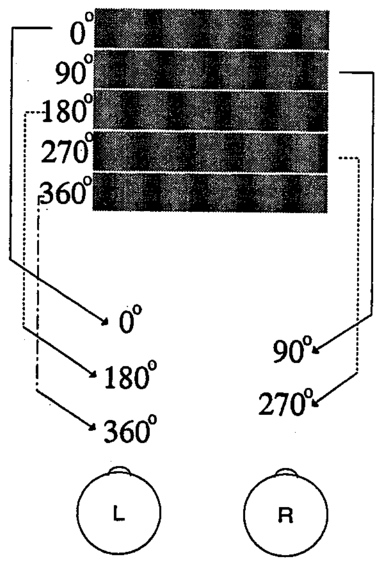


Figure 5.8 Representation of an interocular stimulus presentation in which frames are alternately directed to the left (L) and right (R) eyes. Each successive stimulus has a spatial phase shift of 90 degrees. Within an eye, the stimulus sequence, indicated on the bottom, is ambiguous as to direction of motion. (Reprinted with permission from Lu and Sperling, 1996a.)

when it detects motion in interocular luminance stimuli as it does when it detects motion in stereo-depth and in motion-defined motion stimuli.

On the other hand, interocular presentation of the motion-defined motion stimulus with no pedestal (figure 5.5) produces almost the same threshold for motion-direction discrimination as does monocular presentation of the same stimulus. This result indicates that the motion-defined motion computation is inherently binocular; it is indifferent to the eye of origin of any stimulus frame.

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 as 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 simple 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 system (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 ten 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 motion was thirty times less sensitive than first-order motion. (The relative temporal frequency sensitivities of the systems were given in figure 5.7; 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.

Drastically Different Contrast-Gain Control Properties of the First- and Second-Order Motion Systems

Using pedestal amplitude as the independent variable, we studied the nonlinear gain-control properties of the first-order (luminance) and the second-order (texture-contrast) motion systems, that is, how these systems' responses to motion stimuli are reduced by pedestals and other masking stimuli (Lu and Sperling, 1996a). Motion-direction thresholds were measured for test stimuli consisting of drifting luminance and texture-contrast modulation stimuli superimposed on pedestals of various amplitudes. (The pedestal was always a static sine-wave grating of the same type and same spatial frequency as the moving test grating, as in figure 5.3.)

First-order motion-direction thresholds are unaffected by small pedestals, but at pedestal contrasts above 1 to 2 percent (5× to 10× pedestal threshold), motion thresholds increase proportionally to pedestal amplitude (a Weber Law). For first-order stimuli, pedestal masking is specific to the spatial frequency of the test. On the other hand, motion-direction thresholds for texture-contrast stimuli are independent of pedestal amplitude (no gain control whatever) throughout the accessible pedestal amplitude range (from 0 percent to 40 percent). However, when baseline *carrier contrast* increases (with constant pedestal amplitude), motion thresholds increase.

In first-order motion, gain control is relatively specific to the spatial frequency of motion (e.g., Anderson, Burr, and Morrone, 1991) as indicated in figure 5.1. On the other hand, gain control in second-order motion is completely indifferent to the sine-wave modulator; gain control is determined entirely by the carrier, the amount of baseline contrast, independent of its distribution in space. Baseline contrast in second-order motion is analogous to luminance in first-order motion, so gain control in second-order motion is functionally analogous to luminance adaptation in first-order motion, although, of course, luminance adaptation acts in second-order motion as well.

The drastically different gain-control properties of the two motion systems and prior observations of motion masking and motion saturation are all encompassed in the systems flow chart of figure 5.1. The stimulus inputs to both first- and second-order motion process are normalized by feedforward shunting gain control. The different properties arise because texture grabbers similar to the modulator are used to control the first-order gain, whereas all texture grabbers are used to control the second-order gain (figure 5.1).

