Three-systems theory of human visual motion perception: review and update

Zhong-Lin Lu

Laboratory of Brain Processes, Department of Psychology, University of Southern California, Los Angeles, California 90089-1061

George Sperling

Human Information Processing Laboratory, Departments of Cognitive Sciences and Neurobiology and Behavior, and Institute of Mathematical Behavioral Sciences, University of California, Irvine, Irvine, California 92697

Received January 24, 2001; revised manuscript received May 22, 2001; accepted May 31, 2001

Lu and Sperling [Vision Res. **35**, 2697 (1995)] proposed that human visual motion perception is served by three separate motion systems: a first-order system that responds to moving luminance patterns, a second-order system that responds to moving modulations of feature types—stimuli in which the expected luminance is the same everywhere but an area of higher contrast or of flicker moves, and a third-order system that computes the motion of marked locations in a "salience map," that is, a neural representation of visual space in which the locations of important visual features ("figure") are marked and "ground" is unmarked. Subsequently, there have been some strongly confirmatory reports: different gain-control mechanisms for first- and second-order motion, selective impairment of first- versus second- and/or third-order motion by different brain injuries, and the classification of new third-order motions, e.g., isoluminant chromatic motion. Various procedures have successfully discriminated between second- and third-order motion (when first-order motion is excluded): dual tasks, second-order reversed phi, motion competition, and selective adaptation. Meanwhile, eight apparent contradictions to the three-systems theory have been proposed. A review and reanalysis here of the new evidence, pro and con, resolves the challenges and yields a more clearly defined and significantly strengthened theory.

OCIS codes: 330.4150.

1. INTRODUCTION

The history of multisystem motion theory dates back to the beginning of theoretical study on human visual motion perception.[1] Turn-of-the-century researchers used Greek letters, α,[2] β,[2] δ,[3] γ,[2] and ϕ,[4] to distinguish motion stimuli that produced different perceptual experiences. Recent examples of two-system motion theories are short-range versus long-range motion[5–8]; first-order versus second-order[9,10] and Fourier versus non-Fourier motion.[11] Short-range motion, Fourier motion, and first-order motion seem to refer to the same class of phenomena (here referred to as first-order motion), for which there are several competing computational theories. Some first-order motion theories are quite similar or in many cases equivalent, e.g., Reichardt detectors[12–15] and motion-energy detectors,[16,17] versus some that can be quite different, such as zero-crossing detectors[18] and gradient detectors.[19,20] While there is clearly an element of truth to all these multisystem motion theories, it is impossible to distinguish the theories without (1) explicit computational models that specify the algorithms and (2) experiments that isolate the proposed motion systems.

The first computational model of motion perception was proposed by Hassenstein and Reichardt[21] and further developed by Reichardt[12,13] in studying insect motion perception. The Reichardt model was successfully elaborated as a theory of human perception by van Santen and Sperling.[14] Several apparently different models, e.g., Adelson and Bergen's[16] motion energy model and a linear computation based on Hilbert transforms[22] were proved[15] to be computationally equivalent to a Reichardt model in the sense of having the same overall system input/output function, although the internal sequence of operations differs.

While the Reichardt model in its elaborated form could quite accurately account for motion perception in a wide range of stimuli[15,16,22–25] including some quite complex waveforms and quite counterintuitive predictions,[14] many investigators have reported classes of motion stimuli that are easily perceived by human observers but cannot be detected by Reichardt detectors.[26–32] It had been believed for some time that at least two motion systems are necessary to account for the variety of stimuli in which motion can be perceived. Chubb and Sperling[30] proposed and provided experimental evidence for an explicit computational model to differentiate between a Fourier motion system that extracts motion directly from the luminance modulation (photons) using a Reichardt (or an equivalent motion energy) detector and a non-Fourier motion system that first computes the amount of texture (features) in the stimulus by means of texture grabbers and then submits the outputs of the texture grabbers to Reichardt (or equivalent) computations.[11,30,33] Sperling[34] observed that, although the paradigms were somewhat different, the motion dichotomies first-order versus second-order (e.g., Ref. 9) and Fourier versus

0740-3232/2001/092331-40$15.00

non-Fourier[30] seemed to refer to the same classifications of motion phenomena and suggested that the more neutral terms first- and second-order be used to unify the terminology.[9,11]

Once explicit computations had been proposed for first- and second-order motion, it quickly became evident that a large body of experimental results[5,35–42] did not fit into the proposed scheme of two motion systems. Psychophysical experiments with complex motion stimuli led us to propose a three-system functional architecture for human visual motion perceptual computation[41,43,44]: The first-order system[12,14,16,22] extracts motion from moving luminance modulations. The second-order system[9,11,30] extracts motion from moving stimuli in which the expected luminance is the same everywhere but some features (i.e., contrast and/or spatial frequency) deviate from the background. The third-order system detects movement of feature salience, that is, changes in location of visual areas marked as "important" or as "figure" versus "ground".[40] Both the first- and the second-order systems use a primitive motion-energy algorithm, are primarily monocular, and are fast. The third-order mechanism is binocular, slow, but extremely versatile, and is strongly influenced by attention. A binocular first-order plus separate left- and right-eye computations for first-order and second-order imply a total of at least six perceptual motion computations, all carried out concurrently (Fig. 1).

Considerable evidence from the literature, both prior[5–11,35–38,45–49] and subsequent[50–66] to the publication of the Lu–Sperling three-systems theory seems to support a trichotomy of motion perception.

There have been eight major challenges to both the three-systems theory and its supporting evidence: (1) The apparent success of a single-system model that utilizes an up-front, pointwise, nonlinear transformation of input contrast prior to the motion computation[67] chal-

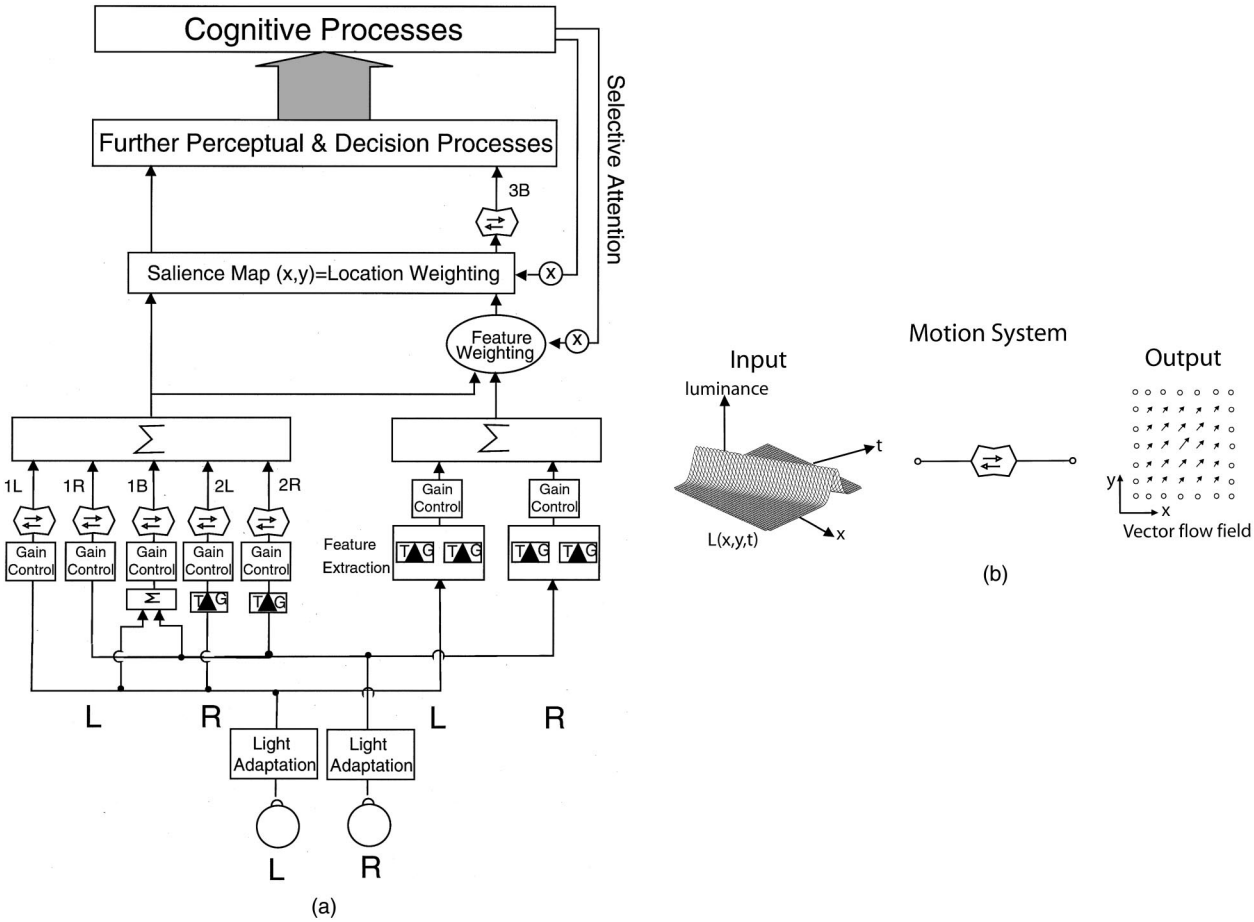


Fig. 1. (a) Schematic functional architecture of the visual motion system. Left, first- and second-order motion systems. Right, third-order motion. L and R, signals originating in the left and the right eye, respectively. Motion-energy detectors ($\rightleftarrows$) acting on the spatially filtered input serve monocular first-order (luminance-modulation) motion (1L and 1R). Σ indicates (possibly complex) summation/combination. The Σ-summed L and R inputs serve binocular (including interocular) first-order motion (1B). Second-order motion (2L, 2R) requires a texture grabber TG (a spatial filter followed by full-wave rectification, e.g., absolute value of each point's difference from mean luminance) before a motion computation. $\times$ represents multiplication—the differential salience-weighting of features determined by selective attention and the selective weighting of locations in the salience field to determine which spatial locations will be subsequently processed; salience ultimately applies to all visual inputs. 3B indicates the binocular third-order motion computation based on salience values. The connecting path from summed motion energy Σ to feature weighting conveys the motion features needed to solve motion-defined motion stimuli. (b) Definition of a motion system: a computation, presumably carried out in a brain nucleus, that takes a space–time representation of the visual field as input and produces a vector flow field that represents motion strength or velocity or some combination. The example illustrates a blurred point (or line) in linear motion and an instantaneous motion-strength flow field.

lenges the existence of independent first- and second-order systems. (2) Predictions of certain second-order motion phenomena using a multichannel gradient model,[20] developed to compute velocity from first-order rigid movements of two-dimensional images, may unify first- and second-order motion systems. (3) The claimed capability of a single "activity" algorithm[68] challenges the necessity of an independent second-order motion system. (4) A postulated qualitative difference between first-order and second-order motion mechanisms (not merely preprocessing): Whereas velocity is the cue to first-order motion, position displacement is asserted to be the cue to second-order motion detection.[69] (5) Patterson[42] concluded that stereoscopic (cyclopean) motion is processed by a motion-sensing system composed of special-purpose mechanisms that function like low-level motion sensors. (6) An interocular stimulus that passed the pedestal test[70] leads to the conclusion that first-order motion is binocular, not just monocular. (7) Failure of pedestal immunity for brief stimuli[71] challenges motion-energy computation as the algorithm for first-order motion extraction. (8) Alleged existence of various first-order (luminance) artifacts in second-order motion stimuli[72] questions the validity of a large body of experiments involving second-order motion stimuli.

In this paper, we first review the Lu–Sperling three-systems theory and the major evidence on which it is based. We then consider new studies: first, the eight most significant challenges to the theory (listed above), and finally, six studies not specifically posed as challenges, which provide useful new information about components of the theory.

2. THE THREE-SYSTEMS THEORY OF MOTION PERCEPTION

A. Definitions and Basic Computational Principles

1. Motion System

A motion system is a computation, presumably carried out in a brain nucleus, that takes as input a space–time representation of the visual field and produces a motion flow field as output. In a motion flow field, each space–time neighborhood of the visual field is represented by a vector that indicates the direction of motion in that neighborhood and one or more quantities that represent estimated speed, and/or motion strength, and/or some combination of motion strength and speed. Motion strength represents the visual system's confidence that motion is actually occurring in the indicated direction. Strength is related, for example, to the amplitude of Fourier coefficients that represent the indicated motion, to the signal-to-noise ratio of the motion stimulus, and to similar quantities. The theme of this article is that the visual system performs three different computations of the directional strength of visual motion. Computations of speed are not considered here.

2. First-Order Motion

The Reichardt detector (Fig. 2), a basic motion detector in computational theories of motion perception, was originally developed for insect vision by Reichardt[12,13] and successfully adapted by van Santen and Sperling[14] for human perception. It consists of two mirror-image subunits (e.g., "Left" and "Right") tuned to opposite directions of motion. Subunit L multiplies the signal at spatial location A with the *delayed* signal from an adjacent spatial location B. Subunit R multiplies the signal at spatial location B with the *delayed* signal from spatial location A. The difference of the outputs of the two subunits indicates the (signed) direction of motion. The Reichardt detector provides an accurate account of motion direction for a wide range of stimuli,[14,23–25] including some quite complex waveforms and quite counterintuitive predictions.[14]

3. Second-Order Motion

There are some obvious problems with Reichardt detectors, as exemplified by the stimulus shown in Fig. 3(c)—a contrast-modulated random-noise pattern in which the random-noise pattern ("carrier") is stationary; the expected luminance is the same everywhere; only the modulator moves. We have yet to find an observer (without a brain lesion) who can see motion in first-order stimuli who does not perceive the obvious apparent motion from such a stimulus. Nevertheless, the motion of the moving contrast modulation [Fig. 3(c)] is invisible to Reichardt detectors[30] (Note: First- and second-order motion are descriptions of perceptual algorithms, not of stimuli. However, for brevity and to better convey the reasoning, the short phrase *first-order stimulus* is used here to designate *a luminance modulation stimulus intended to primarily stimulate the first-order motion system*. The phrase *second-order stimulus* will be used below to designate *a contrast-modulation stimulus directed primarily to the second-order motion system*. Of course, intending a

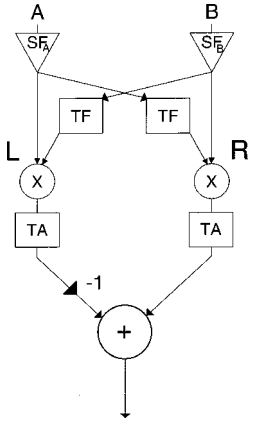


Fig. 2. Elaborated Reichardt detector. It computes motion direction from two inputs that sample the visual stimulus in two spatial areas A and B. SF_A and SF_B denote the linear spatiotemporal filters (receptive fields) that may have different spatial distributions. In the R subunit of the detector, the output of SF_A at A is delayed by a temporal delay filter TF and then multiplied ($\times$) by the direct output of SF_B at B. In the L subunit of the detector, the output of SF_B at B is delayed by a temporal delay filter TF and then multiplied ($\times$) by the direct output of SF_A at B. TA (temporal averaging) represents a low-pass temporal filter. The sign of the difference between the outputs of L and R subunits determines the perceived direction of motion. Outputs greater than zero indicate stimulus motion from A to B; outputs less than zero indicate stimulus motion from B to A.

stimulus to a motion system does not guarantee that it stimulates that system or only that system.)

Chubb and Sperling[30,33] demonstrated clear motion perception in broad classes of [drift-balanced and microbalanced] stimuli constructed of drifting modulations of contrast, spatial frequency, texture type, or flicker (see also Refs. 26–32) whose motion was mathematically proved to be invisible to any Reichardt detector. Such stimuli were said to activate second-order motion analyses.[9,11] According to Chubb and Sperling,[30] the input to the second-order motion system is the modulation of texture contrast. For the most-diagnostic second-order stimulus, texture quilts, Chubb and Sperling[33] proved that no pointwise transformation could make the motion accessible to Reichardt or to motion energy detectors. (In a texture quilt, a patch of a texture type moves consistently from frame to frame, and the successive patches are independent, uncorrelated samples. For example, frame 1 consists of side-by-side patches of right-slanting gratings separated by patches of left-slanting gratings. All patches have the same expected luminance and overall contrast. In frame 2, the patch pattern is translated sideways, and new patch samples are chosen. And so on.)

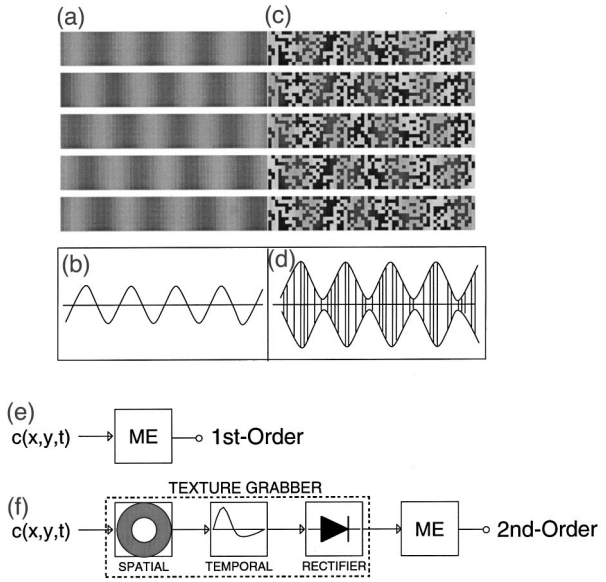


Fig. 3. First- and second-order motion stimuli. (a) Five consecutive frames of a first-order motion stimulus: a moving sine-wave luminance modulation. (b) Graphical representation of the sinusoidal intensity modulation of frame one in (a). (c) Five consecutive frames of a second-order motion stimulus: a moving contrast modulation of a random-texture carrier. (d) Schematic representation of one horizontal line of (c). The random function represents the carrier texture; the envelope is the modulator. To create an impression of motion, the modulators in (b) and (d) translate horizontally from frame to frame. (e) A motion-energy computation, ME, suffices to extract motion from the point-contrast function $c(x, y, t)$ that describes the stimulus in (a). (f) Second-order contrast-modulation motion is detected by first extracting the textural features [by means of a texture grabber represented by the first three boxes in (f)], and then computing motion energy, ME, from the features (delivered by the texture grabber) in the same way as the first-order motion system computes the motion energy of its input. The texture grabber consists of three stages: an isotropic linear spatial filter, SPATIAL; a temporal bandpass filter, TEMPORAL; and rectification (absolute value or square).

No degree of differentiation nor rectification nor any pointwise nonlinearity could extract sufficient information to permit subsequent motion energy detection.

To extract motion from texture quilts, a mechanism first has to extract features, which are an inherently spatial construct. Feature extraction is served by a texture grabber,[11] consisting of a linear spatial filter, a linear temporal filter, and a gross nonlinear transformation (Fig. 3), such as full-wave rectification. (Rectification is defined in terms of point contrast, the deviation of the luminance at a point in an image from the mean luminance of the image divided by the mean luminance. Full-wave rectification refers to any monotonically increasing function of the absolute value of pointwise contrast, typically the absolute value itself or the square of the point contrast). A linear spatial filter, exemplified by a specific size of center–surround configuration, or by an oriented, spatial-frequency-tuned Gabor filter, defines the texture. The stimulus is processed in parallel by a field of such self-similar filters. The outputs of the filters are then full-wave rectified.[73] That is, positive and negative outputs are treated equally.

Processes after full-wave rectification know how much of the feature was present in a neighborhood, but not whether it happened to have a positive or negative phase. Indeed, if inputs were not rectified, the expected output value of bandpass filters for second-order input stimuli would be zero. A texture grabber also includes a temporal bandpass filter that sensitizes it to changes in texture. In second-order motion processing, motion is extracted from the outputs of a field of texture grabbers by Reichardt motion detectors.

4. Third-Order Motion

The existence of purely binocular, interocular, and various other unusual kinds of apparent motion[35–39,42,45,46] has prompted conjectures of complex motion-processing systems,[6,8] but until first- and second-order processing were clearly defined, it was not evident that a third motion-processing system would be required. Lu and Sperling[40,41] demonstrated four classes of stimuli that would be invisible to first- and second-order motion processing as previously defined, including one in which selective attention to a feature determines the direction of apparent motion.

The need for a third motion-perception system is perhaps best illustrated by the example in Fig. 4, in which a square moves from left to right in successive frames. In frame 1 the square is defined by texture—the square contains texture slanted in one way, and the rectangular background contains texture slanted in another way. In frame 2, the square is defined by texture contrast—the square is composed of high-contrast texture, the background of low-contrast texture. In frame 3, the square is defined by a random-dot stereogram—the square is perceived in front of the background. In frame 4 the square is defined by isoluminant color—the square is red, the background is green. Such left-to-right motion is easily and compellingly perceived; i.e., the motion perception itself does not seem qualitatively different to an observer from the motion perception resulting from a single feature, i.e., a texture square, moving across the scene.

The motion between any pair of frames in Fig. 4 is invisible to either the first- or second-order motion systems. To perceive the motion in Fig. 4 with low-level motion detectors would require a motion detector to correlate the square defined by feature 1 in frame 1 with a square defined by feature 2 in frame 2 (so-called interattribute motion, Ref. 38). A low-level motion system that computed interfeature or interattribute motion would be disadvantageous: First, it would be extremely complex because of the large number of *pairs* of attributes, and second, it would automatically and incorrectly compute motion between unrelated objects that happened to be near each other in successive frames. A better solution is to extract a single common attribute from all the frames and then to compute the motion of the common element by means of a similar motion algorithm as the first- and second-order motion computations.

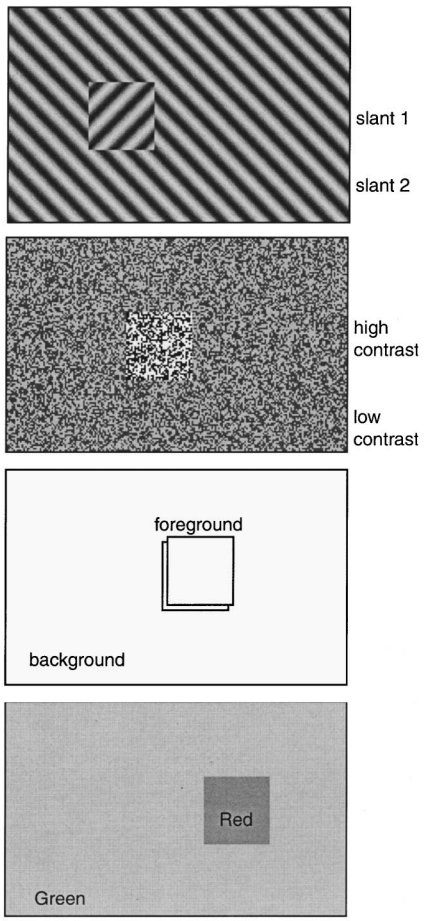


Fig. 4. Alternating-feature stimulus sequence. A square moves from left to right in successive frames. In frame 1, the square is defined by texture orientation: inside the square the texture slant is +45 deg, outside −45 deg. In frame 2, the square is defined by texture contrast: the square is composed of high-contrast texture, the background of low-contrast texture. In frame 3, the square is defined by a random-dot stereogram: the square is perceived in front of the background. In frame 4 the square is defined by isoluminant color: the square is red, the background is green. Such left-to-right motion is easily and compellingly perceived. The motion between pairs of frames is invisible to either the first- or the second-order motion system. To perceive the motion depicted here with low-level motion detectors would require a motion detector to correlate the square defined by the features in frame i with a square defined by the features in frame $i + 1$. The alternative of computing the motion of figure from frame to frame is clearly preferable.

The obvious common element linking all the squares and rectangles of Fig. 4 is figure–ground. The square in all the frames is recognized as figure, the rectangle as ground. The figure–ground computation is itself complex, but the figure–ground is highly prominent in perception, and therefore figure–ground is something that the brain computes. Lu and Sperling[40] proposed a third-order motion system that computes motion between areas that are marked as figure in successive frames. To formalize this it is useful to introduce the concept of salience.

5. *Salience*

We use the term salience to describe the assumed neural process that underlies the perception of figure–ground (and therefore third-order motion). Whereas figure–ground is intrinsically a binary variable (e.g., 1 for figure, 0 for ground), salience is a continuous variable, with larger values more likely to be perceived as figure and smaller values more likely to be perceived as ground. At each moment in time, every x, y location is assumed to have a value of salience (the salience map) in which significant features, those that are computed to be "figure" have values near 1, and areas computed to be "ground" have values near 0. Third-order motion is assumed to be computed by standard algorithms[12,14,16] from the spatiotemporal changes of the map. Computational models have been proposed for the computation of salience, generally, and for how salience in relatively simple stimuli is modified by attention.[74–80]

6. *The Genetic Argument*

An important reason for assuming that third-order motion is computed by an algorithm similar to that for first- and second-order motion is that the genetic code needed to instantiate a computation in the brain is quite complex. The likelihood that a new gene for a motion computation would evolve separately versus the original motion-computation gene being spliced from one location to another is negligibly small. According to this reasoning, insofar as third-order motion appears to have algorithmically different properties from first- or second-order motion (such as failing the pedestal test), it is more likely to be due to preprocessing of the input to the third-order system than to the third-order motion computation itself. At the moment, we know of no good psychophysical procedures that clearly discriminate third-order preprocessing from third-order motion computation itself; this issue is a good candidate for physiological investigation.

The difficulty in measuring the properties of a third-order motion system is that the same stimulus components that influence salience usually also stimulate first- and second-order motion processes. The demonstration of third-order motion requires stimuli that are invisible to first- and second-order motion processes. There are three basic paradigms: (i) In the paradigm illustrated in Fig. 4, features are defined differently in successive frames so that no single-feature motion system can compute motion. (ii) An area defined by a feature that is in-

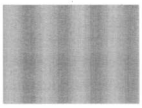
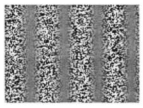
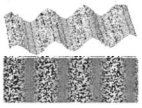
System:	I	II	III	Refs
Property				
EFFECTIVE INPUT	PHOTONS	FEATURES	SALIENCE	Lu & Sperling, 95[40]
CORNER FREQ	12 Hz	12 Hz	3-6 Hz	Lu & Sperling, 95[41]
MONOC/INTEROC	MONOC>INTER	MONOC ?	MONOC=INTER	Lu & Sperling, 95[41]
PEDESTAL IMMUNITY	YES	YES	NO	Lu & Sperling, 95[41]
ATTN INFLUENCE	NO	NO	YES	Solomon & Sperling, 94[73]
DETECTION MECHANISM	ENERGY	ENERGY	ENERGY ?	Lu & Sperling, 95[41]
REVERSED PHI	YES	YES		Nishida, 93[89] Lu & Sperling, 99[90]
MASKING/GAIN-CONTROL				
CARRIER	YES	YES		Lu & Sperling, 96[84]
MODULATOR	YES	NO		Lu & Sperling, 96[84]
SELECTIVE ADAPTATION	YES	YES		Nishida, et al, 97[52]
BRAIN LESIONS	V2-V3	~MST		Vaina, et al, 96, 98[93,96]
KDE (SFM)	YES	NO		Dosher, et al. 89[115]
COLOR SENSITIVE	No		Yes	Lu, et al, 99[58,59]
PHASE INDEPENDENCE	I & II			Lu & Sperling, 95[41]

Fig. 5. Summary of the three motion systems: their preferred stimuli, properties, and key references. Most natural stimuli excite all three systems.

visible to first- and second-order motion computations, such as a depth-defined area in a dynamic random-dot stereogram,[41] is moved. (iii) Selective attention is used to amplify a feature's salience, not the feature itself, and thereby its input to apparent motion.[40]

It should be noted that, like the first- and second-order motion systems, the third-order motion system transmits information about the location, direction, and speed of movement. Information about what is moving—the features themselves—is carried by a pattern-processing system.[32,80]

B. Major Properties of the Three Motion Systems

We summarize the major properties of the three motion systems in Fig. 5.

1. *Effective Input*

The effective input to the first-order motion system is a moving modulation of a photon distribution, detected by cones and rods in the visual system. A typical first-order stimulus is a moving sine-wave luminance modulation, though such a stimulus activates all three motion systems[41] to various degrees.

The effective input to the second-order motion system is a moving modulation of the "activity" of texture-sensitive neurons, i.e., texture grabbers. The typical second-order stimulus is a moving modulation of the contrast of a texture carrier. Such stimuli, without a stationary pedestal, also activate the third-order system.

The effective input to the third-order motion system is a moving modulation of feature salience (foreground/background). All the above stimuli can activate the third-order motion computation. Examples of stimuli that are selectively or exclusively directed to third-order motion are moving depth-defined areas in dynamic random-dot stereograms, motion-from-motion stimuli, alternating feature stimuli, isoluminant color stimuli, and stimuli in which selective attention to a feature is used to disambiguate the motion.

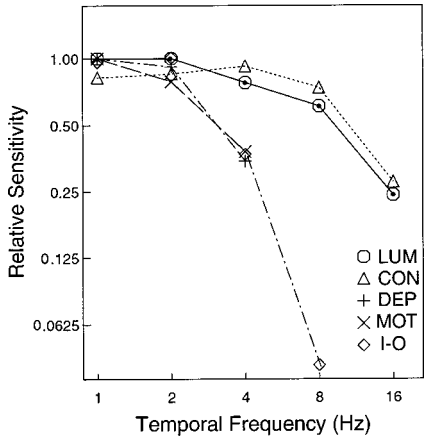


Fig. 6. Temporal-frequency-tuning functions for five types of motion stimuli, for one observer. Abscissa, temporal frequency (Hz) of a moving sinusoidal modulation; ordinate, threshold amplitudes of modulations required for 75% correct left–right motion discrimination. The axes are logarithmic. The curves have been vertically translated to expose their similarity in shape. ○ (LUM), luminance-modulation motion (first order) for pedestaled and for nonpedestaled stimuli (thresholds are identical); △ (CON), texture-contrast modulation motion (second order) for pedestaled and nonpedestaled stimuli (thresholds are identical); + (DEP), nonpedestaled stereoptic depth-modulation motion (third order); × (MOT), nonpedestaled motion-modulated motion (third order); ◇ (I-O) nonpedestaled interocular luminance-modulation motion (third order). The scale value 1.00 on the ordinate represents the following modulation amplitudes: LUM 0.0014, CON 0.027, DEP 0.40 min, MOT 0.11, I-O 0.023.

2. *Temporal Corner Frequencies*

Lu and Sperling[41] measured the temporal tuning functions for the three motion systems (Fig. 6). All the temporal frequency characteristics have typical low-pass-filter shapes; that is, sensitivity decreases monotonically with increasing temporal frequency. The temporal frequency characteristics fall naturally into two groups, fast and slow. The fast group contains the luminance modulation (first-order) and texture-contrast modulation

(second-order); the slow group consists of third-order stimuli: the dynamic stereo-depth modulation, motion-from-motion stimuli, and interocular luminance modulation with observers who have not learned how to perceive first-order interocular movement. Within each group, the shapes of the temporal tuning functions are remarkably similar. Following common engineering practice, we define corner frequency as the frequency at which the sensitivity has dropped to 1/2 of the maximum sensitivity ($-0.3 \log_{10}$). The temporal sensitivity functions for the first- and the second-order stimuli have a corner frequency of ~12 Hz; the temporal sensitivity functions for the third-order stimuli have a corner frequency of ~3 Hz.

3. Pedestal Immunity

Reichardt detectors have two mathematical properties that prove to be useful for characterizing them experimentally. 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.[14] This property is called *pseudolinearity* because it holds only for sine waves and only when they have *different* temporal frequencies. The second is that *static stimuli 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 Reichardt (and hence a motion-energy) detector in response to the moving stimulus. This property is the *pedestal immunity* of motion-energy detectors.

The original proof of these two properties of Reichardt models, pseudolinearity and ignoring static displays, applied to stimuli in continuous motion with exactly n cycles falling within the period of visual temporal integration.[14] Sampled stimuli of finite duration generally have only approximate pedestal immunity, and even that requires a carefully chosen duration.[81] Let m and n be integers, $m \geqslant 4$. For sampled stimuli with m samples per period, the integer number of samples closest to $(n + 0.25)m$ yields the best approximation to pedestal immunity. For example, for a sine wave that moves 90 deg between successive frames, a five-frame stimulus would exhibit approximate pedestal immunity when presented to a Reichardt (or motion energy) detector.

Note that pedestal immunity is a rather unusual property. Figure 7(c) shows five frames of a pedestaled luminance modulation created by linearly superimposing two components: a stationary sine grating [the pedestal, Fig. 7(a)] and a linear moving sine grating [the motion stimulus, Fig. 7(b)] with an amplitude ratio of 2 to 1. In pedestaled motion sequences [Figs. 7(c), 7(f), and 7(i)], the peaks and valleys of the compound stimuli wobble back and forth, moving first one way, then the other. Nevertheless, the output of a motion-energy detector is approximately the same for the compound pedestal-plus-motion stimulus as it is for the motion stimulus alone.[82] The question is, how do human observers perceive the compound stimulus? Do they perceive the wobble of 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. Observers viewed a computer-generated stimulus such as illustrated in Figs. 7(e) and 7(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.) The threshold amplitude for 75% correct motion-direction responses was determined by the method of constant stimuli. A pedestal with twice this measured threshold amplitude was then added to the moving stimulus to produce the pedestaled stimulus [Figs. 7(f) and 7(i)]. If motion-direction judgment were based on the output of a motion-energy detector, we would expect the observer's accuracy of left versus right judgments to be exactly the same 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 observers to judge motion direction.

We performed the pedestal test using four types of stimuli. (1) The moving luminance grating [Fig. 7(e)], which consists of alternating dark and light bars; it is the sort of first-order motion stimulus from which traditional motion psychophysics has evolved. (2) The moving texture-contrast grating [Fig. 7(h)], which is a pure second-order stimulus—a binary noise (carrier) whose texture contrast is subjected to a drifting sinusoidal modulation. (3) The depth grating. A 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. 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. 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. (4) A motion-defined motion grating, which is analogous to a moving sine-wave grating in which areas that are lighter than the surround are replaced by random dots that move a small fixed distance upward between successive frames. Areas that were darker than the surround are replaced with downward moving dots. A test stimulus is produced by drifting the up–down motion pattern horizontally in a consistent direction from frame to frame.[36,39] The ability to perceive this kind of motion-defined motion seems to suggest a hierarchical organization of motion detectors. The movement of the motion modulation (i.e., the movement of the global up–down motion pattern versus the local dot movement) is invisible to first- and second-order systems because there is no consistent global modulation of luminance or contrast.

Observers perceive completely obvious apparent motion in all the motion-stimulus-alone conditions when the modulation amplitude is sufficient. When observers first

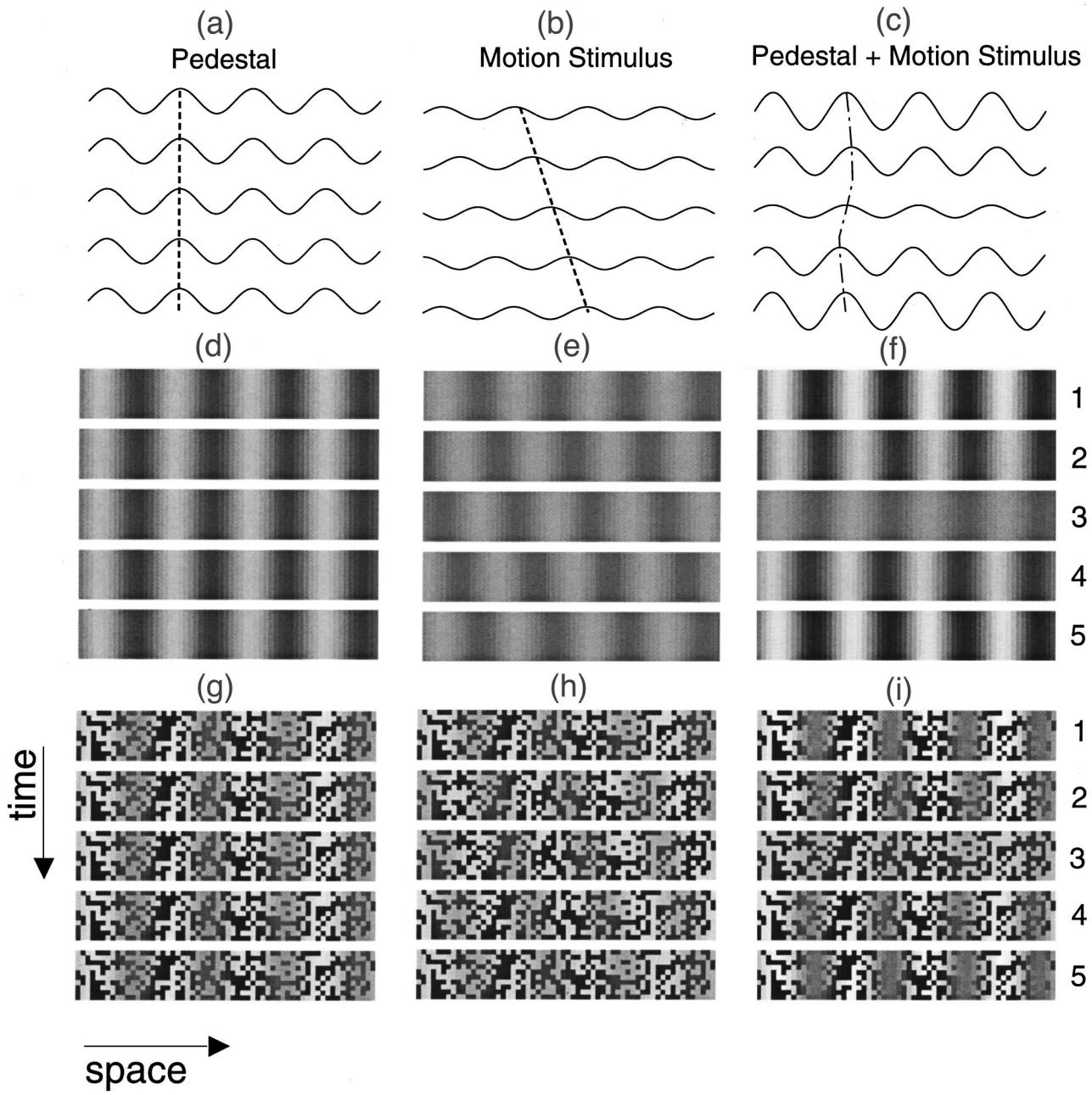


Fig. 7. Pedestal paradigm for first- and second-order stimuli. (a) Schematic representation of five frames of a stationary sine wave (the pedestal). The dashed vertical line indicates a stationary peak. (d) Five stimulus frames of the pedestal in a first-order (luminance-modulation) stimulus. The actual frames were 3.1 × 1.6 deg; only a horizontal slice is shown. Luminance varies sinusoidally as a function of space. (g) Five frames of the pedestal in a second-order texture-contrast modulation stimulus. The expected luminance is the same throughout the texture; texture contrast varies sinusoidally as a function of space. (b) Schematic representation of five frames of a rightward-moving sine wave. The slanting line indicates the rightward movement of the peak. (e) Five frames of a moving sine-wave luminance modulation (first-order motion). From top to bottom, the sinusoid traverses one period. (h) Five frames of a moving second-order texture-contrast modulation. (c) Schematic representation of a pedestal-plus-motion stimulus, summation of the modulations of (a) and (b). The pedestal has twice the amplitude of the moving sine. The dashed–dotted line indicates the peak, which wobbles back and forth 1/6 of a period. (f) A pedestaled first-order motion stimulus, the sum of the modulations of (d) and (e). (i) A pedestaled second-order stimulus, the sum of the modulations of (g) and (h). The accuracy of left–right motion discrimination for (f) equals (e) and for (i) equals (h).

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. In pedestaled stimuli, the linear motion appears like the motion of wind because it is not attached to any particular object—the only objects, the pedestal stripes, are wobbling back and forth or motionless.

Observers quickly learn to ignore the wobble and to attend to the wind. Once the observers judge the direction of the wind, the presence of a pedestal, even with twice the amplitude of the moving stimulus, has no effect what-

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

It is an axiom that to perceive something requires a corresponding brain computation. If the linear motion of the wind is perceived by the first-order and by second-order motion systems, then there must be a motion system to perceive the back-and-forth wobble—obviously, the third-order motion system. The tuning function of third-order motion is consistent with the great reduction in the apparent amplitude of wobble at 8 Hz and its virtual disappearance at 16 Hz.

These results suggest that both the first- and the second-order systems use a motion-energy algorithm, whereas the third-order system extracts motion direction by tracking the important stimulus features (peaks or valleys) over space and time.

4. Monocular Versus Binocular Computation

Consider a simple (not pedestaled) moving luminance-sine-wave stimulus with successive frames separated by 90 deg (Fig. 8). We found that converting such a stimulus from monocular to interocular presentation (e.g., Ref. 37) raises the contrast threshold (at low frequencies) by a factor of 12 (from 0.17 to 2.0%) and decreases the corner frequency from 12 to 3 Hz. The resulting tuning function superimposes exactly on those of stereo depth and of motion-defined-motion stimuli (Fig. 6). This result indicates that the observers perceived the motion of the interocular luminance grating with their third-order systems. (We presume the third-order system exhibits the same corner frequency when it detects motion in interocular luminance stimuli as it does when it detects stereomotion and motion-defined motion stimuli.)

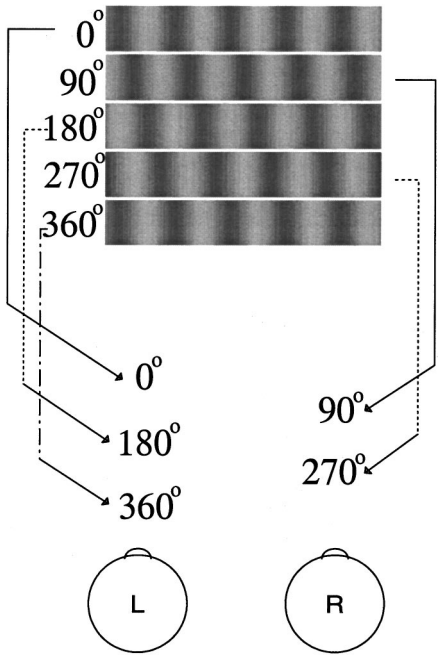


Fig. 8. Representation of an interocular stimulus presentation in which frames are alternately directed to the left (L) and right (R) eyes. Each successive stimulus has a spatial phase shift of 90 deg. Within an eye, the stimulus sequence, indicated on the bottom, is ambiguous as to direction of motion.

On the other hand, interocular presentation of the motion-defined motion stimulus with no pedestal 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 whether stimulus frames alternate between eyes or all frames are directed to the same eye. The same indifference to eye(s) of origin is found in the detection of motion of isoluminant red–green gratings, which are also perceived by the third-order motion system.

We also created interocular pedestaled luminance and texture-contrast stimuli by adding appropriate pedestals to the images in each eye and direct alternating frames into left and right eyes. Though observers perceive motion from such stimuli under normal monocular or binocular viewing conditions, they either could not perceive motion at all or had vastly reduced sensitivity in brief interocular presentations. 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 could combine information from both left and right eyes.

We[41] interpreted these results to mean that the direction of motion of pedestaled luminance and pedestaled texture-contrast stimuli is computed primarily monocularly, but there may be a weak interocular component. In general, for brief stimuli, both the first- and the second-order motion systems are primarily monocular (but see Subsection 3.E). Of course, when a (normal) binocular stimulus is presented, both the left- and the right-eye systems respond. The monocularity is apparent only with interocular stimuli. The third-order system is inherently binocular in the sense that monocular and interocular versions of the same stimulus are perceived equally well.

5. Attentional Influence

Selective attention appears to play little role in the perception of ordinary (nonpedestaled) first- and second-order motion when both types of motion occur transparently at the same location. That is, when a single patch contains two superimposed stimuli of different types, concurrent motion-direction judgments are made as accurately for both types as for just one type of motion stimulus; there is no effect of attention to one type or the other.[73]

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[40] developed ambiguous motion paradigms in which the odd frames (1, 3, 5) contained alternating stripes of white spots and of black spots. The even frames (2, 4) contained a grating in which there is strong salience modulation, e.g., alternating stripes of high-contrast and low-contrast texture (Fig. 9) or a stereogram with a depth grating (the foreground is figure; the back-

ground is ground). With neutral attention, observers fail to perceive a consistent direction of motion because without selective attention, the black spotted and white spotted stripes are equal with respect to salience. Selectively attending to the dark spots produces one direction of apparent motion, and selectively attending to the white spots produces the perception of motion in the opposite direction. Motion is perceived because the third-order motion system computes the motion of the areas of greatest salience; these are the attended spots in odd frames and the high-contrast texture in the even frames. Analogous results were obtained with selective attention to left-slanting gratings versus right-slanting gratings.