Selective Adaptation of Three Motion Systems

To selectively adapt the primarily monocular first-order motion system, we alternately present luminance sine-wave gratings moving in opposite directions in corresponding areas of the left and the right eyes (Lu and Sperling, 1997). Stimuli to the left eye and right eye alternate once per second for 10 seconds. To measure the magnitude of the motion aftereffect (MAE) immediately following adaptation, the observer judges the apparent direction of a monocular, pedestaled, first- or second-order motion stimulus of random amplitude, and gives a confidence rating. Adding stationary pedestals to the test stimuli ensures that the third-order system is not effective (it sees only back-and-forth wobble); and the pedestals themselves serve as MAE inducers.

To adapt and test the second-order motion system, procedures similar to first-order procedures are followed with second-order sine-wave gratings replacing first-order and vice versa. To adapt the binocular third-order system, interocular moving sine-wave gratings are employed. Dynamic random-noise displays with different proportions of "signal" dots moving in particular directions are used as test stimuli to measure the magnitude of the third-order MAE. When viewing these stimuli, observers are not aware of which eye or eyes are being stimulated.

All three types of adaptation stimuli produce significant, highly selective MAEs. The pattern of high-confidence responses indicates that the MAE produced strong perceptual illusions rather than merely instances of decision bias. The first notable finding is that adapting left and right eyes to opposite directions of movement by either a first- or a second-order stimulus in exactly the same perceived location produces an opposite MAE in each eye. Adapting to a first-order stimulus produced an MAE only for first-order test stimuli and not for second-order stimuli. Adapting to a second-order stimulus produced an MAE only for the second-order stimuli; there was absolutely no cross adaptation.

The interocular third-order stimuli had been shown previously to be invisible to first- or second-order systems. Nevertheless, the interocular stimuli produced strong third-order MAEs. These three modes of highly selective adaptation indicate three functionally distinct motion computations carried out in at least five different sites—the neural pathways from the left and right eyes each compute both first- and second-order motion, and there is a binocular site that adapts to third-order stimuli.

ATTENTIONAL INFLUENCES ON MOTION PERCEPTION

Selective attention appears to play no role in the perception of first- and second-order motion—at least for the case of a single location that contains a single (attended or unattended) type of motion stimulus or two superimposed stimuli of different types (Solomon and Sperling, 1994). On the other hand, 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 (Lu and Sperling, 1995a).

A novel alternating-feature paradigm was developed to study attention in third-order motion. Displays consisted of five frames. The odd frames (1, 3, 5) are 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 5.9. In the first example (top row), the odd frames are composed of adjacent areas of white dots and black dots. All the areas have the same average luminance and the same magnitude of microcontrast within each area. 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 systems 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 5.9a, 5.9c, 5.9e cannot stimulate either the first- or the second-order systems.

The even frames are composed of adjacent areas of high and low contrast. These frames are invisible to the first-order system because they all have the same average overall luminance. These frames 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, odd frames are composed of adjacent areas of fine and coarse textures that also differ in

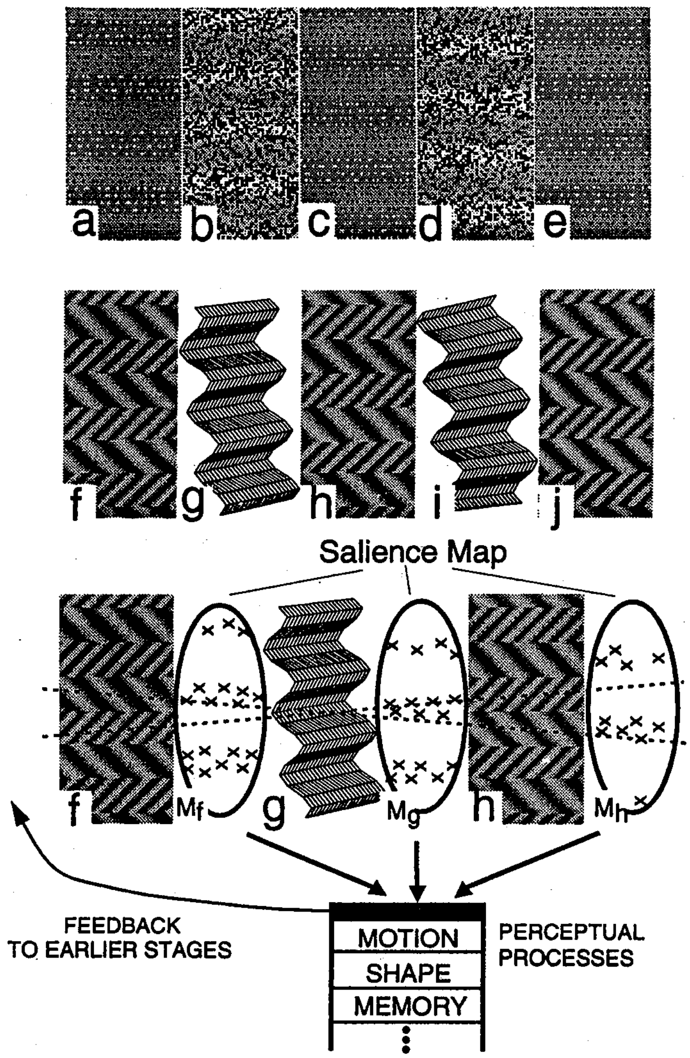


Figure 5.9 Alternating-feature stimulus sequences for attention-generated motion, and their relation to feature salience maps. The top row (a–e) shows an alternating-feature display with frames of black dot textures plus white dot textures (a, c, e) alternating with frames of low- and high-contrast textures (b, d). A sequence of five consecutive frames is shown; each is displaced vertically by 90 degrees 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 and their associated salience maps; the most salient features are marked with Xs. 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 dotted line. There is no support for upward motion (the 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. (Adapted with permission from Lu and Sperling, 1995a.)

orientation. Even frames consist of a stereo-depth grating—adjacent areas of near and far (figure 5.9, bottom).

The alternating-feature paradigm is analogous to the interocular paradigm in that successive frames are displaced 90 degrees, and within even frames or within odd frames, there is no motion signal. Unlike the interocular display, however, there is no motion between even and odd frames in the alternating-feature displays 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 them salient. This is 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 (figure 5.9, row 3). Insofar as images are perceptually segmented into areas of figure and areas of ground, being marked is equivalent to being designated as *figure*, and remaining unmarked is equivalent to *ground*. In high- and low-contrast frames (figure 5.9b and d), 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 (figure 5.9).

In formal experiments, subjects maintain rigid eye fixation while sequences of five successive frames (figure 5.9) are presented at a frequency of 7.5 frames per second. An entire display is completed in 667 milliseconds. 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 percent to 95 percent 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 milliseconds. With extremely brief displays, the attentional effect is reduced to about 70 percent, 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 milliseconds (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 (Cavanagh, 1992) but on an automatic motion computation (Lu and Sperling, 1995a). Indeed, the salience field would be a reasonable mechanism for guiding tracking as well as for computing motion.

Genetic and Computational Advantages of the Salience Motion Computation

In our theory, third-order motion differs from first-order and second-order motion primarily in the preprocessing of the input. The input for first-order motion is *photons*; a first-order stimulus consists of coherent movement of areas of greater concentrations of photons and of lesser concentrations. The input to the second-order motion system is defined by *features*, that is, simple types of patterns. For second-order motion, only the type of pattern needs to move, not the individual pattern (the token). (Movement of the patterns themselves, the tokens, would be first-order motion.) The input domain for third-order motion is the salience field. When an x, y location at time t in the salience field is "marked," this marking has the same status in third-order motion as do photons in first-order and features in second-order motion.

Once a piece of genetic code has developed to produce a neural circuit to compute a complex function, such as motion (i.e., space-time covariance), that genetic code can be repeated and used over and over again at different levels of neural computation. It is much more likely that a piece of genetic code would be reproduced and adapted to other uses than that an equivalent code would be discovered independently. Mini-algorithms, such as gain control, band-pass filtering (center-surround receptive fields), and motion computations recur at many levels in the system flow charts of figures 5.1 and 5.10. So, it makes great sense that the motion computation is repeated at successively more abstract levels to compute space-time pattern covariance: initially on photons, then on features, and, ultimately, on importance (i.e., salience).