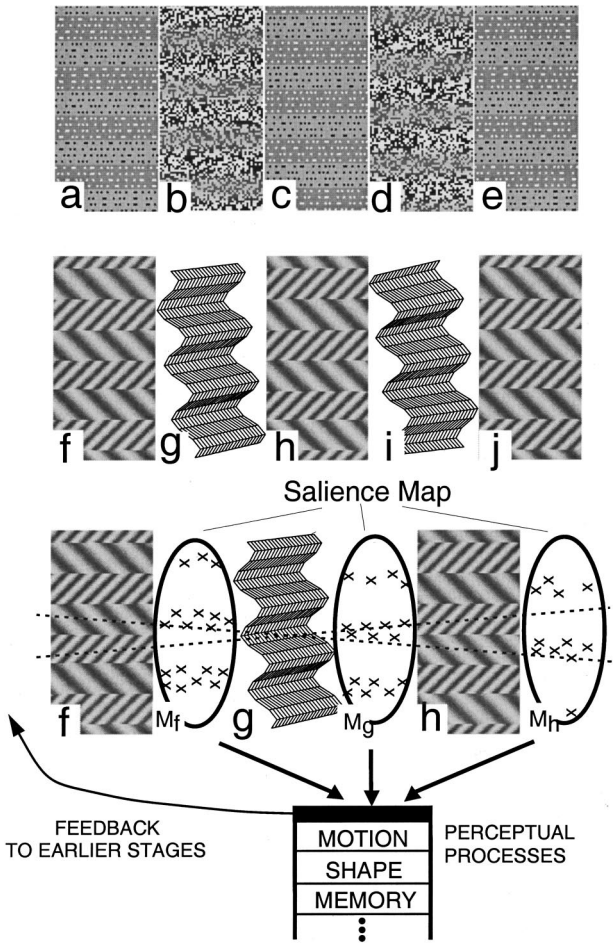


Fig. 9. Alternating-feature stimulus sequences for attention-generated motion and their relation to feature salience maps. The top row (a–e) shows an alternating-feature stimulus with frames of black-dot textures plus white-dot textures (a, c, e) alternating with frames of low- and high-contrast textures (b, d). A sequence of five consecutive frames is shown; each is displaced vertically by 90 deg from the previous one. The high-contrast stripes in b and d are perceived as figure against a low-contrast background. Selective attention to black spots produces downward apparent motion, attention to white spots, upward motion. The second row (f–j) shows a depth/texture alternating feature stimulus. 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 stimuli the near peaks are automatically the most salient. No features in the texture stimuli are automatically salient. When the observer intentionally attends to the coarse grating, the grating's 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 stimulus 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 the transfer to memory.

Blaser *et al.*[56] developed an ambiguous-motion paradigm (similar to Fig. 9), in which selective attention to red stripes produced apparent motion in one direction and selective attention to green stripes produced the opposite direction of perceived motion. They observed that selective attention to red made unsaturated red stripes as efficient as more-saturated red stripes for the purpose of determining the direction of apparent motion. They used this equivalence principle to measure the amplification of attention. Blaser *et al.*[56] determined that, for the salience–motion computation, selective attention to red or to green was equivalent to increasing the saturation of the attended color by ~30%. Only the salience of the attended color changed; its appearance was not observed to change (see Ref. 85). These and similar results[40,86] demonstrate that ambiguous-motion paradigms can provide a delicate assay of the effects of attention.

Using a dual-task paradigm, Ho[53] found that two concurrent second-order motion direction-discrimination tasks (at different locations) interfere with each other; two concurrent third-order motion direction-discrimination tasks also interfere with each other. However, while a concurrent rapid serial visual presentation (RSVP) letter-recognition task interferes with a third-order motion-direction-judgment task, it *does not* interfere with second-order motion-direction discrimination. Ho interpreted her results to mean that the pattern recognition stage of the third-order motion computation shares attentional resources with RSVP letter recognition, while second-order motion is computed in a separate, independent pathway.

6. *Reverse Phi*

When a bright bar on a gray background is flashed first in one location and then ~100 ms later in an adjacent location, the motion perception is called "phi" or phi motion.[4] When the second bar is black instead of white, under some circumstances, the motion can appear to have occurred in the reverse, instead of the forward direction.[87,88] Anstis[87] called this phenomenon "reverse phi." A similar reversal in the direction of apparent motion with the reversal of contrast had been observed by Reichardt in stimuli presented to the beetle eye,[12] and this formed the basis of his correlation model.[13] A positive correlation (white-to-white, black-to-black) indicates the physically presented forward direction. A negative correlation (white-to-black or black-to-white) is perceived in the reverse direction—opposite to what was physically presented. With long interflash delays, the reverse phi illusion fails, and motion is perceived in the forward direction.

Chubb and Sperling[11] developed a reversed-phi type of grating stimulus, GAMMA, that appeared to move in the forward direction when viewed foveally and in the reverse direction when viewed peripherally. They demonstrated that ordinary Fourier analysis actually predicted the

reverse-phi motion, and that the second-order computation (full-wave rectification) predicted the forward direction. This means that reverse phi is a physical, not a psychological phenomenon, in the sense that the physical Fourier stimulus contains the reverse-motion components.

Second-order reverse phi, first observed by Nishida,[89] is the most interesting case. To perceive first-order reverse-phi motion, the sequence is white patch to black patch on a gray background. To perceive second-order reverse-phi motion, the sequence is high-contrast texture to low-contrast texture on a medium-contrast texture background. When such a sequence of texture patches is presented in rapid succession, observers indeed see reverse-phi motion, but only in peripheral viewing. In central viewing they see forward motion.[90] Because these second-order reverse-phi stimuli have no first-order components, and because the reverse-phi direction is exactly what is predicted by second-order motion theory, this leaves the perception of forward motion unaccounted for. It is axiomatic that if something is perceived there must be a corresponding brain computation. In this case it is the third-order motion computation.

In conclusion, perceiving a second-order reverse-phi stimulus in the reverse direction is strong evidence for the full-wave rectification second-order motion computation. Perceiving forward motion in second-order reverse-phi stimuli implies the existence of a third-order motion computation.

7. Masking/Gain-Control Properties

Using pedestal amplitude as the independent variable, Lu and Sperling[84] studied the masking/gain-control properties of the first-and the second-order motion systems, that is, how these systems' responses to motion stimuli are reduced by pedestals and other masking stimuli. Motion-direction thresholds were measured for test stimuli consisting of drifting luminance and texture-contrast modulation stimuli superimposed on pedestals of various amplitudes.

First-order motion-direction thresholds are unaffected by small pedestals, but at pedestal contrasts above 1 to 2% (5× to 10× motion threshold), motion thresholds increase proportionally to pedestal amplitude—a Weber law (Fig. 10). For first-order stimuli, pedestal masking is specific to the spatial frequency of the test (see e.g., Ref. 91). On the other hand, motion-direction thresholds for texture-contrast stimuli are independent of pedestal amplitude, indicating that there is no gain control whatever. The accessible pedestal amplitude range is from 0 to 40%, which is 10× unpedestaled threshold. In first-order motion, thresholds start increasing when the pedestal reaches 5× the unpedestaled threshold. It appears that the biological limitation (no gain control for second-order pedestals) reflects the physical limitation on pedestal size. On the other hand, 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. 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.

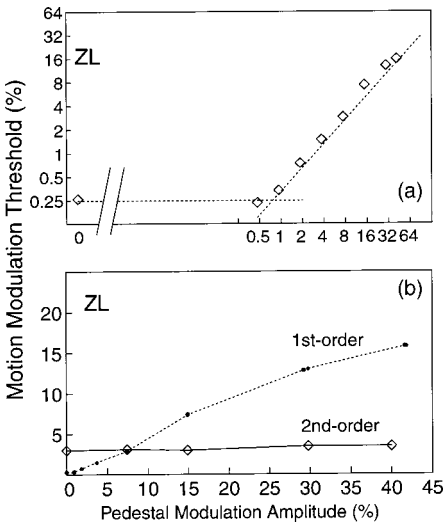


Fig. 10. Masking of sine-wave motion by sine-wave pedestals as a function of pedestal amplitude. (a) First-order stimuli. Contrast-modulation thresholds for 75% correct motion-direction judgments versus pedestal-modulation amplitudes. Both axes are logarithmic. For small pedestal amplitudes (≤1%, ~4× unpedestaled motion threshold), pedestaled thresholds for drifting luminance modulation do not depend on pedestal amplitude. For large pedestal amplitudes, motion threshold for drifting luminance modulation is approximately proportional to the amplitude of the pedestals (a Weber law), as indicated by the dotted line. (b) Second-order stimuli. Texture-contrast modulation thresholds for 75% correct motion-direction judgments versus pedestal-modulation amplitudes. Linear axes are used here to better display the wide range of pedestal amplitudes throughout which motion threshold for a drifting texture-contrast grating is a constant. Data from (a) for first-order motion and pedestal are shown for comparison.

The different gain-control properties of the two motion systems and prior observations of motion masking and motion saturation are encompassed in the flow chart of Fig. 11. The stimulus inputs to both first- and second-order motion processes 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.

In summary, the gain of the first-order system is set by both the carrier (retina light adaptation) and the modulator (pedestal amplitude). The gain of the second-order system is set by light adaptation and the amount of carrier (overall texture contrast) but not by the modulator (pedestal amplitude).

8. Selective Adaptation

A priori, it is probable that each motion system can be separately adapted, including each eye individually of the first- and second-order systems.[92] Moreover, different test stimuli measure adaptation differently. Few adap-

tation experiments admit unambiguous conclusions in the face of these complexities. Nishida *et al.*[52] used selective adaptation to determine whether first- and second-order motion pathways consist of multiple motion-detecting channels that are each narrowly tuned to a different spatial scale. They measured motion-direction-discrimination thresholds for drifting luminance (first-order) and texture-contrast modulation both before and after motion adaptation. The drift direction, spatial frequency and first-order or second-order stimulus type of the adaptation and test stimuli were independently manipulated. When the adaptation and test stimuli were either both first-order gratings or both second-order gratings, robust elevations of direction-identification thresholds were found, and showed both direction selectivity and spatial-frequency selectivity. Cross-over adaptation effects between first-order and second-order gratings were sometimes observed but were weak and not spatial-frequency selective. These findings give direct support for the existence of multiple-scale processing for first-order and second-order motion in the human visual system and provide additional evidence that first- and second-order motion are initially processed by independent pathways.

Lu *et al.*[92] also conducted selective adaptation experiments to investigate the interactions between the 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. Stimuli to the left eye and right eye alternate once per second for 10 s. To measure the magnitude of the motion after effect (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 third-order system, interocular moving sine-wave gratings are employed. Dynamic random-noise stimuli with different proportions of "signal" dots moving in particular directions are used as test stimuli to measure the magnitude of the third-order

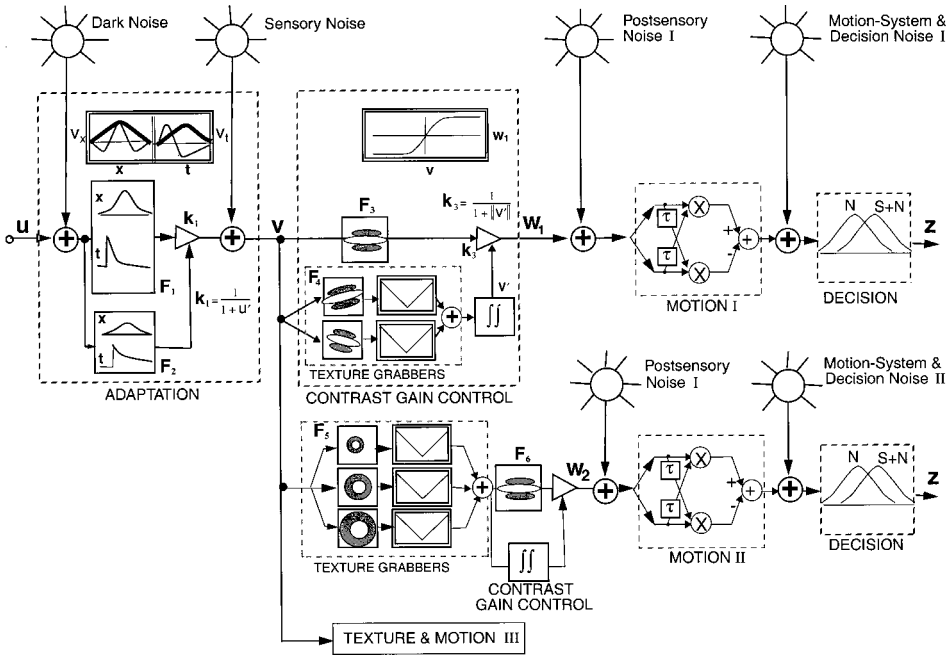


Fig. 11. Model of early visual processing. Suns, stationary-noise sources; rectangular boxes, linear filters; double rectangles enclose graphs of input–output relations. Triangles indicate gain-controlled amplifiers: the input (controlled signal) is the horizontal path; and the gain-controlling path is vertical. *Adaptation*. The visual input $\mathbf{u}$ 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 μ' of the through-signal. Amplifier gain is feedforward controlled by filters F_2, which have greater spatial and greater 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 thick graphic plots in the double boxes. At high luminances, the receptive fields are bandpass, 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 ($\int\int$) and cover a broader range of frequencies than the through-signal filter F_3. The double boxes with a "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-controlling signal is summed over the surrounding area ($\int\int$). The motion computation itself is similar to first-order (Motion I), which is represented here as a Reichardt detector. 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).

MAE. When viewing these stimuli, observers were not aware of which eye or whether both eyes were being stimulated.

All three types of adaptation stimuli produced 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 a MAE only for first-order test stimuli but not for second-order stimuli. Adapting to a second-order stimulus produced a MAE only for the second-order stimuli; there was no significant cross adaptation.

We had previously shown that our observers detected the motion of interocular stimuli with their third-order motion systems. Nevertheless, interocular stimuli produced strong (third-order) MAEs. These three modes of highly selective adaptation indicate three functionally distinct motion computations carried out by at least five clusters of neurons. Neurons selectively sensitive to signals originating in the left eye and in the right eye individually contribute to first- and to second-order motion computations. Furthermore, there is a binocular site to which signals from both eyes contribute equally that adapts to third-order stimuli.

9. Phase-Independence Tests

When two stimuli that move in opposite directions are added, the motion each would have produced individually can be canceled at many different levels of processing. However, when two stimuli both move in the same direction, if there is some relative phase that produces motion cancellation, these stimuli must combine before the motion computation. The combination of two separate motion computations, both in the same direction, cannot be less than either alone [Fig. 12(a)]. This is the underlying basis of the phase-independence test for the independence of processing systems.

To investigate phase independence, 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 just-noticeable differences above threshold. 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 equaled or exceeded 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 before 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-deg phase difference (i.e., opposite sign) would perfectly cancel each other [Fig. 12(d)]. If two stimuli sum before motion is computed [single-channel theory, Fig. 12(b)], we expect some phases that are better than the stronger motion stimulus and other phases that are worse. (1) The absence of any phase dependence[41] combined with (2) 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.

10. Brain Lesions

Recent identification and studies of human patients with highly selective impairment of first-order motion[93] or second-order motion[51,94–96] due to distinctive focal lesions of separate brain regions also lend strong support for an independent first-order and at least one other motion system. In the lesion studies published to date, the authors

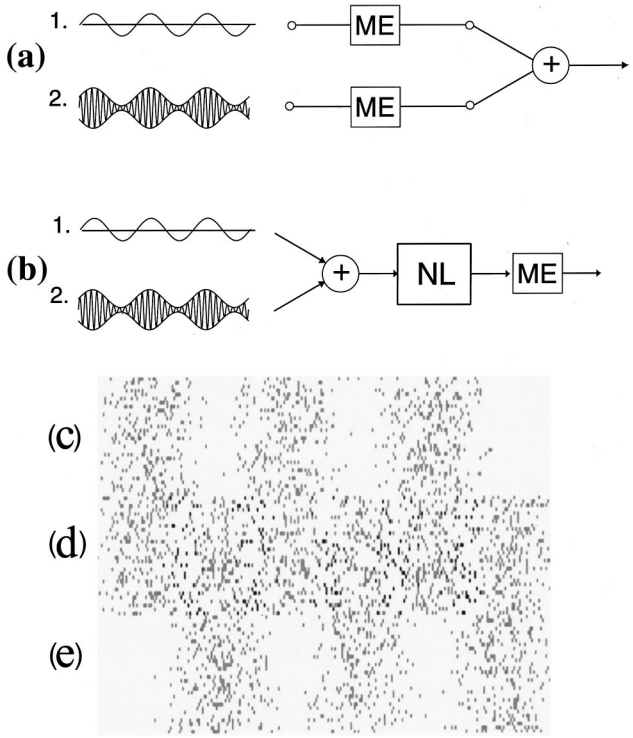


Fig. 12. Phase dependence: stimuli, models, and a demonstration. The first- and second-order input signals represented at the left are physically superimposed and move at the same speed and direction. ME, motion-energy computation for extracting motion direction. (a) A two-mechanism model: First- and second-order motions are computed in separate channels; summation occurs after the two independent motion computations. Motion strength is independent of the relative phases of the two stimuli. (b) A single-channel model: first- and second-order stimuli are combined before motion computation. NL represents a possible pointwise nonlinear transformation on the combined signal. (c), (d), (e). Demonstration of phase-dependence in a single-channel theory: Two signals, (c) and (e), move together 180 out of phase. (d) is the sum of (c) + (e). Because (c) and (e) are 180 deg out of phase they cancel each other; there is no modulation in the sum (e) and therefore no stimulus for motion. Two in-phase stimuli (not shown) would reinforce each other.

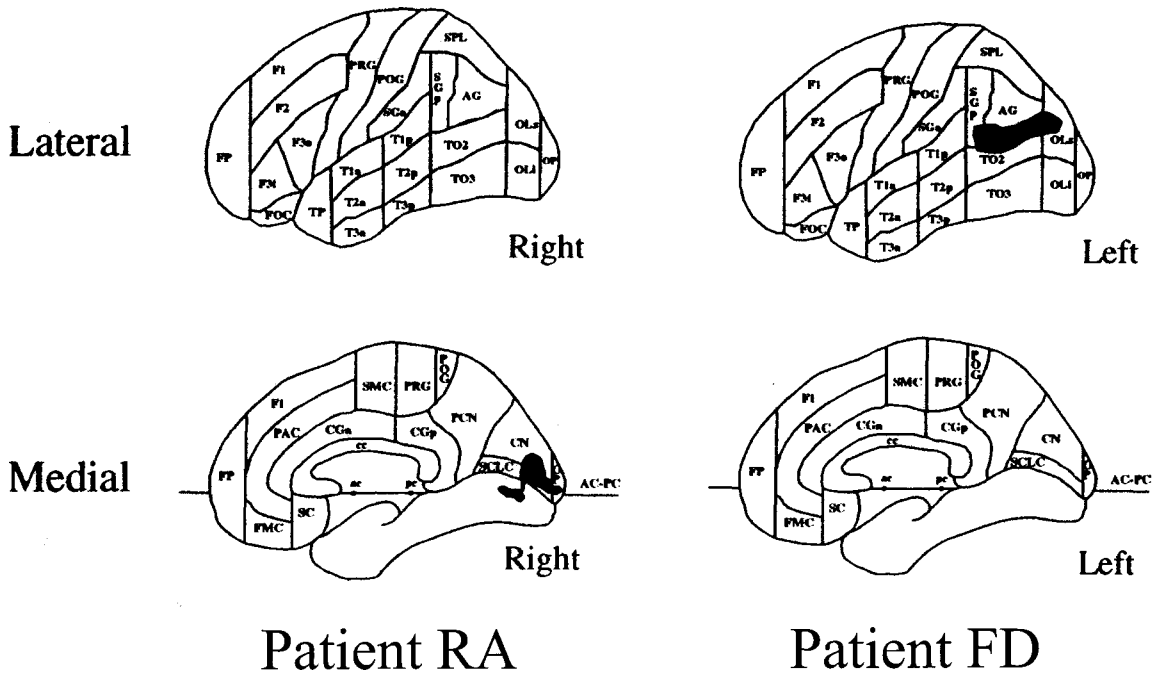


Fig. 13. Schematic anatomic localization of brain lesions in two patients. Left panels (patient RA): An infarct caused a first-order motion deficit in the left hemifield with sparing of second and third-order motion perception. Shown in medial view, it involves cortical parcellation units[97] CALC, SCLC, CN, LG, and OP in the right hemisphere. Right panels (patient FD)[96]: An infarct produced a right hemifield deficit of second and/or third-order motion with sparing of first-order motion perception. Shown in lateral view, it involves cortical parcellation units SGp, AG, and TO2 in the left hemisphere. (From Vaina *et al.*,[98] Fig. 7, p. 75, with permission of Wiley–Liss, Inc.)

have not distinguished between second- and third-order motion systems, nor is a distinction possible on the basis of the available material.

Vaina *et al.*[93] studied a patient RA who had unilateral brain damage centered on putative visual areas V2 and V3 in the medial part of the occipital lobe (Fig. 13). He was seriously impaired in performing first-order motion tasks in the visual field contralateral to the site of the injury. However, he performed second-order (or third-order) visual tasks in the impaired area of the visual field as well as or perhaps even better than normals.

Another patient, FD, with a small unilateral cortical lesion adjacent (dorsal) to human cortical area MT, suffered an apparently permanent disorder in perceiving several forms of second-order (or third-order) motion in the hemifield contralateral to the lesion. But his first-order motion in the same area remained normal.[95] The highly selective nature of the disorders suggests separate neural pathways for first- and second- and /or third-order motion.

While we[65] and others have had moderate success in obtaining different EEG and MEG responses to first- and second-order visual motion stimuli, the brain localization of motion systems by these measures or by fMRI has remained elusive (see Ref. 99 for a review).

3. CRITIQUES AND RESOLUTIONS

A. A Single Motion System with an Early Nonlinearity Can Compute Both First- and Second-Order Motion

Taub, Victor, and Conte[67] proposed that, while there exists a separate "long-range" motion mechanism (presumably related to the third-order motion system), a single "short-range" motion mechanism could account for some critical results of first- and second-order motion perception experiments. Their short-range motion model (referred to here as the TVC model) could also account for the results from a new set of ingenious experiments that used complex motion stimuli.

Here, we reapply the TVC model to the data considered by Taub *et al.*[67] and also consider a wider range of data. We show that, even though the TVC model can fit experimental data quite accurately within a small domain, it fails dramatically and in predictable ways when it is required to use the same parameters to predict data from different domains. Encompassing the range of phenomena the TVC model claims to account for would require at least two independent systems of visual computations, i.e., systems that we previously have designated as first- and second-order motion systems. The new experiments of Taub *et al.*[67] are described first. Then the TVC model is described and its predictions are compared with existing data.

1. The TVC Experiments

Taub, Victor, and Conte[67] produced a set of complex motion stimuli using the principles illustrated in Fig. 14: (I) A first-order stimulus, a vertical sine wave, $f_1(x - t, y) = a + b \sin(x - t)$, (Fig. 15a). The spatial average (E_y, the expected value over y (average over the y dimension) of $f_1(x - t, y)$, $E_y(f_1(x - t, y))$, in this case is simply $f_1(x - t, \cdot)$—a moving (in the x dimension) sine-wave modulation of luminance. One can also compute E_y of $f_1(x - t, y)$ after pointwise transformation by a power function ($\| \cdot \|^p$), $E_y(f_1^p(x - t, y))$. When $p = 1$, $E_y(f_1(x - t, y))$ is a sine-wave modulation in the x dimension

with the same spatial frequency as the original sine wave. When $p = 2$, $E_y(f_1^2(x - t, y))$ is sine function of x with double the spatial frequency (Fig. 14). When $p = 3$, $E_y(f_1^3(x - t, y))$ is a different periodic function of x with the original spatial frequency. When $p = 4$, $E_y(f_1^4(x - t, y))$ is yet another periodic function of x with double spatial frequency. Thus, stimulus $f_1(x - t, y)$ can "survive" (preserved as a nontrivial function of x) pointwise power transformations $\| \cdot \|^p$ when $p = 1$, 2, 3, and 4 (see column 1, Fig. 14). For the case of a vertical sine wave, $\| \cdot \|^p$ with $p = 1$, taking the expectation over y, E_y, leaves matters unchanged.