From a computational point of view, greater abstraction allows greater flexibility. Suppose one wanted to compute the motion of red-coated objects in a complex field of objects. One might develop a special motion system designed for red (vs. other colors). Alternatively, one might develop a motion system to compute the motion of "important" objects, and another system to designate arbitrary objects, such as red coats, as "important." Specialized systems will develop when there is persistent evolutionary pressure, e.g., the first- and second-order systems. The abstract computation performed by the third-order motion system would accomplish many spur-of-the-moment motion tasks for which there may be an occasional need. It would be ex-

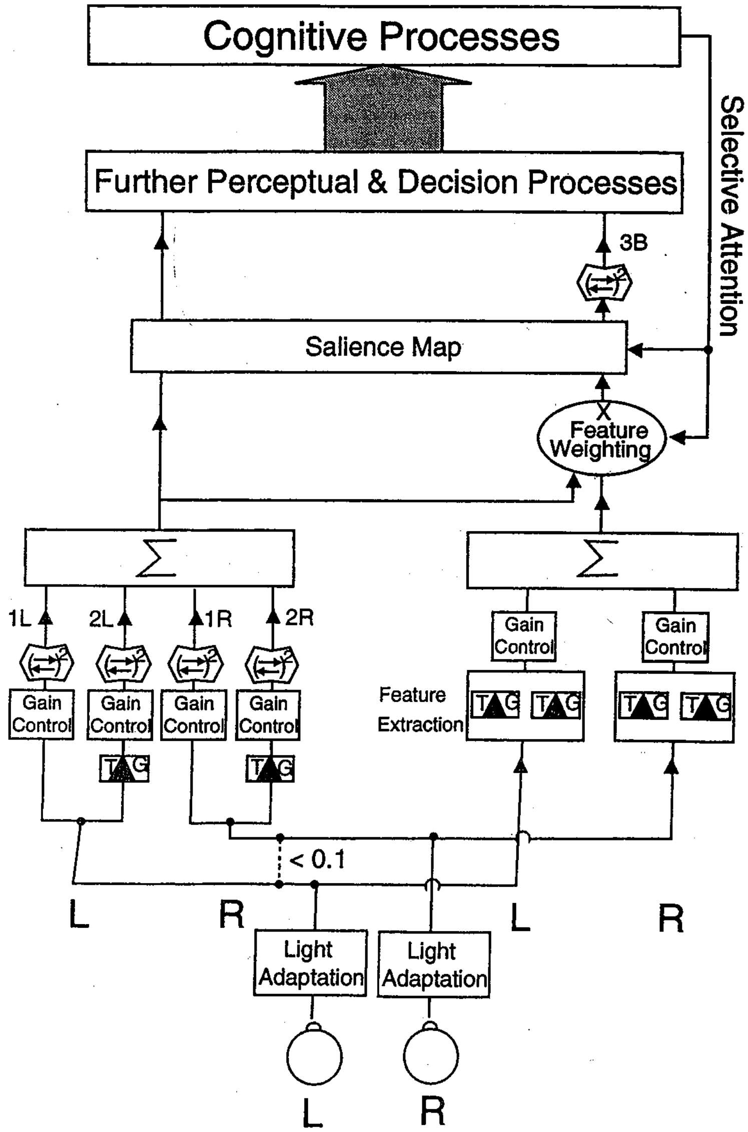


Figure 5.10 Functional architecture of the visual motion system. The fast, primarily monocular first- and second-order motion systems are represented on the left; third-order motion is represented on the right. L and R indicate left and right eye signals, respectively. The dotted line adjacent to 0.1 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 full-wave 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 salience weighting of features determined by selective attention. Attention also acts directly on the salience field to determine the spatial locations that will determine subsequent processing; this applies to all visual inputs. 3B indicates the binocular third-order motion computation based on the salience field. The connecting path from motion-energy Σ to feature weighting conveys the motion features needed to solve motion-defined motion stimuli. (Based on Lu and Sperling, 1995b, with permission.)

tremely wasteful, and ultimately computationally disadvantageous, to build too many special-purpose motion mechanisms. Each new system adds noise —accidental correlations—to the overall computation, while adding genetic complexity and metabolic costs. More is not necessarily better.

THE BIG PICTURE

The organization of the three motion systems is summarized in figure 5.10. This schematic flowchart is built around motion-direction discrimination which is preliminary to many other motion computations. The first- and second-order motion systems provide raw motion data that require further perceptual processing in order to extract velocity (e.g., Heeger, 1987), to determine heading direction, to infer a 3D structure from 2D motion (Sperling et al., 1989), and so forth. Ordinarily, only the first-order system has sufficient resolution to compute structure from motion (Dosher, Landy, and Sperling, 1989). All these subsequent processes are encompassed in the box entitled *Further Perceptual and Decision Processes* of figure 5.10.

How Many Components Are There in the Big Picture?

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 5.10. The figure contains about two dozen computational "boxes." Many of these represent computations that can be analyzed into components. For example, the motion boxes $(\rightleftarrows)^2$ represent Reichardt detectors which consist of eight components (figure 5.2). The gain-control mechanisms are described in figure 5.1 and contain four (or more) components. Texture grabbers (Chubb and Sperling, 1988, 1989a, 1989b, 1991) contain three components (figure 5.4). Much is known about decision processes, although their analysis into components is beyond the present scope. However, the combination of motion signals from different sources (indicated by the Σ in figure 5.10) is still quite mysterious.

The two dozen boxes of figure 5.10 contain about one hundred known subcomponents. Figure 5.10 represents processing in only one horizontal slice through the pyramid that represents the different spatial scales. All the complexity of figure 5.10 is reproduced at finer and coarser resolutions. Thus, a systems model requires on the order of many hundreds of individual components to represent motion-direction processing within a small neighborhood. This complexity must again be reproduced in all neighborhoods, central and peripheral, of visual space. If there is a moral in this, it is that perceptionists have tended to grossly underestimate the complexity of visual computation.

Salience Map Operations

Figure-Ground Segregation The salience map is the focal point of the mechanisms of spatial attention; its utility for motion is secondary. Figure-ground marking in the salience map determines what is admitted to shape recognition processes. For example, mapmakers long ago discovered how to color and to fill in continents so that when we look at a global map, we see the continents as figure and the oceans as ground. In systems terms, the areas of the salience field that represent continents are marked automatically, and therefore information from these areas is forwarded to shape recognition processes. The unmarked area, the ground, is not processed further—at least not to the same extent as figure. Because the shape of the ground is not computed, we do not recognize the shape of oceans. Similarly, we compute the shapes of trees and not the shapes of spaces between trees. However, when we are seeking a place to escape, the space between trees becomes attentionally relevant, and we can compute the shape of that space to determine whether or not we would fit through.

Visual Search 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 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 (Shih and Sperling, 1996). 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 searching for particular features, selective attention is mediated via locations of the attended features.

Partial Reports 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., number 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 rapid dynamic access to visual short-term memory is mediated via spatial location, not by properties of the material to be stored (Sperling, 1960).

The Salience Model All these phenomena are encompassed within the salience model 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). Selectively attended features have a larger influence in determining what is marked in the salience map. Top-down and bottom-up interaction produce the current salience map, a dynamic map of the locations of the most salient stimulus features. The salience map can be used directly to compute salience motion. Or the salience map can be used to guide bottom-up processing (as in figure-ground segregation and attention-guided search), or to control access to visual memory (as in partial report experiments).