(II) A TVC second-order stimulus, $f_2(x - t, y)$ was created by sampling from two out-of-phase sine-wave modulations (e.g., Ref. 100), $L_0(1 + m \sin(x - t))$ and $L_0(1 + m \sin(x - t + \pi))$, with equal probability (Fig. 15b). Figure 14 illustrates that $f_2(x - t, y)$ survives (with doubled spatial frequency) pointwise transformations $\| \cdot \|^p$ followed by spatial averaging in the y dimension when $p = 2$, and 4 but not when p is odd.

(III) A TVC third-order (notice the different meaning of "third-order" here) stimulus, $f_3(x - t, y)$, is created by sampling from three 120-deg out-of-phase sine-wave modulations, $L_0(1 + m \sin(fx))$, $L_0(1 + m \sin(fx + \frac{2}{3}\pi))$, and $L_0(1 + m \sin(fx + \frac{4}{3}\pi))$, with equal probability (Fig. 15c). It is apparent that for $p = 1, \ldots 4$, $f_3(x - t, y)$ "survives" the pointwise transformation $\| \cdot \|^p$ followed by spatial averaging in the y dimension only when $p = 3$. In this case, $f_3^3(x - t, y)$ has triple the original spatial frequency.

(IV) A TVC fourth-order stimulus, $f_4(x - t, y)$ is created by sampling from four 90-deg out-of-phase sine-wave modulations, $L_0(1 + m \sin(fx))$, $L_0(1 + m \sin(fx + \frac{1}{2}\pi))$, $L_0(1 + m \sin(fx + \pi))$, and $L_0(1 + m \sin(fx + \frac{3}{2}\pi))$, with equal probability (Fig. 15d). Here, $f_4(x, y)$ "survives" the pointwise transformation $\| \cdot \|^p$ followed by spatial averaging in the y dimension only when $p = 4$. In this case, $f_4^4(x, y)$ has quadruple the original spatial frequency.

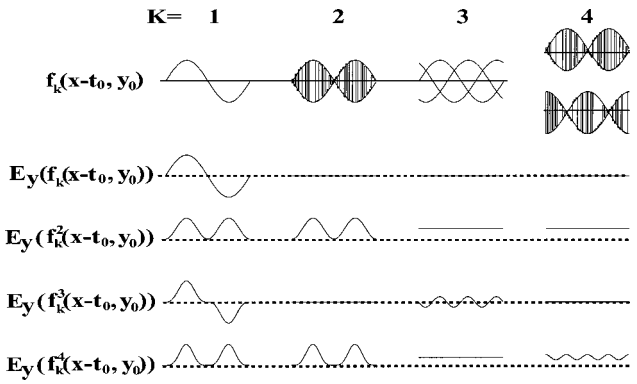


Fig. 14. Schematic stimulus generation principles of Taub *et al.*[67] Stimuli are $f_K(x - t, y) = \underset{k=1,K}{\Lambda} [a + b \sin(x - t + \theta_k)]$ where $K = 1, \ldots, 4$ and $\theta_k = (k - 1)2\pi/K$. Λ designates a random sample at each point (x, y, t) from one of the K sine waves. E_y is the expectation over y of $f_K^p(x - t, y)$. For $t = t_0$, and for particular values of p and K, E_y is a function of x that is shown in row p column K. (See text for details).

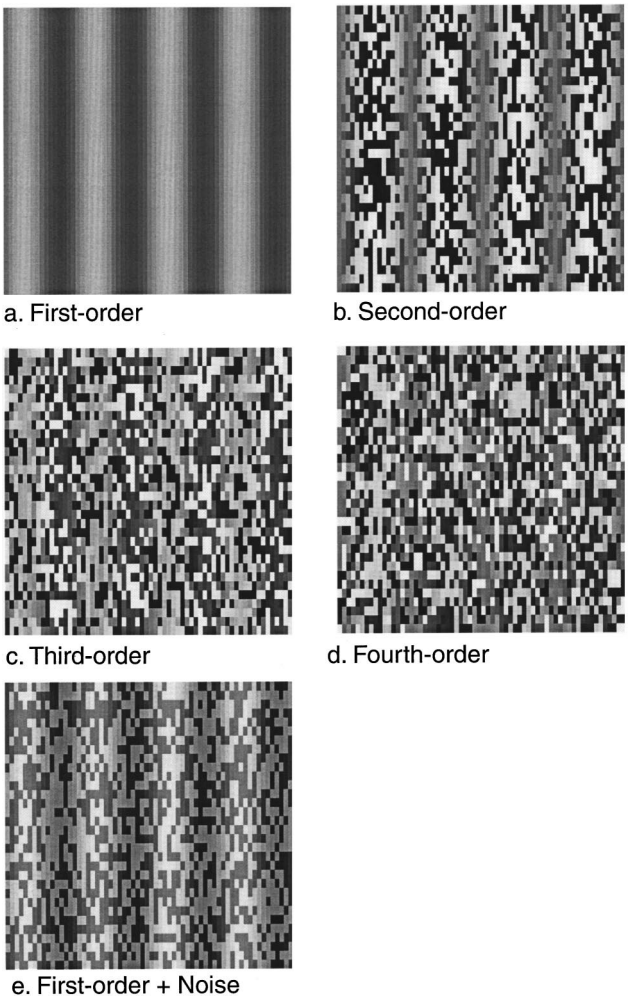


Fig. 15. Single frames similar to stimuli of Taub *et al.*[67] a, TVC first-order stimulus; b, TVC second-order stimulus; c, TVC "third-order" stimulus; d, TVC fourth-order stimulus; e, TVC first-order stimulus with a noise carrier. The TVC terminology to describe stimuli is quite different from the terminology used by Lu and Sperling [41] to describe motion systems. The TVC first-order stimulus is primarily a stimulus to the first-order motion system. The other TVC stimuli stimulate both second- and third-order motion systems to a degree that depends on the rectifying nonlinearity of the texture grabbers in the second-order motion system and on the figure–ground process in the third-order motion system.

To generate motion stimuli from $f_K(x, y)$ illustrated in Fig. 15, each component sine-wave modulation was moved at exactly the same speed and in the same direction. Because each frame involved new random samples, there was no frame-to-frame correlation and hence no first-order motion stimulus for TVC-stimuli with $K = 2, 3, 4$.[101]

2. *The TVC Model*

The TVC model consists of two consecutive stages: an initial pointwise compressive nonlinearity followed by a Reichardt-type covariance motion computation (Fig. 16). Specifically, the pointwise compressive nonlinearity T is a half-wave rectifier (versus the full-wave rectifier in the Chubb–Sperling[30] second-order motion model). The pointwise nonlinearity acts directly on the point contrast (defined below) in the stimulus.

Let $l(x, y, t)$ be the luminance of a point x, y at time t, and l_0 be the mean luminance of a stimulus. Then the point contrast $P(x, y, t)$ at point x, y at time t is

$$P(x, y, t) = \frac{l(x, y, t) - l_0}{l_0}. \quad (1)$$

The compressive nonlinearity T is defined as

$$T[P(x, y, t)] = \begin{cases} P(x, y, t)^{\alpha} & \text{if } P(x, y, t) > 0 \\ 0 & \text{otherwise} \end{cases}. \quad (2)$$

The nonlinearly transformed input $T[P(x, y, t)]$ is submitted to a modified Reichardt-type motion covariance computation[12–15] to derive the net directional motion strength MS_{TVC} along the x axis.[102] Specifically, motion strength in the TVC model is defined as

$$\begin{aligned} MS_{TVC}[P(x, y, t)] = \sum_x \sum_y \sum_t \{ & T[P(x, y, t)]T[P(x + \Delta x, y, t + \Delta t)] \\ & - T[P(x, y, t)]T[P(x - \Delta x, y, t + \Delta t)]\} \end{aligned} \quad (3)$$

where Δt is the duration of each frame of the stimulus and Δx is the distance the image moves from one frame to the next.[103]

3. TVC Predictions and Data: Ratio of First- to Second-Order Thresholds

We apply the TVC motion model to derive threshold amplitude ratio between first- and second-order motion. The logic is simple. Suppose a single motion system, i.e., the TVC short-range motion model, extracts motion from both first-order and second-order motion stimuli. Then, at threshold, the computed motion strength should be exactly the same for first- and second-order motion stimuli provided that these stimuli have the same spatial and temporal frequency, spatial layout, and duration.

In our simulation, we concentrated on a particular pair of first-order and second-order stimuli, one in which, after initial nonlinear preprocessing in the TVC motion model [Eq. (2)], will have the same fundamental spatiotemporal frequency (Ref. 41, Figs. 10a and 10b, p. 2713). The first-order stimulus is a drifting sine grating with spatial frequency ω_x, temporal frequency ω_t, and initial phase θ_1. Its contrast C is defined by

$$C_1(x, y, t, \theta_1) = m_1 \cos\{2[2\pi(\omega_x x - 2\omega_t t)] + \theta_1\}. \quad (4)$$

The second-order stimulus is a drifting contrast modulation of static binary spatial noise. It is defined by

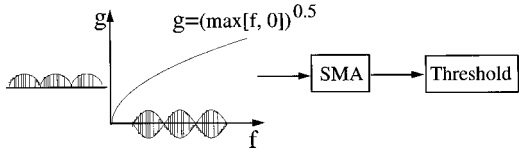


Fig. 16. Taub *et al.*[67] single-channel model: The input to the motion system first goes through a pointwise half-wave compressive rectification ($g = f^{0.5}$). Motion is computed from the output of the half-wave rectifier by a standard motion analyzer, SMA, which is similar to a Reichardt detector.

$$C_2(x, y, t, \theta_2) = R(x, y)m_2 \cos[2\pi(\omega_x x - \omega_t t) + \theta_2], \quad (5)$$

where θ_2 is the initial phase and $R(x, y)$ is a random variable that assumes values $+1$ and -1 with equal probability (Fig. 17).

Set the spatial period ω_x^{-1} to 32 pixels, and the temporal period ω_t^{-1} to 16 frames and the frame duration 9 ms. The simulation shows that in order for the first- and the second-order stimuli to produce the same motion strengths, $m_2/m_1 = 12.4$. In Lu and Sperling,[41] Experiment 4, Figs. 10a and 10b, the measured luminance-modulation motion threshold is 0.0042 for observer ZL and 0.008 for observer EB; and the measured contrast-modulation motion threshold is 0.024 for observer ZL and 0.029 for observer EB. At threshold, the measured m_2/m_1 is 5.6 for observer ZL and 3.6 for observer EB. These measured threshold ratios differ by factors of 2.2 and 3.4 from the predicted value. The TVC theory fails to predict the measured threshold ratio between first- and second-order stimuli that are quite similar to the first- and second-order stimuli in the TVC experiment. Our attempts to correct the problem by adding linear spatial averaging before the nonlinear preprocessing stage in the TVC model made the problem worse, not better.

4. TVC Predictions and Data: Phase-Independent Summation of First- and Second-Order Motion

Lu and Sperling[41] developed a relative phase-dependence test to examine whether first- and second-order motion is computed in the same or separate motion systems. If two stimuli are processed by the same motion system, then there must be a way to add them while they are moving in the same direction so that there is a phase θ in which the individual motion strengths cancel each other (in the perception of apparent motion) and another phase $\theta + \pi$ in which the individual motion strength add. Specifically, when superimposed, equal-strength luminance and texture-contrast modulations of the same temporal and spatial frequencies [described by Eqs. (4) and (5) above] with different relative phases were moved in the same direction, Lu and Sperling[41] found that there was no phase dependence in observers' performance in motion-direction-judgment experiments. Moreover, the probability of correct motion-direction judgments in the combined stimulus followed probability summation of each component stimulus alone (as would be expected from two independent motion computations). If motion from first- and second-order stimuli had been computed in one single system, strong phase dependence would have resulted. The observed phase-independent summation of first- and second-order motion led to the conclusion that motion from first-order luminance- and second-order texture-contrast modulations is computed in separate motion systems (Ref. 41, p. 2714).

Taub *et al.*[67] simulated the results of the the Lu–Sperling phase-dependent test using the TVC model. Their simulations showed that "for a stimulus consisting of a superimposed Fourier and nonFourier grating in which the envelope spatiotemporal frequency of the non-Fourier grating matches the spatiotemporal frequency of the Fourier grating, motion energy varies by 16% (aver-

age fractional deviation) with relative spatial phase (corresponding to Ref. 41, Figs. 10a, b)". They argued that the phase-dependence in terms of motion strength (MS_{TVC}) is "surprisingly small" (p. 1474). Because Lu and Sperling[41] measured fraction of correct motion-direction judgments (not motion energy), the phase-dependence would be further compressed. Taub *et al.*[67] thus asserted that the observed phase-dependence in observers' performance (ranging from 85% to 88.3% for observer ZL, and 93.3% to 95% for observer EB) can be accounted for by their model.

Before going into the details of our simulation of phase effects in the TVC model, we need to clarify the problem. The TVC single-channel model produced small threshold variation with phase, similar to the almost null phase effect in the Lu–Sperling[41] data. However, Taub *et al.*[67] overlooked the fact that, unlike the TVC model, for all phases the Lu–Sperling data show a large improvement in observers' performance (Ref. 41, Figs. 10a and 10b). For observer ZL, fraction correct for first-order alone was 0.72 and second-order alone was 0.70; for observer EB, fraction correct for first-order alone was 0.77 and second-order alone was 0.80. However, the range of performance for the combined stimulus over eight different phases was 0.85 to 0.88 for observer ZL and 0.93 to 0.95 for observer EB. There was not a single instance in which observers' combined-component performance was not significantly greater than that of either single component alone. A single-channel system would show improved direction discrimination in half the range of phases of a joint stimulus and impaired discrimination in the other half of the range of phases.

The second issue to be clarified concerns the small variation in performance with a combined first- and second-order stimulus as the phase between them is varied. A small variation in a single model could mean that one of the stimuli is not effective, i.e., that the stimuli have not been equated for motion-energy output. When the first- and second-order stimuli are both set to produce equal outputs (i.e., both are at threshold), we expect the TVC model to produce the dramatic threshold variation that is characteristic of all single-channel models.

To simulate the relative phase-dependence experiment within the framework of the TVC single motion system theory, we have to first create luminance modulation (first-order) and contrast modulation (second-order) stimuli that are of equal strength (in terms of MS_{TVC}). In the TVC model, for the stimulus pair described in Eqs. (4) and (5), we have found that equal first- and second-order stimuli require the ratio $m_2/m_1 = 12.40$. The computation of the predicted fraction of correct response in motion-direction-judgment tasks requires a psychometric function to convert motion strength MS_{TVC} to fraction correct. Figure 17 illustrates our simulation procedure (described in detail below) and the results.

1. Create a luminance-modulation (first-order) stimulus $C_1(x, y, t)$ according to Eq. (4). Set $m_1 = 0.0042$, the threshold for observer ZL.[41]
2. Implement nonlinear preprocessing. Use Eq. (2) to compute $T[P(x, y, t)]$.
3. Compute motion strength $MS_{1,TVC}$ of $T[P(x, y, t)]$ using Eq. (3).
4. Create a contrast modulation (second-order) stimulus $C_1(x, y, t)$ according to Eq. (5). Preliminary explorations show that a contrast modulation m_2 such that $m_2 = 12.4m_1$ will produce a motion strength $M_{2,TVC} = M_{1,TVC}$.
5. Implement nonlinear preprocessing. Use Eq. (2) to compute $T[P(x, y, t)]$.
6. Compute the motion strength of $T[P(x, y, t)]$ using Eq. (3).

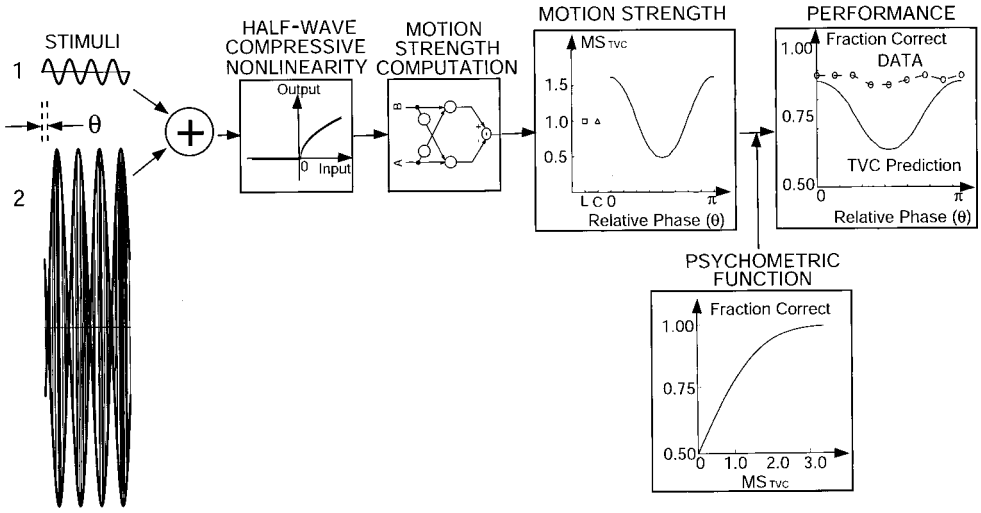


Fig. 17. Phase dependence in the TVC single-channel model: stimuli, model, and predictions. A drifting luminance sine wave and a drifting texture-contrast modulation are produced with an amplitude ratio 1:12.4 (in order to produce equal outputs after the compressive nonlinearity). The inputs are summed with relative phase θ, then are linearly summed (+). The summed signal is transformed by a half-wave compressive pointwise nonlinearity (see Fig. 16), after which the TVC motion energy extraction algorithm computes motion strength MS_{TVC}. MS_{TVC} is plotted against θ, $0 \leq \theta \leq \pi$. MS_{TVC} for each stimulus component, individually, is also shown (square, triangle). A psychometric function (from Ref. 41) converts motion strength MS_{TVC} to fraction correct in the motion-direction-judgment task. TVC predictions are compared with experimental data from Lu and Sperling (Ref. 41, Fig. 10a, observer ZL). The TVC single-motion-system theory obviously fails to capture these data.

7. Repeat steps (4, 5, 6) 256 times with new random stimuli. Find the average motion strength $M_{2,TVC}$ for the second-order stimulus, and verify that indeed $M_{2,TVC} = M_{1,TVC}$.

8. Create a large number of combined stimuli consisting of the sum of (1) the luminance-modulation (first-order) stimulus and (4) the contrast-modulation (second-order) stimulus with different relative phases, θ, $0 \leqslant \theta \leqslant 2$, where $\theta = \theta_2 - \theta_1$ [defined in Eqs. (4) and (5)].

9. Implement nonlinear preprocessing. Use Eq. (2) to compute $T[P(x, y, t)]$ for each combined stimulus.

10. Compute motion strength of the combined stimulus for each value θ using Eq. (3).

11. Repeat steps (8, 9, 10) 256 times to find average motion strength $M_{1+2,TVC}(\theta)$ for the combined stimuli.

12. To convert motion strength $M_{1+2,TVC}(\theta)$ into the probability of a correct motion-direction judgment assume that zero-mean Gaussian noise is added to motion strength. A net value greater than 0 indicates the positive motion direction; a value less than zero indicates the opposite direction. The standard deviation σ of the Gaussian distribution is defined by the requirements that a net motion strength of zero ($MS_{TVC} = 0$) produces 50% correct motion-direction judgments and that $M_{1,TVC} = M_{2,TVC}$ produces 75% correct motion-direction judgments.

The simulation results are shown in Fig. 17. For the single-component stimuli, $MS_{1,TVC}$ and $MS_{2,TVC}$ individually produce 75% correct responses. $MS_{1+2,TVC}(\theta)$ for the combined stimulus varies systematically; it achieves its minimum when $\theta = \pi/2$, its maximum when $\theta = 0, \pi$. Converting MS_{TVC} to fraction correct in motion-direction-judgment tasks, this motion-strength range corresponds to 62 to 86% correct, with 50% being chance level. The average data for observer ZL from Lu and Sperling[41] in the precisely comparable condition are shown as the nearly horizontal line at 86%.

It is apparent that the TVC single-system motion theory, as would any single-system theory, makes dramatically incorrect predictions for phase-dependence in two-component experiments. TVC obtained different results in their simulations because, instead of matching the strength of the first- and the second-order stimuli at the output level, i.e., after the nonlinear pointwise transformation and motion-strength computation, they matched the strength (amplitudes) of the first- and the second-order stimuli at the input level. When the first- and the second-order stimuli were of equal motion strength at threshold, their amplitude ratio was more than an order of magnitude apart. The small phase dependence predicted in their simulation was due to the far-below-threshold strength of the second-order stimulus. To reiterate our main point: No matter what detailed assumptions one may make, the signature of a single-channel model is that performance on a two-component stimulus is worse than that of the better component individually in some relative phases and better in others. To the contrary, the actual data show phase independence and a uniform improvement.