True Objectless Motion

Pedestaled Displays without Wobble When subjects view a pedestaled display, their first- or second-order motion systems compute the linear motion of the moving grating, and the third-order system computes the wobble of the combined moving and pedestal stimulus (figure 5.3). Because gain control in the second-order system is insensitive to pedestal amplitude, one can produce pedestals with amplitudes limited only by the inherent optics of images, and that have no effect whatsoever on threshold of the moving grating (Lu and Sperling, 1996b). With a pedestal that is 5x or 10x the amplitude of the motion signal, there is no detectable wobble. Similarly, because third-order motion is relatively insensitive to high temporal frequencies, pedestaled motion displays of 10 Hz gratings rest on an apparently unmoving pedestal. Nevertheless, the first- and second-order motion systems produce their normal signal. The third-order motion system (and the shape-recognition system) correctly report that there is a stationary grating—the pedestal. These cases represent true objectless motion because there is no object to which the motion signal can be attached. The subject simply reports the output of both the shape-recognition system (a stationary grating) and the output of the first- or second-order motion system (motion to the left or right) but cannot say *what* is moving. The sensation of movement is attached to an area of the visual field but not to an object because the only object visible there is reported as "stationary" by the third-order system, and its report has priority.

Blindsight True objectless motion is reminiscent of blindsight. Certain brain injuries can selectively destroy the visual shape-recognition system. Nevertheless, subjects may report sensations of motion because motion is computed in another area of the brain, and the motion system's output may still wend its way to higher-level perceptual and motor computations. Such patients cannot say what is moving because they cannot compute shape; but they can say where in the visual field movement might be occurring.

Forward Motion 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. As in the case of atoms, we expect that, in the future, these processes will be further subdivided and additional ones discovered.

ACKNOWLEDGMENTS

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

NOTES

1. This article incorporates a partial reprinting (with permission of the American Psychological Society) of an earlier article (Lu and Sperling, 1996a) together with elaborations and additions. For a more detailed account of the work described herein, and for additional references, the reader should consult two recent summary articles: Lu and Sperling, 1995a; 1995b.

2. The original proof of two properties of Reichardt models, pseudolinearity and ignoring static displays, applied to continuous displays of infinite duration (van Santen and Sperling, 1984). Only a restricted version of this proof can be extended to sampled displays of finite duration. Specifically, for a sinewave stimulus that moves forward 90 degrees between successive frames, any number of full cycles (of four frames) plus one extra frame provides pedestal resistance. Pedestal resistance becomes full immunity whenever the aliased backward motion component with one-third the forward fundamental temporal frequency can be neglected. When a different number of frames or a different shift between frames is used, the output of a Reichardt detector depends complexly on the relative phases of the moving and pedestal components.

REFERENCES

Adelson, E. H., and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, A 2*, 284–299.

Albrecht, D. G., and Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience, 7*, 531–546.

Albrecht, D. G., and Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology, 48*, 217–237.

Anderson, S. J., Burr, D. C., and Morrone, M. C. (1991). Two-dimensional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision. *Journal of the Optical Society of America, A 8*, 1340–1351.

Blackwell, H. R. (1946). Contrast thresholds of the human eye. *Journal of the Optical Society of America, 36*, 624–646.

Blakemore, C. B., and Campbell, F. W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retina images. *Journal of Physiology, London, 203*, 237–260.

Bonds, A. B. (1991). Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Visual Neuroscience, 6*, 239–255.

Boulton, J. C., and Baker, C. L. (1993). Dependence on stimulus onset asynchrony in apparent motion: Evidence for two mechanisms. *Vision Research, 33,* 2013–2019.

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research, 14,* 519–527.

Burr, D. C., and Ross, J. (1982). Contrast sensitivity at high velocities. *Vision Research, 22,* 479–484.

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

Cavanagh, P., Arguin, M., and von Grünau, M. (1989). Interattribute apparent motion. *Vision Research, 29,* 1197–1204.

Cavanagh, P., and Mather, G. (1989). Motion: The long and the short of it. *Spatial Vision, 4,* 103–129.

Chubb, C., and Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America, A 5,* 1986–2007.

Chubb, C., and Sperling, G. (1989a). Two motion-perception mechanisms revealed through distance-driven reversal of apparent motion. *Proceedings of the National Academy of Science, 86,* 2985–2989.