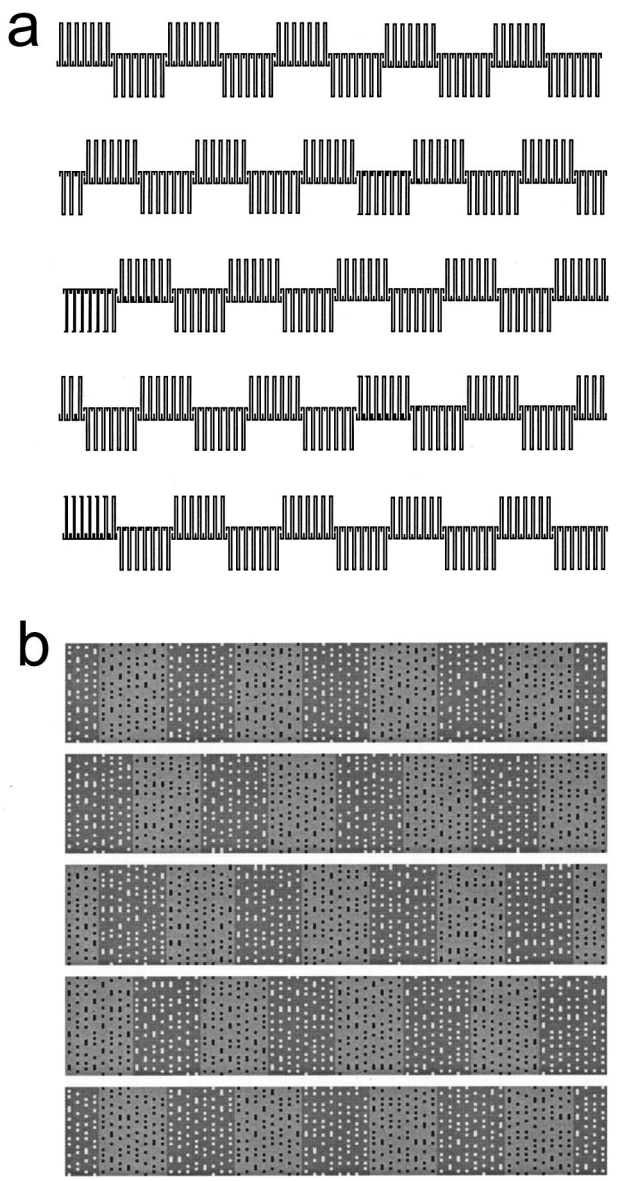


Fig. 18. Displays to stimulate motion systems that use half-wave rectification. Half-wave stimuli are composed of "Mexican hats:" A micropattern consists of 3×3 pixels in which the middle pixel has contrast c and the other eight pixels have a contrast $-c/8$, so the mean contrast of the hat micropattern is 0. c is positive for white hats and negative for black hats. a, Schematic cross sections, from top to bottom, of five frames of a half-wave stimulus moving rightward. b, The actual images (partial views). Intensity compression in early vision distorts intensity relations, producing first-order and second-order contaminations. When these are compensated (removed), the motion of the half-wave stimulus can be detected only by the third-order motion system, indicating that second-order motion uses full-wave, not half-wave, rectification.[73]

5. TVC Predictions and Data: Relative Efficiency of Half-Wave and Full-Wave Motion

Solomon and Sperling[73] measured threshold amplitude in motion-direction judgment using full-wave and half-wave stimuli (Fig. 18). They found that observers were much more efficient in extracting motion from full-wave than from half-wave stimuli. Only one third of their observers (total number of observers = 18) could reliably perceive motion from the half-wave motion stimuli under optimal

conditions. All their observers easily perceived motion from the full-wave stimuli. This is a gross contradiction to the TVC half-wave model, which predicts an advantage of the half-wave stimuli over the full-wave stimuli.

6. Full-Wave Versus Half-Wave Rectification

Chubb and Sperling[11] studied motion in a Gaussian windowed reverse-phi stimulus, which they called Γ. Γ has the interesting property that the "forward" direction of motion, which is automatically and easily perceived by every observer viewing Γ in foveal vision, would be invisible to a system that performed half-wave rectification. However, full-wave rectification exposes the forward (second-order) direction of motion very efficiently. Therefore Chubb and Sperling concluded that the rectification in second-order motion probably was full-wave, and certainly not exclusively half-wave.

The full-wave versus half-wave question was taken up again by Solomon and Sperling,[73] who produced stimuli composed of Mexican hat micropatterns which were designed to produce apparent motion only in visual systems that contained full-wave rectifiers or only in half-wave systems. They called these full-wave and half-wave motion stimuli. The TVC model assumes half-wave rectification. Because threshold data for all observers and both full-wave and half-wave conditions are available (Ref. 73, Table 1), we compute the responses of the TVC model (MS_{TVC} at threshold) in both full-wave and half-wave conditions, and compare them with the actual data for each observer. Table 1 shows the ratio of half-wave versus full-wave MS_{TVC} for each observer.

If the TVC single motion system theory were correct, the MS_{TVC} half-wave/full-wave threshold ratio would have been 0.5 for all observers because, after the positive half-wave rectification of TVC model, the Solomon–Sperling full-wave stimuli contain only 1/2 as many positive micropatterns as the half-wave stimuli. It is apparent that, at threshold, MS_{TVC} for half-wave stimulus is at least 3.6 times that of the full-wave stimulus. The half-wave stimuli were extremely inefficient in Solomon and Sperling's[73] experiments, though half-wave stimuli should have been twice the strength of full-wave stimuli if TVC's theory held.

In summary, the failure to predict the phase independence of motion-direction judgments of combined same-direction first-order and second-order motion stimuli falsifies the single-channel assumption of the TVC model. The gross failure of the half-wave/full-wave threshold predictions falsifies the half-wave rectification assumption of the TVC model.

B. A Single Multichannel-Gradient Model for First- and Second-Order Motion

Johnston and associates[20,104–106] proposed multi-channel gradient models that could account for all first-order phenomena they investigated and for many second-order motion phenomena. They did not attempt to deal with third-order motion phenomena. Although a gradient model is inherently a velocity model, it was adapted to make predictions of motion strength. The computation of gradients involves only linear processes (i.e., differentiation). The computation of velocity from the ratio of two gradients involves a nonlinear process (division, in this case of a temporal by a spatial gradient). The computation of velocity or motion strength from the combinations of velocities computed in different spatial frequency bands and from different points in space and time can be quite complex. We have not been able to fully understand the nonlinear processes proposed by Johnston *et al.* nor how these processes accomplish motion detection. A description of the basic gradient computations underlying their model is offered here for completeness.

A rigidly moving pattern $I(x, y, t)$ that at a moment in time t moves with velocity u in the x dimension and with velocity v in the y dimension satisfies the following:

$$l(x, y, t) = I(x + u\,dt, y + v\,dt, t + dt). \tag{6}$$

The full derivative of $I(x, y, t)$ with respect to t is zero:

$$\frac{dI(x, y, t)}{dt} = u\frac{\partial I(x, y, t)}{\partial x} + v\frac{\partial I(x, y, t)}{\partial y} + \frac{\partial I(x, y, t)}{\partial t} = 0. \tag{7}$$

In an infinitesimal neighborhood around a point x, y, t, the surface $I(x, y, t)$ can be regarded as a plane. Equation (7) describes infinitesimal movements of a planar surface. Consider any point x, y in the image $I(x, y, t)$ at time t. A moment later that point has moved an amount in x that is proportional to the first right-hand term, an amount in y that is proportional to the second right-hand term, during an interval t proportional to the third term. Since it is the same point at both times, there is no

Table 1. Responses of the Selected Observers of Solomon and Sperling[a] and of the TVC Model[b] to Full-Wave and Half-Wave Stimuli

Observer	Full-Wave Threshold[c]	Half-Wave Threshold	Data (hw/fw)	MS_{TVC} (hw/fw)[d]	
				Observed	Predicted
PS	0.1225	0.185	1.7	3.6	0.5
JS	0.0683	0.121	1.8	4.3	0.5
AH	0.0603	0.12	2.0	4.7	0.5
AKH	0.0361	0.187	5.2	12.0	0.5
RK	0.0524	0.357	6.8	16.2	0.5

[a] Ref. 73.
[b] Ref. 67.
[c] Contrast amplitude of center point of Mexican hat micropatterns.
[d] MS_{TVC} is the ratio of motion strengths of the TVC model corresponding to the indicated stimuli.

change in intensity; therefore dI/dt is zero. That Eq. (7) holds for every point on an object is the standard definition of rigid motion.

Equation (7) is formulated for a two-dimensional surface.[106] However, it is simpler to explain the theory with one-dimensional stimuli (stimuli that do not change in the y direction), and we shall do so here (as is done in Ref. 20).

For a one-dimensional stimulus [in which the middle right-hand term in Eq. (7) is omitted], we designate the velocity as u_0, and it is given simply by

$$u_0 = -\frac{\partial I(x, y, t)/\partial t}{\partial I(x, y, t)/\partial x}. \tag{8}$$

Suppose every image in a space–time cube is subjected to the same linear spatial transformation, for example, differentiation with respect to x. If the original object translates rigidly, then the image-transformed object also translates rigidly. That is, changing how the stripes are painted on a zebra does not change the translatory motion of the zebra. Johnston *et al.*[20] differentiate the original images with respect to x K times. They define vectors x_k, t_k, $k = 0, ..., K$,

$$x_k = \frac{\partial^{k+1} I(x, y, t)}{\partial x^{k+1}}, \qquad t_k = \frac{\partial^{k+1} I(x, y, t)}{\partial x^k \partial t}. \tag{9}$$

Again, if the original image translates rigidly, each of these k-differentiated images follows precisely the same trajectory. The velocity u_k of a k-differentiated image is given by substituting it, instead of the original image, into Eq. (8). At certain image points, it may happen that the velocity computation is degenerate (zero over zero) for the original image or for some of the k transformed images but is well defined for other transformations. This is a potential advantage of considering transformed images, not just the original image in computing velocity.

If measurement were perfect, all nondegenerate estimates of velocity u_k, $k = 0, ..., K$, would be identical. In actual practice, measurement is not perfect, sampling is not perfect, so even in the case of perfectly rigid motion, the computed velocity values may differ. To combine the different estimates of velocity $u_k(x)$ at a point x, Johnston *et al.*[20] first make a graph of the numerator versus the denominator of Eq. (8) for the original and each of the k-transformed images. For any one image, therefore, velocity is given by the slope of the line from the origin to the plotted point. When there are several estimates, Johnston *et al.* compute the slope of the best-fitting line (in a least-squares sense) through all the points and use that slope as the estimated velocity.

Finding the slope of the best-fitting line is an elementary statistical procedure. Applied to these data, the slope is given by

$$u = -\frac{\sum \dfrac{\partial^{k+1} I(x, y, t)}{\partial x^{k+1}} \dfrac{\partial^{k+1} I(x, y, t)}{\partial x^k \partial t}}{\sum \dfrac{\partial^{k+1} I(x, y, t)}{\partial x^{k+1}} \dfrac{\partial^{k+1} I(x, y, t)}{\partial x^{k+1}}} = -\frac{\sum_{k=0}^{K} x_k t_k}{\sum_{k=0}^{K} x_k^2} \tag{10}$$

The denominator is a normalization factor that depends only on the contrast energy in the original and the differential images. The numerator is proportional to the sum of the individual slopes. In other words, the slope estimate is just the average of the individual slope estimates. Although, this is an appropriate procedure for first-order motion, it is unclear why it might work for second-order motion.

In conclusion, gradient models can recover the velocity component of first-order motion quite well. This is not surprising because a long-standing physical theory indicates that the gradient-based computation is an appropriate velocity-extraction algorithm. Problems arise when gradient models appear to recover second-order motion, because no formal theory has yet developed to explain how they might do so.

C. A Single "Activity"-Based System for Both First- and Second-Order Motion

Grzywacz *et al.*[68] proposed a temporal coherence theory to account for first-order motion integration over space and time. The authors claimed that activity at $(x1, t1)$ followed by activity at $(x2, t2)$ could induce motion from $x1$ to $x2$. Because their model could potentially compute activity from certain second-order stimuli, their model could explain second-order motion results without a need for two separate motion systems.

The crux of all motion models is the definition of the input to motion detectors—the "activity." For Grzywacz *et al.*,[68] "activity" appears to be the output of Reichardt (or similar) motion detectors. Insofar as these first-stage motion detectors function similarly to the texture grabbers in the Chubb–Sperling[30] theory, they effectively provide a contrast-modulation input to what is then a second-order motion mechanism. However, insofar as there is only one motion mechanism, the Grzywacz *et al.*[68] theory will fail the phase-independence test just as badly as the TVC model.[67] This theory, like the gradient and the initial-stage nonlinearity theories considered above, does not detect third-order motion.

D. Position Displacement, Not Velocity, Is the Cue to Second-Order Motion Detection

Seiffert and Cavanagh[69] report a qualitative difference between first-order and what they call second-order motion. They studied back-and-forth oscillatory motion in a variety of stimuli. For first-order motion stimuli, the minimum displacement threshold is inversely proportional to the temporal frequency of back-and-forth oscillation (between 0.1 and 1 Hz). They therefore conclude that first-order motion threshold is determined by velocity. For their various second-order motion stimuli, displacement threshold is approximately constant between 0.1 and 4 Hz (Fig. 19). (However, in one experiment where observers are required to discriminate direction—as well as presence—of motion, there is a marked decrease in sensitivity to motion at 4 Hz, the highest back-and-forth frequency tested). They conclude that second-order motion threshold is determined by displacement only. The overall conclusion is that first-order and second-order motion involve different mechanisms: Velocity is the cue for first-order motion; position displace-

ment is the cue to second-order motion. We have no argument with their data except that their label "second-order" is inappropriate and that they studied too small a range of conditions to define a mechanism.

There are several reasons for believing Seiffert and Cavanagh's[69] stimuli were stimulating primarily a third-order system. One of their so-called second-order stimuli was a stereo-defined grating, and it had threshold characteristics similar to their other second-order stimuli. In the review above, the temporal tuning function for stereo-motion was shown to have a sharp decline in sensitivity at 4 Hz, which was typical of third-order motion (but not first- or second-order).

Other of Seiffert and Cavanagh's[69] second-order stimuli were contrast modulated textures. These, indeed, are stimuli capable of stimulating the second-order motion system although, like most above-threshold first- and second-order stimuli, they are also clearly seen by the third-order system. Pedestaled contrast-modulated motion (Fig. 7i) is easily seen. It is perceived as a moving wind above a wobbling pedestal. Perception of the wobble motion of the pedestal could be position based (it is a third-order motion computation); the motion of the wind—the second-order motion—obviously is not position based. Our conclusion is that second-order motion (the wind) is not position based, that third-order motion (which is what we presume Seiffert and Cavanagh[69] studied) might depend on perceived position but not necessarily according to the invariance rule proposed by them.

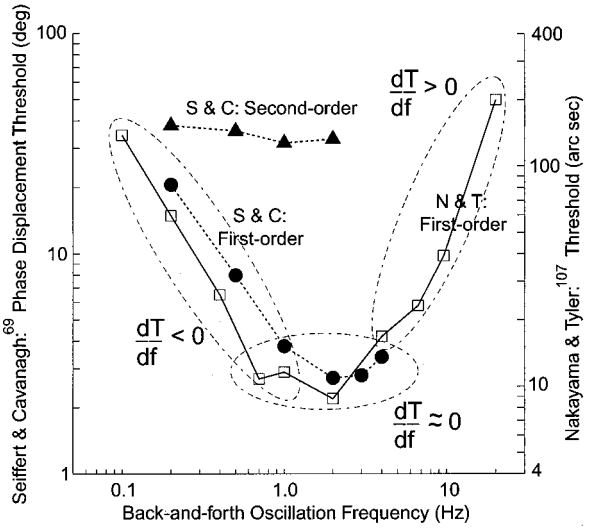


Fig. 19. Summary of the results of Seiffert and Cavanagh[69] and of Nakayama and Tyler (Ref. 104, Experiment 1, Fig. 3) determinations of displacement thresholds for back-and-forth oscillating stimuli. Seiffert and Cavanagh[69] interpret data with $dT/df \approx -1$ as indicating a velocity mechanism and data with $dT/df \approx 0$ as indicating a position mechanism. They conclude that first-order motion involves velocity sensing and second-order motion involves position sensing. Applying their criteria to their data, supplemented by Nakayama and Tyler's data we would have to conclude that first-order motion is served by three mechanisms, a different one in each of the circled ranges. An alternative interpretation is that the shape of threshold-versus-frequency functions is determined primarily by temporal filtering before the motion computation itself. (After Ref. 69, with permission of Elsevier Science Ltd.)

There is a problem with inferring the basis of a motion computation from observations of a limited frequency range of back-and-forth oscillation. For example, as measured by Seiffert and Cavanagh[69] and by Nakayama and Tyler,[107] the minimum displacement threshold for first-order motion is also a constant between about 1 and 4 Hz. In a more extensive study, Nakayama and Tyler[107] find that the minimum displacement threshold of first-order back-and-forth oscillatory motion is a U-shaped function of temporal frequency (Fig. 19). Following the logic above, one would have to conclude that first-order motion is based on a velocity mechanism below 1 Hz (where minimum displacement is inversely related to temporal frequency), is based on a position mechanism between 1 and 4 Hz (where minimum displacement versus frequency is flat), and is based on yet a third mechanism above 4 Hz (where minimum displacement increases with temporal frequency).

The data of Seiffert and Cavanagh[69] can be explained more parsimoniously when we recognize that their procedure does not differentiate spatiotemporal filtering before motion detection from frequency selectivity in the motion detector itself or from the nature of the motion algorithm. It is well known that the temporal tuning function for first-order motion has a bandpass characteristic.[14,108,109] This would tend to produce a U-shaped curve for displacement threshold versus frequency. The temporal tuning function for second-order motion is low pass[41,80] in the range in which it has been studied. This would tend to produce a flat curve for displacement threshold versus frequency and lead to a false conclusion that second-order motion perception was based on a displacement mechanism when, in fact, second-order motion perception is based on a motion-energy type of algorithm (see below).

There are experimental data that are more informative about motion algorithms than temporal tuning functions: For example, (1) pedestal immunity and (2) the fact that in sandwich stimuli, motion strength depends on the product of amplitudes in even and odd frames. Such data suggest that first- and second-order motion systems each utilize a Reichardt or an equivalent motion energy algorithm. The observed qualitative differences between first- and second-order motion systems reported by Seiffert and Cavanagh[69] may reflect the different temporal dynamics of processing before the two systems or perhaps the dynamic properties of the motion detectors themselves. But such data, by themselves, do not constrain the algorithms embodied in the motion detectors.

E. An Interocular Stimulus That Passes the Pedestal Test

Lu and Sperling[41] concluded that first-order motion is primarily monocular because luminance sine-wave gratings to either or both eyes pass the pedestal test,[41] have a temporal corner frequency of 10–12 Hz,[41] and exhibit reverse phi.[8] (Passing a pedestal test means that sine-wave motion direction is perceived equally with or without a stationary sine-wave pedestal of the same spatial frequency and 2× amplitude.) Sine waves presented alternately to the two eyes so that interocular combination is necessary to perceive motion (Fig. 8) fail the pedestal test,[41] have a corner frequency of ~4 Hz,[41] fail reverse phi,[8,90] and have

at least 10× lower contrast sensitivity and spatial resolution.[41] Therefore Lu and Sperling[41] concluded that interocular motion perception is mediated by the third-order, not the first-order system.

1. Carney's Interocular Pedestal Stimulus

Carney[70] found that he and his other observers could perceive the direction of interocular motion at temporal frequencies of 30 Hz. This is far beyond what would be possible for the third-order system. He also discovered an interocular sine-wave stimulus that passed the pedestal test. He concluded, naturally, that there was a first-order binocular motion-sensing mechanism. We show here how the apparently contradictory observations of Lu and Sperling[41] and Carney[70] can be reconciled. Selective attention and adaptation are the clues.

Why were Lu and Sperling's[41] observers unable to perceive interocular motion above 8 Hz? At 8 Hz, the interocularly physically moving sine wave seems to stand still or to wobble, and its overall motion direction is almost impossible to discern. This is a manifestation of the perceptual phenomenon of motion standstill, which occurs when the third-order motion system fails, but the form–color–texture systems are still able to deliver a reasonably accurate representation of the moving stimulus. (The task of the form–texture–color systems is to extract a perceptually stable object shapes from images that jitter widely over the retina; motion systems provide the perception of movement. Motion standstill does not occur with sine-wave stimuli in first- or second-order motion because these motion systems are as sensitive as or more sensitive than shape–color–texture systems.)

In fact, Lu and Sperling's[41] observers accurately reported the output of their third-order motion system. What they failed to observe then, but have now learned to observe with selective attention, is that, in addition to the wobbly sine wave there is a wind blowing over the grating that represents motion direction quite accurately. This perceptual wind is the output of an interocular first-order motion system. By selectively attending to it, observers can judge the motion direction of even extremely brief stimuli at temporal frequencies up to 30 Hz.[110,111] The interocular first-order system is less sensitive than the monocular system by a factor of 2–3 in five-frame (1.25-cycle) stimuli. In 2-s stimuli, interocular contrast sensitivity for motion direction approaches monocular sensitivity.

On the other hand, interocular motion cannot pass the pedestal test. In brief sine-wave motion stimuli, no matter how discriminable the direction may be, adding a 2× pedestal reduces interocular direction discrimination to chance. To perceive interocular motion in the presence of a 2× pedestal requires quite long durations. Interocular motion-direction discrimination in the presence of pedestals improves continuously from ~0.5 s to 3 s. That such a long exposure is required to perceive pedestaled motion suggests that the pedestal is being filtered out by adaptation processes before the motion computation.

In agreement with Carney, Sperling *et al.*[110] conclude that there indeed is an interocular first-order motion computation. It is less sensitive than monocular first-order motion. But it is reasonable to call it first-order because it is very sensitive to small contrasts (0.5%) and resolves extremely high temporal frequencies (30 Hz). Unlike monocular first-order (and second-order) motion computations, interocular first-order motion fails the pedestal test. Why this should be so is unresolved.

F. A Low-Level Special Mechanism for Stereomotion?

Stereomotion is motion that is produced in such a way that a stereoscopic depth mechanism is required to decode the sensory input, and no artifactual motion signals are available to other motion detection systems. The classical example is a dynamic random-dot stereogram (DRDS) that defines a stereo-depth grating. The depth-defined grating translates in successive frames. As there is no correlation between successive frames, there is no motion signal for first- or second-order motion. Stereomotion, defined by a DRDS, follows the typical third-order-motion temporal tuning function (falloff corner frequency is 4 Hz).

It is usually assumed that the only motion signal in DRDS motion is the depth signal. However, there often is a concomitant motion signal produced by interocular rivalry, which can be of equal strength to the true stereomotion signal.[112] Rivalry motion is a common confound in stereomotion experiments. However, as rivalry motion appears to have the same temporal dynamics as stereomotion, probably it too is perceived by the third-order motion system.

In a minireview of stereomotion sensing, Patterson[42] concludes that "stereoscopic motion is processed by a motion-sensing system composed of special-purpose mechanisms that function like low-level motion sensors" (p. 3329). He contrasts this to Lu and Sperling[41] who conclude that the third-order motion system mediates the perception of stereomotion. What is the evidence?

Patterson[42] considers motion in the Z direction (toward and away from the observer) as well as motion in the $X-Y$ direction, perpendicular to the line of sight. Only movement in the $X-Y$ plane is considered here. Patterson's arguments for a stereo-specific motion system hinge on (1) motion adaptation, (2) a difference between speed and position discrimination, (3) the assertion that failure of stereomotion to pass a pedestal test is not conclusive evidence against a specialized stereomotion mechanism.

In principle, motion adaptation paradigms could reveal mechanisms for perceiving stereomotion. For example, if there were a low-level stereomotion mechanism, adapting it should not affect nonstereo third-order motion stimuli. On the other hand, if third-order motion were the mechanism for perceiving stereomotion, then stereomotion should cross adapt with other third-order motion stimuli. The problem is that there undoubtedly is a final common path for all motion mechanisms. So cross adaptation between motion systems is the rule, and finding cross adaptation (as Patterson does) is uninformative. The hard problem is avoiding cross adaptation and adapting only one specific system. To learn about stereomotion mechanisms, it would be useful to cancel the common adaptation and thereby to expose stereo-specific or third-order-specific motion adaptation. One must also demonstrate that adaptation is of stereomotion versus adaptation of a more general stereo-depth mechanism that incidentally

weakens the stereomotion response. These adaptation experiments are possible in principle, but none of the experiments cited by Patterson[42] grapples successfully with these complexities. All admit alternative interpretations.

With respect to speed discrimination versus position discrimination, Patterson[42] advances no general principle that could used to discriminate third-order from a stereospecific motion mechanism. That arguments based on temporal-frequency-tuning functions are inconclusive with respect to identifying motion algorithms (versus motion systems) has been pointed out above in the reanalysis of Seiffert and Cavanagh.[69]

Patterson's final argument is based on an analogy to interocular motion perception. Pedestaled interocular motion perception failed in short-duration stimuli but apparently succeeded in long-duration stimuli. Patterson[42] writes that Carney[70] concludes "cyclopean motion was processed by a binocular motion-energy system without feature tracking" (p. 3340). Continuing by analogy, Patterson reasons "We should not infer that a cyclopean motion system does not exist on the basis of a lack of evidence for its existence" (p. 3340).

Some indirect evidence for third order as the only mechanism for perceiving DRDS stereo motion is the phenomenon of motion standstill—the perceptual standstill of an objectively rapidly moving stimulus. Originally, Julesz and Payne[35] observed that a stereo-defined bar oscillating back and forth at oscillation rates of ~7–9 Hz appeared to stand still. Tseng *et al.*[113] observed motion standstill in DRDS stereomotion even with continuous linear motion in a consistent direction (versus back-and-forth) for temporal frequencies of 5–8 Hz. The significance of motion standstill is that a stimulus is moving too quickly for motion mechanisms to resolve but slowly enough for texture-shape mechanisms to resolve clearly.[59] Until now, only the third-order motion system has been slow enough to fail before shape–color–texture mechanisms and thereby to make motion standstill possible.

In conclusion, the typical third-order temporal tuning function, the ease with which motion standstill can be produced, and the lack of evidence to the contrary, suggest that stereomotion is perceived only by the third-order motion system.

G. Apparent Failure of Pedestal Immunity for Brief Stimuli

A motion pedestal stimulus consists of a standing sine grating (the pedestal) on which a moving grating of the same spatial frequency is superimposed. Lu and Sperling[41] observed that motion direction thresholds for sine gratings are unaffected by pedestals of twice the amplitude of the moving sine (pedestal immunity). This observation of pedestal immunity is important because (1) a pedestal with an amplitude of 2× that of the moving grating is amply sufficient to foil any motion algorithm that

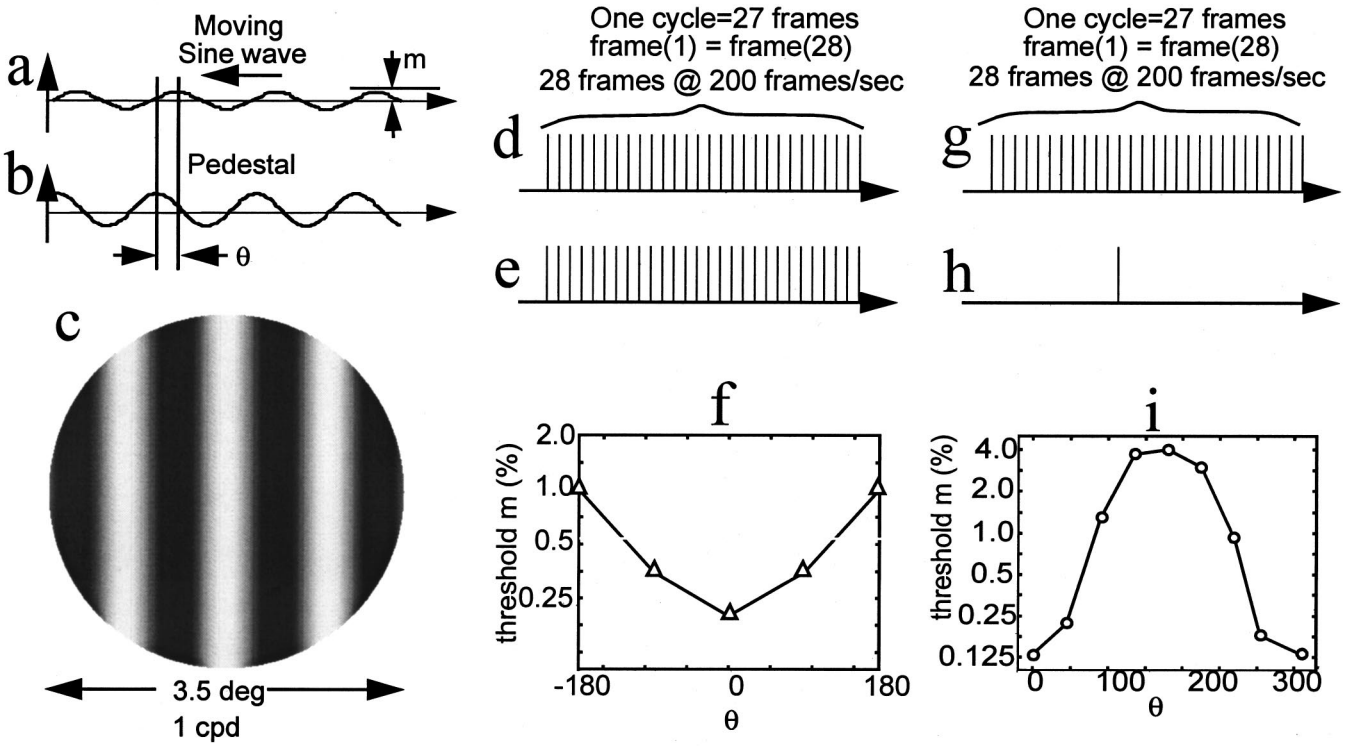


Fig. 20. Stimuli and results from the pedestaled motion experiments of Zemany *et al.*[71] They measured contrast threshold for discriminating left-versus-right motion of a vertical, 1-c/deg luminance sine-wave grating a, superimposed on a 1 c/deg stationary luminance grating (b, the pedestal) with twice motion-threshold amplitude. a and b schematically illustrate the phase θ between the moving sine wave and the stationary pedestal in frame 1. d, e, f, Equal-duration motion stimulus and pedestal. Both the motion stimulus and the pedestal contain one full cycle (27 frames) of the motion stimulus plus one extra frame. The graph f shows the systematic dependence of the motion-direction threshold on the initial phase θ. g, h, i One-cycle motion stimulus, impulse pedestal. The pedestal was presented for only one frame. The motion stimulus contained 28 frames (one full cycle+one frame). i, Systematic dependence of the motion-direction threshold on the phase θ.

depends on feature tracking (Fig. 7) and because (2) pedestal immunity is a nonintuitive prediction of a Reichardt (and hence motion-energy) detector.

Zemany *et al.*[71] reported apparent counterexamples to pedestal immunity. They measured contrast threshold for discriminating left versus right motion of a vertical, 1-c/deg luminance sine-wave grating superimposed on a 1-c/deg stationary luminance grating (pedestal) with twice-threshold amplitude, lasting one cycle of the motion stimulus. They[71] observed large phase-dependent effects of the stationary pedestal on motion direction discrimination at 7 and 15 Hz; i.e., the motion direction discrimination threshold depended systematically on the relative (initial) phase θ of the pedestal and the moving sine wave (Figs. 20a–20f). They also observed strong phase dependence in motion-direction-discrimination thresholds with very brief one-frame ("flash") pedestals (Figs. 20g–20i).

Zemany *et al.*'s[71] experimental results pose two challenges to Lu and Sperling.[41] (1) In Ref. 41, pedestal immunity of first-order motion was observed in experiments in which different relative phases of the pedestal and the moving sine wave were mixed and averaged. Their apparent pedestal immunity might have resulted from averaging across random phases. (2) Pedestal immunity is a mathematical property of the Reichardt (or equivalent motion-energy) detectors. Gross violation of pedestal immunity would suggest that computational theories of human motion perception (such as the Lu–Sperling three-motion-systems theory) based on Reichardt (or equivalent motion energy) detectors are wrong.

Reanalysis of the original pedestal data of Lu and Sperling[41] in terms of the relative initial phase between the moving luminance sine wave and the pedestal found no systematic phase dependence whatever for any of the five temporal frequencies of motion stimuli that were tested (0.94, 1.88, 3.75, 7.50, and 15.0 Hz). How is it possible to account for pedestal immunity and phase independence in these data[41] and the gross violation of pedestal immunity and the big phase dependence in the data of Zemany *et al.*?[71]

Examination of the stimuli used by Lu and Sperling[41] and by Zemany *et al.*[71] suggests the major, critical difference in the two experiments: In the experiments of Lu and Sperling,[41] the pedestaled stimuli were composed of five frames with a 90-deg phase shift between successive frames. The period of the motion sequence was four frames, and the fifth frame was a repeat of the first. The total phase shift was 450 deg. In the experiments of Zemany *et al.*[71] the pedestaled stimuli were composed of 28 frames in a motion sequence with a period of 27 frames (a total phase shift of 373 deg in their 7.43-Hz condition).[114] In this section we show that the Reichardt (or equivalent motion-energy) detector actually predicts the two disparate data sets that result from these seemingly minor differences in stimulus durations.

This section is organized in three parts. (1) We first show that, for finite and/or sampled stimuli, Reichardt detectors do not have perfect pedestal immunity, and we outline the conditions under which a high degree of immunity obtains. This can best be understood in terms of the Fourier analysis of the stimuli. (2) We illustrate a simple algebraic approximation that can predict both the basic pedestal immunity observed by Lu and Sperling[41] and the apparently contradictory results of Zemany *et al.*[71] (3) We make some new predictions—for example, that prolonging the Zemany *et al.* stimulus to $90 + n \times 360$ deg, where n is a positive integer, would make it relatively immune to pedestals under a wide range of conditions.

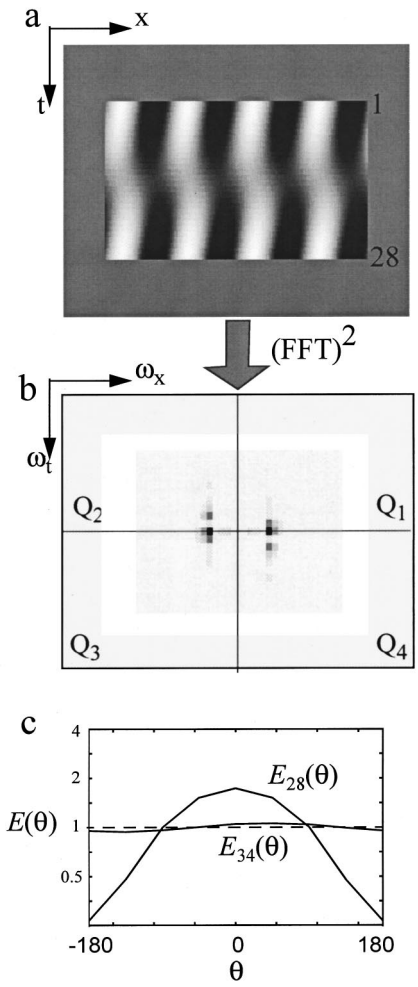


Fig. 21. Fourier analysis of the pedestaled stimuli of Zemany *et al.*[71] a, Space–time representation of a pedestaled motion stimulus. To minimize edge effects in Fourier analysis, zero's are padded in the x, t space around the visual stimulus. b, Power spectrum of the pedestal + motion stimulus in a. c, Directional energy ratio for a pedestaled motion stimulus as a function of the initial angle θ between the motion component and the 2× pedestal on which it is superimposed. The subscripts 28 and 34 indicate the number of stimulus frames (27 frames is one full cycle). Direction energy $E_{28}(\theta)$ is computed as follows: $E_{28}^{P+M}(i, \theta)$, the motion energy in quadrant i, is the integral of the energies of all the motion components in i. Directional energy is given by energy in quadrants 1 + 3 minus 2 + 4. $E_{28}(\theta)$ is the ratio of directional motion energy of the pedestaled stimulus divided by that of the nonpedestaled motion stimulus, computed for each initial relative phase θ of the pedestal. Graphs show $E_{28}(\theta)$, 1.04 cycles; and $E_{34}(\theta)$, 1.26 cycles. $E_{28}(\theta)$ varies greatly with phase θ, including a change in direction of motion. $E_{34}(\theta)$ has only a small variation with θ.

1. *Reformulation of Pedestal Immunity*

When a continuous periodic stimulus composed of a sum of sine waves of *different* temporal frequencies is pre-

sented to a Reichardt detector that averages either over one period exactly or over a very long interval that contains many periods, its response is equal to the sum of its responses to each of the frequencies individually.[14] This property has been called pseudolinearity.[14]

Because of pseudolinearity, Fourier analysis of the stimuli gives great insight into the behavior of Reichardt detectors. In particular, Fig. 21b shows the complex Fourier power spectrum of a pedestaled stimulus. Of the four most intense points (largest amplitudes) in the power spectrum, two that represent the pedestal are on the spatial frequency axis, and two that represent the moving grating are between the axes. The response of a Reichardt detector to any stimulus on either axis is zero. Therefore, whenever pseudolinearity applies, a pedestal has no influence on the output of a Reichardt detector.

Consider the Fourier spectrum of finite duration and sampled stimuli. The continuous pedestaled stimulus is multiplied by a window to produce a finite-duration stimulus or by a sampling function to produce a sampled stimulus, or by both. Multiplying the function by a window is equivalent to convolving the Fourier transform with the transform of the window. For example, the transform of a rectangular temporal windowing function is a sync function, $\sin(\omega_t)/\omega_t$, that splatters the energy of both the moving grating and the pedestal along a vertical line above and below the horizontal axes (Fig. 21b). The transform of a regular sampling function is itself a sampling grid, which splatters energy over the entire temporal frequency space. The usual computer-generated stimulus is a windowed sampling grid, whose transform is the convolution of the two transforms above, which has even more frequency splatter.

The motion of a moving grating is assumed to be sensed by detectors that are sensitive to spatiotemporal frequencies in a neighborhood around the location of the stimulus in the Fourier domain (and also around its mirror-image locations, which represent movement in the opposite direction).

For brief-duration stimuli, because of energy splatter, pseudolinearity does not hold exactly. Therefore, Reichardt detectors presented with pedestaled motion in brief-duration stimuli do not have absolute pedestal immunity. But for certain brief stimuli, such as a 2× pedestaled motion stimulus that is sampled every 90 deg and is presented for $4n + 1$ frames, the net directional motion energy (quadrants $Q_1 + Q_3$ minus $Q_2 + Q_4$) is independent of the pedestal. Therefore we expect and we observe pedestal immunity for these particular stimuli.

2. *Analytic Predictions of a Simplified Reichardt Detector*

Consider a simplified "point-delay" Reichardt detector (Fig. 22a) which has (1) a spatial separation of Δx between the left and the right detector units, (2) a temporal delay Δt that is exactly equal to the frame-to-frame time, and (3) an integration time that is long relative to the duration of the stimulus (a property originally assumed by Reichardt[13]). At time $t + \Delta t$, the subunit R of this Reichardt detector multiplies the input at location $x + \Delta x$, $I(x + \Delta x, t + \Delta t)$ with the delayed input at location x, $I(x, t)$; the subunit L multiplies the input at location x, $I(x, t + \Delta t)$ with the delayed input at location $x + \Delta x$, $I(x + \Delta x, t)$. The output of the detector $E(x, t + \Delta t)$ is the difference between the right subunit R and the left subunit L:

$$E(x, t + \Delta t) = I(x, t)I(x + \Delta x, t + \Delta t) - I(x + \Delta x, t)I(x, t + \Delta t). \quad (11)$$

For a stimulus with a duration of $K\Delta t$, the output of the Reichardt detector at time $t + k\Delta t$, $k = 1, ..., K$, is

$$E(x, t + k\Delta t) = I(x, t + (k - 1)\Delta t)I(x + \Delta x, t + k\Delta t) - I(x + \Delta x, t + (k - 1)\Delta t)I(x, t + k\Delta t). \quad (12)$$

Motion direction of the stimulus is determined by the integration of the Reichardt detector throughout the duration of the stimulus (assumption 3):

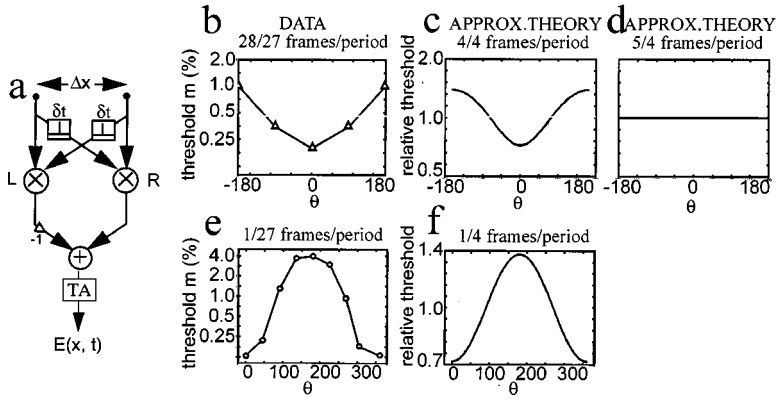


Fig. 22. Analytic predictions of a simplified Reichardt detector for four-frame and five-frame pedestaled (I^{P+M}) and nonpedestaled (I^{M}) motion stimuli. a, Simplified Reichardt detector composed of (1) a spatial separation of Δx between the left and the right detector units, (2) a temporal delay Δt that is exactly equal to the frame-to-frame time, and (3) an integration time (temporal averaging, TA) that is long relative to the duration of the stimulus. b, Data from Zemany *et al.*[71] showing that motion threshold depends strongly on the relative phase of the pedestal and the motion stimulus. c, Theoretical prediction by the simplified Reichardt detector that matches the data of b very well. d, Theoretical prediction by the simplified Reichardt detector for a five-frame stimulus: complete pedestal immunity (matches the data of Lu and Sperling[41]). e, One-frame pedestal data of Zemany *et al.*[71] f, Theoretical prediction of e by the simplified Reichardt detector.

$$E(x) = \sum_{k=1}^{K} E(x, k\Delta t). \quad (13)$$

If $E(x) > 0$, motion direction is to the right; if $E(x) < 0$, motion direction is to the left.

We use the simplified Reichardt detector as an approximation to the most sensitive Reichardt detector of the following stimuli: (1) A sampled moving sine-wave grating M

$$I^M(x, t) = m \sin(fx + \omega t), \quad (14)$$

with spatial frequency f and temporal frequency ω. The grating jumps 90 deg from frame to frame in a consistent direction every Δt. (2) A pedestaled motion stimulus I^{P+M} in which the moving luminance grating is identical to (1) and the pedestal lasts from the beginning to the end of the motion stimulus:

$$I^{P+M}(x, t) = m \sin(fx + \omega t) + p \sin(fx + \pi + \theta), \quad (15)$$

where θ is defined relative to the central frame of the motion sequence; i.e., $\theta = 0$ if the pedestal and the moving frame are in phase at the middle of the motion sequence. (3) A one-frame pedestal stimulus I^{F+M} in which the moving luminance grating is the same as (1) but the pedestal only occurs in one frame of the stimulus at time t_0:

$$I^{F+M}(x, t) = m \sin(fx + \omega t) + p \sin(fx + \pi + \theta)\delta(t - t_0), \quad (16)$$

where δ denotes a Kronecker delta function.

We derive motion threshold as a function of the relative phase of the pedestal and the moving component in three conditions: (1) a pedestaled motion stimulus with a duration of four frames (exactly one cycle) to approximate Zemany *et al.*'s[71] pedestaled stimulus (28 frames with a period of 27, 1.04 cycles); (2) a pedestaled motion stimulus with a duration of five frames, as used in Lu and Sperling[41]; (3) a motion stimulus with a duration of four frames and a one-frame pedestal occurring in frame three to approximate one of Zemany *et al.*'s[71] stimulus conditions.

1. Four-frame pedestaled motion stimulus. The output of the Reichardt detector to a pedestaled four-frame motion stimulus is given by substituting Eq. (15) (a four-frame pedestaled motion stimulus) into Eq. (13) (the total output of the Reichardt detector), which gives

$$E_4^{P+M} = 3 \sin(f\Delta x)\left[m^2 + \frac{\sqrt{2}}{3} mp \sin\left(\frac{3}{4}\pi + \theta\right)\right]. \quad (17)$$

Similarly, substituting Eq. (14) into Eq. (13) gives the output of the Reichardt detector to a nonpedestaled four-frame motion stimulus:

$$E_4^M = 3m^2 \sin(f\Delta x). \quad (18)$$

Let motion threshold for the nonpedestaled four-frame stimulus be m_0. Choose the pedestal amplitude to be $2m_0$. Given these choices, what amplitude m of the motion component in the pedestaled motion stimulus is required to bring it to threshold? The threshold motion energy (for left–right motion discrimination of nonpedestaled motion) of the Reichardt detector is given by Eq. (18) with m_0 substituted for m. The motion energy for pedestaled motion is given by Eq. (17); at threshold this is equal to Eq. (18). Setting Eq. (17) equal to Eq. (18) gives

$$3 \sin(f\Delta x)\left[m^2 + \frac{\sqrt{2}}{2} m 2m_0 \sin\left(\frac{3}{4}\pi + \theta\right)\right] = 3m_0^2 \sin(f\Delta x). \quad (19)$$

A complication in the Zemany *et al.*[71] experiments that the authors apparently overlooked is that superimposing a rightward moving sine-wave grating on a pedestaled motion stimulus with relative phase θ is not a symmetric time reversal of superimposing a leftward-moving sine-wave grating with relative phase θ. Rightward and leftward motion must be computed separately. (In the ideal case, motion-direction discrimination is computed from true mirror-image stimuli, which of course do have the same theoretical threshold.) Since Zemany *et al.*[71] provide only the average of leftward and rightward motion thresholds, we combine the leftward and rightward predicted motion thresholds. Given Eq. (19), the definition of θ, and a rightward moving four-frame pedestaled stimulus, the threshold amplitude $m_\tau^r(\theta)$ for motion-direction discrimination (versus its mirror image) is

$$m_\tau^r(\theta) = \frac{m_0}{6}\left\{-2^{3/2} \sin\left(\frac{3}{4}\pi + \theta\right) + \left[48 + 8 \sin^2\left(\frac{3}{4}\pi + \theta\right)\right]^{1/2}\right\}. \quad (20)$$

Threshold amplitude for direction discrimination for leftward motion, $m_\tau^l(\theta)$ for the four-frame pedestaled stimulus, is

$$m_\tau^l(\theta) = \frac{m_0}{6}\left\{-2^{3/2} \sin\left(\frac{3}{4}\pi - \theta\right) + \left[48 + 8 \sin^2\left(\frac{3}{4}\pi - \theta\right)\right]^{1/2}\right\}. \quad (21)$$

We take the mean of $m_\tau^l(\theta)$ and $m_\tau^r(\theta)$ as the motion threshold $m_\tau(\theta)$:

$$m_\tau(\theta) = \frac{m_0}{2}[m_\tau^l(\theta) + m_\tau^r(\theta)]. \quad (22)$$

Figure 22c plots relative motion direction discrimination threshold in a four-frame pedestaled stimulus, $m_\tau(\theta)/m_0$, as a function of the relative phase θ. It provides a fairly good account of Zemany *et al.*'s[71] results (Fig. 22b).