Chubb, C., and Sperling, G. (1989b). Second-order motion perception: Space-time separable mechanisms. In *Proceedings: IEEE Workshop on Visual Motion,* 126–138. Washington D. C.: IEEE Computer Society Press.

Chubb, C., and Sperling, G. (1991). Texture quilts: Basic tools for studying motion from texture. *Journal of Mathematical Psychology, 35,* 411–442.

Dean, A. F. (1981). The relationship between response amplitude and contrast for cat striate cortical neurons. *Journal of Physiology, London, 318,* 413–427.

Derrington, A. M., and Badcock, D. R. (1985). Separate detectors for simple and complex grating patterns? *Vision Research, 25,* 1869–1878.

Derrington, A. M., and Lennie, P. (1981). Spatial and temporal contrast sensitivities of neurons in lateral geniculate nucleus of macaque. *Journal of Physiology, London, 357,* 219–240.

Dosher, B. A., Landy, M. S., and Sperling, G. (1989). Kinetic depth effect and optic flow. I: 3-D shape from Fourier motion. *Vision Research, 29,* 1789–1813.

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

Fleet, D. J., and Jepson, A. D. (1985). On the hierarchical construction of orientation and velocity selective filters. Technical Report RBCV-TR-85-8. Toronto: University of Toronto, Computer Science Department.

Heeger, D. J. (1987). A model for the extraction of image flow. *Journal of the Optical Society of America, A 4,* 1455–1471.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9,* 181–197.

Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing and test field luminances. *Journal of Experimental Psychology, 50,* 89–96.

Heinemann, E. G. (1972). Simultaneous brightness induction. In D. Jameson and L. M. Hurvich, eds., *Handbook of Sensory Physiology.* Berlin: Springer Verlag, vol. 4, 146–149.

Hood, D. C., and Finkelstein, M. A. (1986). Sensitivity to light. In K. R. Boff, L. Kaufman, and J. P. Thomas, eds., *Handbook of Perception and Human Performance.* New York: Wiley, vol. 1, chapter 5.

Jamar, J. H., and Koenderink, J. J. (1985). Contrast detection and detection of contrast modulation for noise gratings. *Vision Research, 25,* 511–521.

Kaplan, E., and Shapley, R. M. (1982). X and Y cells in the lateral geniculate nucleus of macaque monkeys. *Journal of Physiology, 330,* 125–143.

Kelly, D. H. (1979). Motion and vision. II: Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America, 69,* 1340–1349.

Kelly, D. H. (1982). Motion and vision. IV: Isotropic and anisotropic spatial responses. *Journal of the Optical Society of America, 72,* 432–439.

Koenderink, J. J., and van Doorn, A. J. (1979). Spatiotemporal contrast detection threshold surface is bimodal. *Optical Letter, 4,* 32–34.

Lu, Z.-L., and Sperling, G. (1995a). Attention-generated apparent motion. *Nature, 377,* 237–239.

Lu, Z.-L., and Sperling, G. (1995b). The functional architecture of human visual motion perception. *Vision Research, 35,* 2697–2722.

Lu, Z.-L., and Sperling, G. (1996a). Three systems for visual motion perception. *Current Directions in Psychological Science, 5,* 44–53.

Lu, Z.-L., and Sperling, G. (1996b). Contrast-gain control in first- and second-order motion perception. *Journal of the Optical Society of America, A 13,* 2305–2318.

Lu, Z.-L., and Sperling, G. (1997). Selective adaptation of three motion systems. *Investigative Ophthalmology and Visual Science, 38,* ARVO Abstract Book, part 1, S237.

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

Mather, G., Cavanagh, P., and Anstis, A. M. (1985). A moving display which opposes short-range and long-range signals. *Perception, 14,* 163–166.

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

Nakayama, K., and Silverman, G. H. (1985). Detection and discrimination of sinusoidal grating displacements. *Journal of the Optical Society of America, A 2,* 267–273.

Ohzawa, I., Sclar, G., and Freeman, R. D. (1982). Contrast-gain control in the cat visual cortex. *Nature, 298,* 266–268.