Another important aspect of the phase-dependence effect observed by Zemany *et al.*[71] is that the size of the effect depends on the temporal frequency of the motion stimulus. This is due to the effective size of the pedestal in the visual system. Because the envelope of the pedestal has lower temporal frequency components than the motion stimulus, the higher the temporal frequency, the larger is the effective amplitude of the pedestal relative to

the amplitude of the motion stimulus and therefore the larger is the relative phase effect induced by the pedestal.

2. Five-frame pedestaled motion display. Equations (11)–(13) yield the following output of the Reichardt detector to a five-frame pedestaled motion stimulus:

$$E_5^{P+M} = 4m^2 \sin(f\Delta x) \tag{23}$$

As stated in Ref. 41, the output of the simplified Reichardt detector with temporal delay equal to the frame-to-frame time depends on neither the relative phase θ nor the amplitude of the pedestal p. For comparison, the output of the Reichardt detector for a five-frame nonpedestaled motion stimulus is

$$E_5^{M} = 4m^2 \sin(f\Delta x). \tag{24}$$

E_5^{P+M} and E_5^{M} are actually identical—the simplified Reichardt detector is *completely immune* to the stationary pedestal.[41]

3. Four-frame motion stimulus with a one-frame pedestal. For this stimulus, the output of the Reichardt detector is

$$E_4^{F+M} = 3m^2 \sin(f\Delta x)\left[1 + \frac{p}{m}\cos(\theta)\right]. \tag{25}$$

Let motion threshold for the nonpedestaled four-frame stimulus be m_0 and the pedestal amplitude be $2m_0$. For the motion stimulus with a one-frame pedestal, at threshold,

$$3m^2 \sin(f\Delta x)\left[1 + \frac{2m_0}{m}\cos(\theta)\right] = 3m_0^2 \sin(f\Delta x). \tag{26}$$

Thus motion direction discrimination threshold $m_{\tau_F}(\theta)$ for the motion stimulus with a one-frame pedestal is

$$m_{\tau_F}(\theta) = \frac{1}{6}\{-2\cos(\theta) + [36 + 4\cos^2(\theta)]^{1/2}\}. \tag{27}$$

Figure 22f shows that Eq. (27), which was derived from a simplified Reichardt detector, provides a good account of the one-frame pedestal data of Zemany *et al.*[71]

3. Fourier Analysis

In this section we use Fourier analysis of x, t space to estimate the directional motion power (e.g., Ref. 115) of the pedestaled motion stimulus used by Zemany *et al.*[71] We then extend our analysis to a slightly different stimulus (with a duration of 34 frames instead of 28 frames) to demonstrate that it (unlike the 28-frame stimulus) has a high degree of pedestal immunity. These results are illustrated in Fig. 21.

We first generate stimuli in two-dimensional x, t space (Fig. 21a) in which the x dimension is sampled continuously and the t dimension is discretely sampled with a period of 27 frames. Two types of stimuli are generated: pure-motion stimuli $I_N^M(x, t)$ and pedestaled-motion stimuli $I_N^{P+M}(x, t, \theta)$, where N is the number of frames in the stimulus. (In Zemany *et al.*[71] $N = 28$.) We also pad zero's around the visual stimuli in x, t space to minimize the edge effects in Fourier analysis.

Second, we compute the Fourier power spectrum of each of the stimuli: $F^2(I_N^{P+M}(x, t, \theta))$ and $F^2(I_N^M(x, t))$, using a MATLAB[116] program. We then define motion energy in a particular quadrant i as the integration of Fourier power in that part of the Fourier space:

$$E_N^M(i) = \int_{Q_i} F^2(I_N^M(x, t))\mathrm{d}\omega_t \mathrm{d}\omega_x, \tag{28}$$

$$E_N^{P+M}(i, \theta) = \int_{Q_i} F^2(I_N^{P+M}(x, t, \theta))\mathrm{d}\omega_t \mathrm{d}\omega_x. \tag{29}$$

The total motion energy, whose sign determines the direction of motion, is defined as

$$E_N^M = [E_N^M(1) + E_N^M(3)] - [E_N^M(2) + E_N^M(4)], \tag{30}$$

$$E_N^{P+M}(\theta) = [E_N^{P+M}(1, \theta) + E_N^{P+M}(3, \theta)] - [E_N^{P+M}(2, \theta) + E_N^{P+M}(4, \theta)]. \tag{31}$$

Equations (30) and (31) permit the computation of the motion energy $E_N(\theta)$ of an N-frame pedestaled motion stimulus with pedestal-to-motion phase θ relative to the equivalent nonpedestaled stimulus:

$$E_N(\theta) = \frac{E_N^{P+M}(\theta)}{E_N^M}. \tag{32}$$

The ratio of pedestaled to nonpedestaled motion energy $E_N(\theta)$ as a function of θ is shown in Fig. 21c for two different N's: $N = 28$ and $N = 34$. When $N = 28$, corresponding to the stimulus used by Zemany *et al.*,[71] $E_{28}(\theta)$ achieves its minimum value 0.2654 at $\theta = -180$ deg and 180 deg, its maximum value 1.73 at $\theta = 0$ deg. In other words, when $N = 28$, the resulting relative motion energy computed from the pedestaled stimulus has very strong phase dependence. Thus the corresponding motion-direction-discrimination threshold—which is assumed to require motion energy to reach a certain value relative to the internal noise in the visual system—is predicted to exhibit the very strong phase dependence shown in Fig. 20f.

On the other hand, when $N = 34$, corresponding to a condition in which the duration of the pedestaled motion stimulus is one full cycle plus ~90 deg, $E_{34}(\theta)$ has only a very weak dependence on θ: $E_{34}(\theta)$ reaches its minimum value 0.941 at $\theta = -135$ deg and its maximum value 1.059 at $\theta = 45$ deg. Overall, $E_{34}(\theta)$ is quite close to 1.0. Whereas perfect pedestal immunity requires the ratio to be exactly 1.0, using 34 frames yields a maximum deviation of 6% and an average deviation from 1.0 less than 3%. So we predict that if Zemany *et al.*[71] were to use 34 frames instead of 28, they would observe approximate pedestal immunity.

4. Conclusion

The apparent inconsistencies between Zemany *et al.*[71] and Lu and Sperling[41] on the effects of pedestals on direction-of-motion thresholds are resolved within the framework of the Reichardt (or equivalent motion-energy) detector theory, which predicts both sets of results. By means of this resolution, instead of contradicting the Reichardt model (as the authors suggested), the Zemany

et al.[71] data provide strong support for the Reichardt detector as a computational algorithm for first-order motion perception.

H. First-Order Artifacts in Second-Order Motion Stimuli

Light adaptation and contrast gain control in the visual system before motion detection (e.g., Fig. 11) may distort the intensity values of the signal presented to the motion system. When motion systems receive a distorted signal, and the experimenter or theorist believes the systems to have received an undistorted signal, the conclusions about motion processing may be incorrect. The distortion itself may be regarded as producing a contamination.

Smith and Ledgeway[72] allege that imposing a moving texture-contrast modulation on a static instead of a dynamic carrier (both second-order stimuli), artifactually produces a first-order motion stimulus and that therefore results based on such stimuli (e.g., Ref. 41) are erroneous. The basis of their claim was that second-order stimuli with static and dynamic carriers produced different temporal-frequency-tuning functions: The one with static carrier mimicked first-order motion, whereas the one with dynamic carrier cut off at a much lower temporal frequency. They concluded that the observed similarity between the temporal tuning characteristics of first- and second-order motion systems by Lu and Sperling[41] was due to a first-order artifact in the second-order stimuli because these stimuli used static carriers. They attributed the first-order artifacts to two different sources: (1) Distortion products (visible to first-order mechanisms) occur in second-order stimuli made of static carriers. (2) Selective attention to homogeneous areas permits operation of first-order mechanisms. We discuss their experimental results and each of the alleged artifacts in turn.

1. Different Temporal Tuning Functions: Dynamic versus Static Carrier

A second-order temporal tuning function typically is measured by modulating a random-noise carrier with a contrast-modulated sine wave (e.g., Fig. 7h). The modulator translates (typically 90 deg) from frame to frame in a consistent direction. The static-noise carrier remains unchanged; only the contrast modulation changes (translates) from frame to frame. With a dynamic carrier, a new noise sample is chosen for each frame. The modulator is the same in both static and dynamic cases. The tuning function (e.g., Fig. 6) is measured by varying the frame rate and determining the amplitude threshold of the modulation for 75% correct left–right motion discrimination.

Smith and Ledgeway[117] found that static and dynamic carriers produced different temporal frequency tuning functions for second-order motion. Whereas the static-carrier tuning function was similar to the first-order motion tuning function, the dynamic-carrier tuning function had a lower temporal corner frequency (Fig. 23b) and greatly reduced high-frequency sensitivity. On the basis of this observation, they concluded that there were first-order artifacts in second-order motion stimuli and that these first-order artifacts caused the supposed second-order tuning function to resemble a first-order motion tuning function. They alleged that second-order stimuli with static carriers, i.e., the stimuli used by Lu and Sperling,[41] produced artifactual data and that only dynamic carriers produce legitimate second-order motion (Ref. 117, p. 404).

We demonstrate below (Experiment 1) that Smith and Ledgeway's[117] different temporal tuning functions for first- and second-order motion are a consequence of the vastly different stimuli used to measure them, not of the different systems to which they are directed. When stimuli that contain equivalent amounts of high-frequency noise are used to measure tuning for first- and second-order motion, the first- and second-order tuning functions are remarkably similar.

2. Experiment 1: Temporal Tuning Functions with and without Noise Masking

Display apparatus. The stimuli were created off-line, displayed on a personal computer with a 60-Hz refresh rate, a white phosphor, and 12 bits of intensity resolution. Mean luminance was 169 cd/m^2, and a linearized, high-

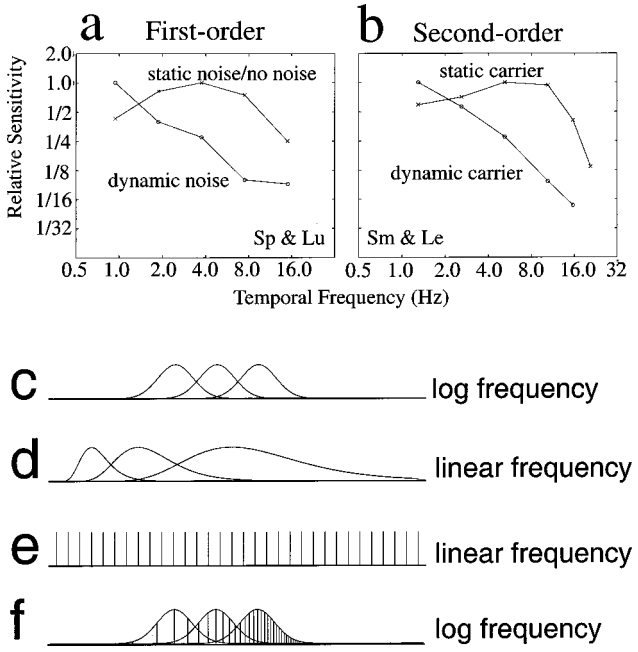


Fig. 23. Tuning functions for luminance-modulation and contrast-modulation motion stimuli masked by broad-spectrum white static and dynamic noise. a, Temporal tuning functions for luminance modulation motion with static noise (same as without noise)[41] and with dynamic noise.[81] For first-order motion, this static noise has no effect. The dynamic carrier impairs performance approximately in proportion to temporal frequency. b, Temporal tuning functions for contrast modulation motion obtained by Smith and Ledgeway (Ref. 115, Fig. 2) for a static and a dynamic carrier. The data from the static carrier are similar to those of Lu and Sperling (1995, Fig. 7).[41] c, Temporal frequency channels with equal bandwidth in log frequency. d, Channels of c on a linear frequency scale. e, The expected Fourier power spectrum of a dynamic second-order carrier has equal Fourier power at all temporal frequencies within the achievable range. f, The temporal frequency channels of c contain more noise frequencies and hence more noise energy at higher frequencies. Doubling the temporal frequency doubles the amount of noise within the channel, thereby producing a loss of sensitivity that is inversely proportional to temporal frequency.

resolution intensity scale was produced.[118] All the stimuli were viewed binocularly with natural pupils in a dimly lighted room.

Stimuli. The luminance-modulation stimulus is a rigidly translating sine-wave grating, with spatial frequency $\alpha = 2.55$ cpd, modulation amplitude m, and temporal frequency f_j, $f_j = 0.94, 1.88, 3.75, 7.50, 15.0$ Hz:

$$L(x, y, t, \beta, j) = L_0\{1.0 + m \sin[2\pi(\alpha x + \beta f_j t)]\}, \quad (33)$$

where $\beta = \pm 1$ indicates the direction of motion ($\beta = +1$ for leftward motion and $\beta = -1$ for rightward motion).

All luminance modulations extended 3.13 deg horizontally and 1.57 deg vertically centered in a uniform background extending to 17.2×11.3 deg. The motion stimulus started with a random temporal phase. It always lasted a full temporal cycle plus one extra frame. The extra frame was added so that the last frame was identical to the first frame. In this way, we removed any positional cue on which observers could base their judgments.

Dynamic binary masking noise of either 0 or $\pm 20\%$ was added to the motion stimulus. The size of noise pixels was 6×6 min^2. A new frame of external noise was presented at every frame refresh (60 Hz).

Determining the frequency characteristic requires that a threshold be measured at each of the frequencies to be tested. The method of constant stimuli[121] was used to generate a psychometric function; the estimated 75% correct point is the threshold. Psychometric functions were obtained for the two noise conditions, for each of the temporal frequencies tested, and for each observer. Eighty observations were made by each observer at every point on the psychometric functions.

For a given noise condition n and temporal frequency f_j, the observer's threshold is designated as $m_{75}(n, f_j)$. Temporal sensitivity functions were generated by plotting $\log_{10}[1/m_{75}(n, f_j)]$ as a function of $\log_{10}(f_j)$ for different stimulus types.

Observers. A UCI graduate student (JM), naïve to the purposes of the experiments, and the first author served as observers in all the experiments. Both have corrected-to-normal vision.

Results. Comparison with Smith and Ledgeway.[117] The data for the two observers were highly similar, the data for one are shown in Fig. 23a (dynamic noise). The data labeled "static noise/no noise" were previously obtained from the same observer with static noise (Ref. 41, Fig. 7) and within statistical error are identical to data obtained with no noise. Figure 23b shows the temporal tuning functions for second-order motion obtained by Smith and Ledgeway (Ref. 117, Fig. 2) for a static and a dynamic carrier. For luminance modulation, the static carrier is like a pedestal—it is invisible to the motion system and has no effect as long as the carrier contrast is insufficient to activate contrast gain-control mechanisms. These luminance modulation thresholds are the same with or without the carrier. With a static carrier, the first-order and second-order motion system tuning functions are essentially the same. The tuning functions obtained with dynamic carriers also are the same for luminance modulation motion that primarily addresses the first-order system and for contrast-modulation motion that addresses the second-order motion system.

Conclusion. Figure 23 shows that the first-order and second-order tuning functions are virtually identical, all the more remarkable because they were obtained in different laboratories with different observers. The conclusion is that the difference that Smith and Ledgeway observed between first- and second-order tuning functions is not due to an "artifact" as they asserted in the title of their paper but to the different properties of the vastly different first- and second-order carriers they were comparing. When the carrier is made the same for first- and second-order motion, the tuning functions are remarkably similar.

3. Basis of the High-Frequency Falloff with Dynamic Carriers

Suppose that the temporal frequency channels of the visual system, like the spatial frequency channels, are of more or less equal bandwidth in log frequency, i.e., in octaves (Fig. 23c).[122,123] Then, in linear frequency space, the channel bandwidth increases with the central frequency of the channel (Fig. 23d). The expected Fourier power spectrum of the dynamic second-order carrier is white in the linear frequency space, i.e., has equal Fourier power at all spatial and temporal frequencies (Fig. 23e) within the achievable range. When bandwidth is measured in octaves, temporal channels with high central frequencies contain more noise energy (Fig. 23f) than low-frequency channels. Doubling the temporal frequency doubles the amount of noise within such a channel. In fact, the slope of the dynamic-noise tuning function is very nearly -1, which is precisely what would be predicted from a white-noise stimulus in channels with a constant octave bandwidth (within the temporal frequency range in which channel sensitivity is approximately constant, as measured by static noise). The conclusion is that the difference between the tuning functions for static and dynamic noise is predictable from the basic principles of temporal receptive fields and the stimulus construction. Therefore it is not surprising that the same relation holds for both first- and second-order motion stimuli.

4. First-Order Distortion Products in Second-Order Stimuli: Three Regimes

Consider the alleged distortion products in three different regimes of second-order motion stimuli (Fig. 24): (1) small carrier with small modulation, (2) large carrier with large modulation, and (3) large carrier with small modulation.

(1) Small carrier with small modulation. The relative phase method and the results in Lu and Sperling (Ref. 41, Experiment 4) provide a sensitive estimate of the amount of distortion products in this regime (none). Suppose there were a first-order distortion product in a second-order stimulus such as Fig. 24a. Then, adding a first-order stimulus of the same phase should reinforce the distortion product in some phases and cancel it in others. The distortion product could, in principle be of the same frequency as the second-order stimulus (most likely) or of double frequency (less likely). When near-threshold

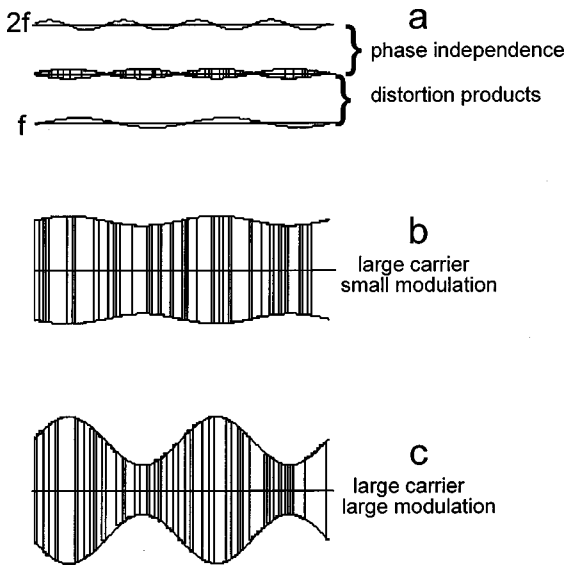


Fig. 24. Contrast-modulation stimuli used to stimulate the second-order motion system. Early visual processing produces different distortion products for the three regimes: a, small carrier with small modulation; b, large carrier with small modulation, and c, large carrier with large modulation.

first-order stimuli of equal or half spatial frequency were added to a threshold second-order stimuli moving in the same direction (Fig. 24a) with different initial relative spatial phases, Lu and Sperling[41] showed that observer's performance did not depend on the relative phase of the two types of stimuli; their performance was equal to or better than the prediction of probability summation from two independent motion computations. The absence of a phase effect indicates that there are no first-order distortion products. The improved performance with two stimuli versus one indicates that two different channels are contributing to performance. Conclusion: No phase dependence implies no distortion of near-threshold second-order stimuli.

(2) Large carrier with large modulation. Second-order stimuli with large-amplitude carrier and modulators, the sort often used for demonstrations of second-order motion, are certainly not transmitted accurately to the motion systems by the early visual system. There is compressive intensity nonlinearity that results in significant distortion products. One can quickly estimate the amount of distortion products by adding a first-order stimulus with slightly different spatial and temporal frequencies and observing beats[81] or use the phase methods outlined above. Scott-Samuels and Georgeson,[61] and Lu and Sperling[124] discuss calibration procedures. The main problem is that the corrections (for early visual system intensity distortion) are slightly different for different observers, so it is not possible to guarantee that a particular high-amplitude second-order stimulus will be distortion free for all observers.[125]

(3) Large carrier with small modulation. Measurements of second-order tuning functions are typically carried out with large-amplitude carriers (see Ref. 84) and small-amplitude modulations for most of the frequency range. To evaluate the amount of first-order distortion products in such near-threshold second-order motion stimuli requires indirect measurements of amplitude because the visually produced distortion products are below motion threshold, as will be shown below.

5. Experiment 2: Measuring First-Order Contamination in a Threshold Second-Order Stimulus

Display apparatus. The stimuli were created dynamically on a Power PC Macintosh. Stimuli were displayed on monitor with a white phosphor, 120 frames/s, 12.6 bits of intensity resolution, and a carefully linearized intensity scale. The background and the mean stimulus luminance was 27 cd/m^2. All the stimuli were viewed binocularly with natural pupils in a dimly lighted room.[126,127]

Stimuli and procedure: moving luminance grating (Fig. 25a). A square-wave luminance grating (1.28 c/deg) moved 90 deg from frame to frame at 7.5 Hz and lasted five frames. To make this condition similar to subsequent texture-modulation conditions, the same static binary noise (pixel size 3 × 3 min) at ±50% contrast that was used subsequently was added to the luminance grating. Threshold modulation at 75% correct motion-direction discrimination was measured by the method of constant stimuli and found to be ~1.0% (A, Fig. 25d).

Moving texture-contrast grating (Fig. 25b). The square-wave texture-contrast grating (1.28 c/deg) moved 90 deg from frame to frame at 7.5 Hz and lasted five frames. The carrier texture was made of static binary noise (pixel size 3 × 3 min) at ±50% contrast. The grating is made of *Carrier* × (1 + *B* · *modulator*). Threshold modulation *B* was measured by the method of constant stimuli and found to be ~4.5% (Fig. 25e).

Sandwich stimulus (Fig. 25c). In the sandwich stimulus, odd frames were made of square-wave texture-contrast gratings at threshold *B* (4.5% modulation depth for ZL). Even frames were made of luminance square-wave gratings. The phase shift between odd frames was 180 deg; the phase shift between even frames was 180 deg. The phase shift between successive odd and even frames was 90 deg. When there was no (physical or perceptually produced) luminance contamination, no motion system was activated by the stimulus because the odd frames addressed only the second-order system, the even frames, only the first-order system. The direction of apparent motion, when it was perceived, indicated the phase of the (physical or perceptually produced) luminance contamination.

Luminance gratings of small amplitude were added to the odd frames, either "in phase" (align the bright region in the first-order stimulus with the low-contrast region of the second-order stimulus), or "out of phase" (align the bright region in the first-order stimulus to the high-contrast region of the second-order stimulus). The method of constant stimuli was used to determine the amplitude *C* of the luminance grating that reduced motion-direction-discrimination performance to chance. In the sandwich stimulus, *C* was estimated to be 0.40%.