Pantle, A., and Picciano L. (1976). A multistable movement display: Evidence for two separate motion systems in human vision. *Science, 193,* 500–502.

Parish, D. H., and Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research, 31,* 1399–1415.

Pavel, M., Sperling, G., Riedl, T., and Vanderbeek, A. (1987). Limits of visual communication: The effect of signal-to-noise ratio on the intelligibility of American Sign Language. *Journal of the Optical Society of America, A 4,* 2355–2365.

Pelli, D. G. (1981). Effects of visual noise, Ph.D. dissertation. Cambridge, U.K.: University of Cambridge.

Petersik, J. T., Hicks, K. I., and Pantle, A. J. (1978). Apparent movement of successively generated subjective figures. *Perception, 7,* 371–383.

Ramachandran, V. S., Rao, M. V., and Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Research, 13,* 1399–1401.

Reichardt, W. (1957). Autokorrelationsauswertung als funktionsprinzip des zentralnervensystems, *Zeitschrift Naturforschung, 12b,* 447–457.

Reichardt, W. (1961). Autocorrelation: A principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith, ed., *Sensory Communication,* 303–317. Cambridge: MIT Press and New York: Wiley and Sons.

Robson, J. G. (1966). Spatial and temporal contrast-sensitivity functions of the visual system. *Journal of the Optical Society of America, 56,* 1141–1142.

Sachs, M. B., Nachmias, J., and Robson, J. G. (1971). Spatial frequency channels in human vision. *Journal of the Optical Society of America, 61,* 1176–1186.

Sclar, G., Maunsell, J. H., and Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research, 30,* 1–10.

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

Shapley, R., and Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. In N. Osborne and G. Chader, eds., *Progress in Retinal Research, vol. 3,* 263–346. Oxford and New York: Pergamon Press.

Shih, S., and Sperling, G. (1996). Is there feature-based stimulus selection in visual search? *Journal of Experimental Psychology: Human Perception and Performance, 22,* 758–779.

Smith, A. T. (1994). Correspondence-based and energy-based detection of second-order motion in human vision. *Journal of the Optical Society of America, A 11,* 1940–1948.

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

Sperling, G. (1960). The information available in brief visual presentations. *Psychology Monograph, 74,* no. 11, (whole no. 498).

Sperling, G. (1970). Model of visual adaptation and contrast detection. *Perception and Psychophysics, 8,* 143–157.

Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behavior Research Methods and Instrumentation, 8,* 144–151.

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

Sperling, G., and Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of Optical Society of America, 58,* 1133–1145.

Sperling, G., Landy, M. S., Dosher, B. A., and Perkins, M. E. (1989). The kinetic depth effect and the identification of shape. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 826–840.

Turano, K., and Pantle, A. (1989). On the mechanism that encodes the movement of contrast variations. Velocity discrimination. *Vision Research, 29,* 207–221.

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

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

Victor, J. D., and Conte, M. M. (1990). Motion mechanisms have only limited access to form information. *Vision Research, 30,* 289–301.

Watson, A. B., and Robson, J. G. (1981). Discrimination at threshold: Labelled detectors in human vision. *Vision Research, 21,* 1115–1122.

Watson, A. B., Barlow, H. B., and Robson, J. G. (1983). What does the eye see best? *Nature, 302,* 419–422.

Watt, R. J., and Morgan M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25,* 1661–1674.

Werkhoven, P., Sperling, G., and Chubb, C. (1994). Perception of apparent motion between dissimilar gratings: Spatiotemporal properties. *Vision Research, 34,* 2741–2759.

Whittle, P., and Challands, P. D. C. (1969). The effect of background luminance on the brightness of flashes. *Vision Research, 9,* 1095–1110.

Wilson, H. R., and Bergen, J. (1979). A four-mechanism model for threshold spatial vision. *Vision Research, 19,* 19–32.

Wilson, H. R., Ferrera, V. P., and Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience, 9,* 79–97.

Zanker, J. M. (1994). What is the elementary mechanism underlying secondary motion processing? *Investigative Ophthalmology and Visual Science, Suppl. 35,* 1405.