When there is appreciable luminance contamination in the high-contrast texture grating, the sandwich grating in Fig. 25c will appear to move consistently in one direction. Luminance modulation is added to find two points: the 75% threshold for rightward apparent movement and the 75% threshold for leftward movement (i.e., 25% rightward

movement). We estimate the point of perfect motion cancellation (50%) as being midway between the 25% and the 75% point, and designate its contrast as C (Fig. 25f). C is the contrast of a luminance (first-order) contamination in the contrast-modulation (second-order) stimulus. Referring C to the prior estimate of luminance thresholds with the same carrier (Fig. 25d) yields an estimate of the magnitude of the first-order distortion product (contamination) in the second-order stimulus. The contamination is ~1/2 the amplitude of a first-order stimulus that achieves 75% threshold. The conclusion is that a second-order stimulus with a large static carrier and a near-threshold second-order modulation will produce a visual first-order distortion product that is 1/2 threshold and, by itself, would seldom be visible.

A second phenomenon illustrated in Fig. 25f is that by appropriately choosing the amplitude of the even (luminance) frames in the sandwich display (Fig. 25c) reduces the distance between the 25% and the 75% points of the psychometric function. This is motion amplification.[124] Amplification factors greater than five are easily obtained in such sandwich displays and are useful for making quick, very accurate estimates of possible (below-threshold) contamination components.

6. Selective Attending

Smith and Ledgeway[72] also claimed that selective attention to homogeneous areas permits operation of first-order mechanisms. However, introspectively, in the stimuli that we use (e.g., Refs. 40 and 41) we do not see such areas; they are too small and too evanescent. Moreover, all the stimuli that we use (e.g., Ref. 41), and all the stimuli described by Smith and Ledgeway[72] are microbalanced.[30] This means that in any area—large or small—there is equal expected first-order motion energy in opposite directions. There is no area in any microbalanced stimulus where selective attention to that area could make possible a first-order computation to extract the second-order motion direction.

7. First-Order Artifacts in Second-Order Stimuli: Conclusions

Because the visual system does not represent stimulus intensity perfectly, all stimuli are distorted to some degree before they reach a motion detector. Measurements of these distortions (described above) showed that, for second-order motion stimuli with low texture contrast, these distortions do not produce any visible first-order contamination. For texture stimuli with high contrasts but small-amplitude motion modulations, such as might be used to measure motion thresholds, the first-order distortion product can be 1/2 the amplitude of a threshold motion threshold stimulus; i.e., by itself it would rarely be visible, but it could reduce the value of an estimated second-order motion threshold. Using sensitive methods to measure and thereby to compensate perceptual distortions of second-order stimuli leaves second-order motion perfectly intact and does not change the appearance of second-order motion.[61,124] However, demonstrations of second-order motion that involve large-amplitude motion modulations should be individually calibrated to assure that perceptual distortions do not produce visible first-order motion components. The perceived motion in the second-order stimuli that have been described so far is not just caused by first-order contamination; it is true second-order motion.

The apparent differences in the temporal tuning functions for first- and second-order motion reported by Smith and Ledgeway[117] were due to the different carriers that they used in the two measurements, not to a first-order

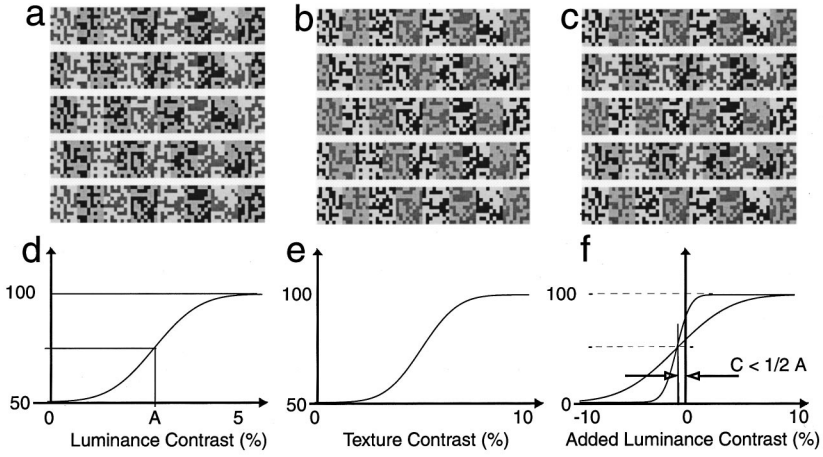


Fig. 25. Procedure for estimating first-order contamination in the second-order stimuli.[124] a, First-order (luminance modulation) motion stimulus on static-random-noise carrier with ±50% contrast. d, Psychometric function (percent correct motion-direction judgment versus luminance-modulation amplitude for stimulus a). "A" is the 75% correct motion-direction threshold. b, Moving contrast modulation imposed on the same ±50% static random carrier as in a. e, Psychometric function for b for determining the contrast-modulation amplitude B that yields 75%-correct motion-direction thresholds. c, Sandwich method[124] for measuring the amount of first-order contamination C in the second-order stimulus at its threshold as determined in b. Odd frames are the same as for the 75% threshold in b. Even frames have pure luminance modulation. Direction bias (ordinate) depends on added luminance contrast in odd frames and on luminance amplitude in even frames. [On even frames, luminance modulations m_e of 6–8× motion threshold produce the steepest psychometric functions (greatest motion amplification[124])]. Two psychometric functions are shown for two values of m_e. Luminance modulation C is added to odd frames to bring performance to chance (50%). The canceling modulation reveals the magnitude of the contamination product. Typically $C < (1/2)A$; i.e., a contrast modulation stimulus with a large static carrier and a near-threshold second-order modulation contains a first-order distortion product that is 1/2 threshold and would normally be invisible.

artifact as they claimed in the title of their publication. When the same carrier was used in the determination of temporal tuning functions, first- and second-order had essentially identical tuning functions. The effects of the carrier on the tuning function could be understood in terms of the temporal bandwidth of visual detectors. Stimuli that stimulate the second-order motion system without first-order contamination can be produced with either dynamic or static texture carriers.

4. SOME EVIDENCE IN SUPPORT OF ASPECTS OF THE THREE-SYSTEMS THEORY

A. Second-Order Stimuli Are Not Transformed into First-Order Stimuli by an Early Nonlinearity

Scott-Samuel and Georgeson[61] used a nulling method to measure the distortion product and then asked whether this early distortion could account for perception of second-order motion. The stimulus sequence consisted of alternate frames of contrast-modulated (CM, 100% modulation) and luminance-modulated (LM) patterns. The texture carrier was either two dimensional binary noise or a 4-c/deg grating. The moving modulation was 0.6 c/deg. The carrier remained stationary while the phase of the modulating signal (LM alternating with CM) stepped successively through 90 deg to the left or right. Motion was seen in a direction opposite to the phase stepping, consistent with early compressive distortion that induces an out-of-phase LM component into the CM stimulus. The authors measured distortion amplitude by adding LM to the CM frames to null the perceived motion. Distortion increased as the square of carrier contrast, as predicted by the compressive transducer they proposed. Distortion also increased with modulation drift rate, implying that the transducer is time dependent, not static. They concluded that early compressive nonlinearity does induce measurable first-order artifacts into second-order stimuli. Nevertheless, this does not account for second-order motion, because perceived motion of a pure second-order sequence (CM in every frame) could not be nulled by adding LM components. This implies that two pathways for motion (first-order, second-order) do exist.

It is worth noting that, in addition to studying second-order motion produced by imposing moving modulations on noise carriers, Scott-Samuels and Georgeson[61] also used high-spatial-frequency sine-wave carriers. The disadvantage of a sine-wave carrier, noted by the authors, is that the modulator produces real, physical first-order movement components that interfere with the measurement of the visually produced first-order motion components. The random noise carriers and modulators used in the experiments described herein have been proved not to contain systematic first-order components,[30] so any first-order motion that can be canceled must have been visually produced.

B. Two Mechanisms That Extract Motion from Contrast-Modulated Sine-Wave Gratings

Ukkonen and Derrington[128] used the pedestal paradigm (Refs. 14 and 41 and Subsection 2.B.3 above) to investigate the nature of the motion computation involved in discriminating the direction of motion of a moving contrast-modulation imposed on a static sine-wave grating (the carrier). Their stimuli consisted of a carrier, a modulator and a pedestal as described by Eq. (35):

$$L(x,t) = L_0[1 + C\cos(2\pi f_c x + \phi_c) \times \{1 + m\cos[2\pi(f_m x + \omega t) + \phi_m] + p\cos[2\pi f_m x + \phi_p]\}]. \quad (34)$$

The mean luminance of the stimuli was L_0. The carrier was a static sine-wave grating with a relatively high spatial frequency ($f_c = 5$ c/deg) of contrast C. The moving modulator was a sine wave of relatively low spatial frequency ($f_m = 1$ c/deg), temporal frequency ω, and contrast m. The stationary pedestal also was of frequency f_m; it had contrast p. Psychometric functions (percent correct in discriminating motion direction consistent with that of the modulator as a function of modulator contrast m) were determined in different conditions, defined by the contrast of the carrier C, the contrast of the pedestal p, and the temporal frequency of the moving modulator ω.

Ukkonen and Derrington[128] found that, when the contrast of the static carrier sine-wave gratings was high ($C = 45\%$), the presence of pedestals did not affect performance in judging motion direction (pedestal immunity); and the modulation sensitivity remained high up to 12 Hz. When the contrast of the carrier sine-wave gratings was low ($C = 4.5\%$), the presence of pedestals greatly impaired the observer's performance in judging motion direction, and it was impossible to judge motion direction even in the absence of pedestals at 4 Hz. Pedestal immunity and relative high corner frequency are characteristics of second-order motion system; pedestal vulnerability and low corner-frequency (4 Hz) are characteristics of a third-order motion system.[41]

Ukkonen and Derrington[128] concluded that motion of contrast-modulated sine-wave gratings, which are normally considered to be second-order stimuli, can be extracted by two different mechanisms. When the contrast of the carrier sine-wave gratings is low, motion of the contrast envelopes (the modulator) is computed by feature tracking[129] (i.e., the third-order motion system). When the contrast of the carrier is high, motion of the modulators is extracted by a second-order mechanism.

By manipulating the contrast of the carrier, Ukkonen and Derrington[128] clarified the range of operation of two motion systems. Their study illustrates the importance of three often stated principles: (1) In principle, any motion stimulus could (and unless it is highly artificial usually does) activate all three motion systems. (2) Which motion system is most useful in extracting motion from a particular stimulus is an interesting empirical question. (It may even depend on attention. In interocular sine-wave motion, selective attention to the wind was necessary to perceive the first-order motion component.) (3) The distinction between first-, second- and third-order motion systems is computation based, not stimulus based.

C. Determining the Monocularity and/or Binocularity of Motion Systems

Solomon and Morgan[62] developed a clever paradigm to determine whether a type of motion could be perceived monocularly. Adopting a strategy originally used by

Kolb and Braun,[130] they created monocular textures that conveyed a motion signal. The textures were composed of arrays of small texture elements: Gabor patches. In Fig. 26a, the texture consists of two rows of patches slanted +45 alternating with two rows oriented at −45. In the next frame, the 2 + 2 row pattern translates down one row, i.e., 90 deg. Because the second-order texture grabbers are isotropic, insofar as motion is detected it must be by the third-order system. Although the authors do not state so explicitly, we expect that, to see good motion, observers would have to attend to either the +45 or the −45 deg slant and that the motion would then be visible monocularly, binocularly, or interocularly. Indeed, observers had no difficulty in determining the direction of motion of these stimuli. Stimuli of this type are the basis for their experiments.

1. A Stimulus That Contains Monocular but Not Binocular Pattern Motion

Solomon and Morgan[62] present an easily perceivable motion stimulus (as described above) to the left eye and a similar stimulus to the right eye that moves in precisely the same direction. The two stimuli move in perfect synchrony. Each eye individually has a clearly defined motion stimulus, the two stimuli moving in the same direction and at the same speed. The trick is that the left-eye and right-eye stimuli are constructed 180 deg out of phase so that if they were physically summed, the summed stimulus would not contain a motion stimulus. Therefore, if the left- and right-eye stimuli were equally weighted and summed in a cyclopean image in the brain, that cyclopean image would be completely ambiguous with respect to motion direction. Solomon and Morgan[62] find that indeed, observers can easily perceive motion direction in the monocular task but fail completely in the binocular task. This proves that this particular kind of third-order texture motion is not computed monocularly; it is computed only from binocular neurons. These neurons can be driven by either eye to produce a motion signal.

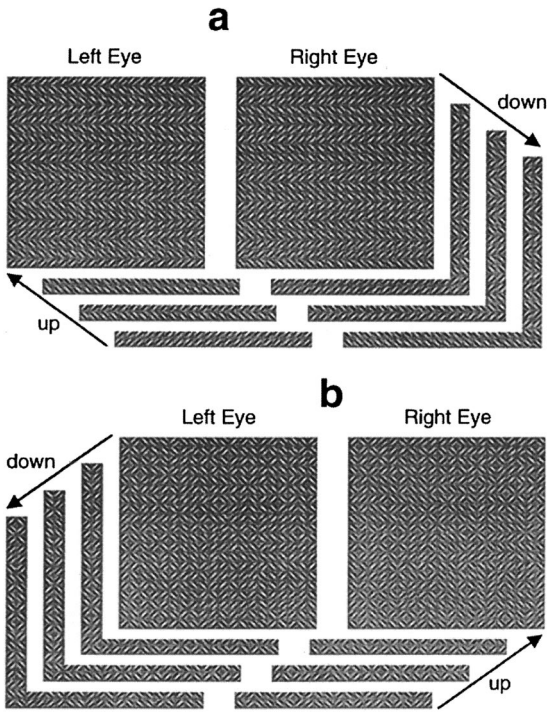


Fig. 26. Dichoptically canceled motion.[62] Five consecutive frames to either the left or the right eye produce one complete cycle of motion defined either by slant, a, or by flicker, b. Motion disappears when left- and right-eye images are physically summed. When viewed dichoptically so the left- and right-eye images are summed in a perceptual cyclopean image, the motion disappears unless it can be computed monocularly before interocular combination. Only the flicker motion is visible dichoptically (i.e., computed monocularly). (After Solomon and Morgan,[62] Fig. 2, p. 2295, with permission of Elsevier Science Ltd.)

2. A Stimulus That Contains Monocular but not Binocular Flicker Motion

The stimulus in Fig. 26b is constructed on the same principles as that in Fig. 26a except that here a flicker-defined (instead of an orientation-defined) grating moves. With each frame change, all the elements in one pair of rows change their orientation (i.e., they produce flicker); all the elements in the alternate set of rows remain unchanged. This flicker-versus-silence pattern translates one row (in the vertical direction) from frame to frame. This is classical second-order motion, the movement of "activity."[80] When such a pattern is viewed by either or both eyes, the direction of motion is completely obvious.

Again, Solomon and Morgan[62] construct two similarly moving stimuli, one in the left and the other in the right eye so that if the stimuli were perfectly summed, the sum would be ambiguous with respect to motion direction. Now, when the observers simultaneously view the binocular stimulus, the motion is easily perceived. This indicates that, unlike translating-orientation motion, translating-flicker motion is computed monocularly before binocular combination.

3. Conclusions

Third-order motion is computed only after binocular combination; second-order motion is computed before binocular combination. Solomon and Morgan's[62] observations, based on a completely different paradigm, are strongly confirmatory of the same conclusion reached by Lu and Sperling.[41]

D. Competition between the Second- and Third-Order Motion Systems

One method of determining whether there are indeed different motion algorithms for second- and third-order motion is to put them into competition in a compound motion stimulus. That one could perceive either of two directions of motion is suggestive but not proof of two systems. It is more important that other factors such as temporal frequency, interference tasks, monocular versus interocular viewing, and individual differences have predictable, differential effects on the outcome of the competition. These differences, in turn, help to define the actual algorithms and more general properties, such as temporal tuning functions, that characterize the second- and third-order motion systems.

When Ho[53] created ambiguous motion stimuli similar to those of Werkhoven *et al.*,[80] she discovered that the

identical stimulus was perceived to move in opposite directions by different observers. In one direction, the second-order direction, successive pattern elements along the motion path have a greater effective contrast but vary in spatial frequency. In the other direction, the third-order direction, successive elements along the motion path have lower average contrast but have the same spatial frequency. Because attention (to a pattern) can strongly influence third-order motion[40] but not second-order motion,[73] Ho introduced a concurrent letter search task (which requires attention).[73] As predicted, the attention-demanding task[40] interfered with perceiving the third-order motion direction—but not the second-order direction. The inverse relation also held: A concurrent third-order (but not second-order) motion task interfered with the letter search task.

The Werkhoven *et al.*[80] and Ho[53] second-versus-third-order motion paradigms were extended to examine the effect of temporal frequency and of interocular versus monocular viewing.[57,66] To study temporal frequency, the exact same sequence of frames is shown but the frame-to-frame time is varied. In monocular viewing, one eye, randomly chosen on each trial, sees all the frames. In interocular viewing, even-numbered frames in the sequence are directed to one eye and odd-numbered frames to the other eye. Because there is a 90-deg phase shift from frame to frame, successive frames within one eye have a phase shift of 180 deg, which is direction ambiguous. To perceive a consistent direction of motion in the interocular conditions, information from the two eyes must be combined.

Figure 27a shows an ambiguous motion stimulus in which the motion direction of the higher-contrast elements is rightward (second order), whereas the motion direction of elements that have the same slant is leftward. At a temporal frequency of 1 Hz, by attending to the slant, all observers perceive this stimulus as moving leftward virtually 100% of the time. When temporal frequency is increased to 3–4 Hz, the stimulus becomes directionally ambiguous; above 5 Hz the predominant perception of motion is rightward. Above ~15 Hz, the temporal frequency is too high for consistent motion perception.

To relate the results of the temporal frequency manipulation to previously observed characteristics of second- and third-order motion, recall that the corner (or critical) frequency above which amplitude attention becomes significant typically is ~4 Hz for third-order motion and 10–12 Hz for second-order motion. Therefore it was anticipated that above 3–4 Hz, perception of the third-order motion direction would be greatly reduced (as it was).

Viewing the same motion sequence (Fig. 27a) interocularly instead of monocularly greatly enhances the perceptual strength of third-order direction versus the second-order direction relative to monocular viewing. Again, this is consistent with previous observations[41] that second-order motion is monocular whereas third-order motion is indifferent to the eye of origin.

Very similar results for both temporal frequency and interocular viewing are obtained when the first-order motion system is put into competition with the third-order system, as illustrated in Fig. 27b. That is, at frequencies of 1 Hz on 5%-contrast gratings, observers universally perceive the stimulus to be moving in the third-order direction; from ~5 Hz to the high frequency cutoff, the stimulus moves in the first-order direction. Interocular stimuli almost eliminate motion perception in the first-order direction and thereby greatly increase the probability of perceiving the third-order direction.

A third manipulation to influence second- versus third-order motion perception for stimuli as in Fig. 27 is making the slant of all the grating patches identical. This eliminates the third-order cue, because a particular direction of slant is uninformative about motion direction. Motion at all temporal frequencies is now perceived in the

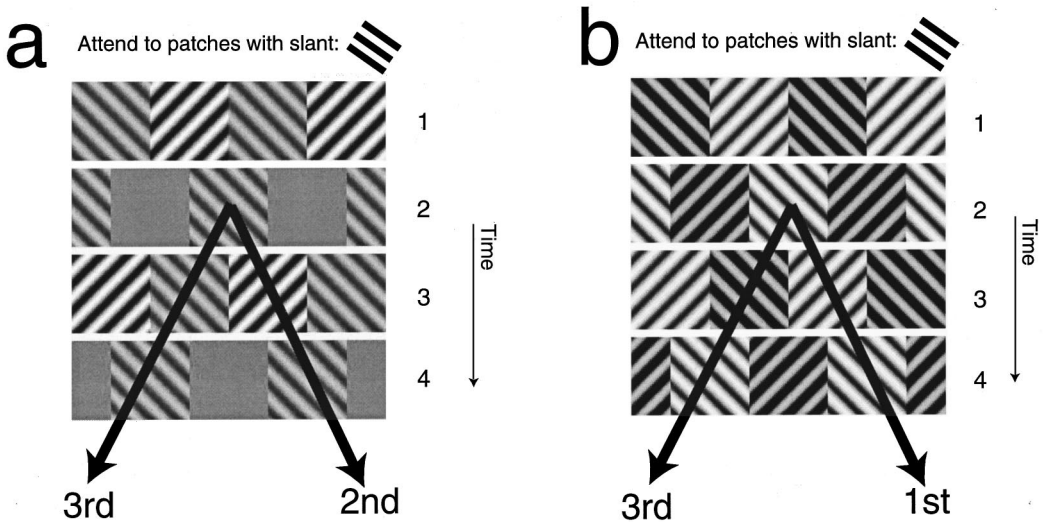


Fig. 27. Motion competition.[66] a, Second- versus third-order motion stimulus with orthogonal-slant patches. Four frames of a longer motion sequence of texture-defined motion stimulus are shown. Two trajectories (left, right) are shown. Along the leftward trajectory (third-order), all frames have the same slant and the same low contrast (0.2). Along the rightward trajectory (second-order), patches with contrast 0.2 alternate with patches with contrast 0.4. Attending the common feature favors the third-order direction, whereas the rightward direction has higher second-order motion energy. b, Luminance-defined motion stimulus. Four frames of a longer motion sequence in which first- and third-order motions are in opposite directions. Frames along the leftward motion trajectory (third order) have the same slant feature. The rightward trajectory has first-order motion energy. For both stimuli, below 3Hz, leftward (third-order) motion is dominant; above 3 Hz, rightward motion dominates.

second-order direction (Fig. 27a), and in the first-order versus third-order direction for the stimuli shown in Fig. 27b. The same-slant data are especially informative because they show that, at some level, both directions of motion are computed over a wide range of frequencies even though, typically, only one becomes conscious.[131] That is, the stronger-direction one wins, and the observer is often completely unaware of the weaker motion-direction computation. These observations are analogous to observations from Burt and Sperling's[131] paradigm. Their observers rated the motion strength of the weaker of two alternative trajectories in an ambiguous-motion paradigm. In many conditions, a trajectory that consistently received ratings of zero (invisible) received high ratings when the competing trajectory was removed. Such observations are consistent with the perceptual computation of all motion strengths and a winner-take-all selection procedure.

For the ambiguous-motion stimuli of Fig. 27, the results of three manipulations (temporal frequency variation, monocular versus interocular viewing, and presence of absence of the third-order slant cue) could be incorporated into a formal mathematical theory. The theory yields the temporal tuning functions of the first-, second-, and third-order systems derived from this class of ambiguous-competition stimuli. The three system tuning functions are quite similar to those derived from pedestaled (and unpedestaled) sine waves in quite different conditions.[41]

E. Motion of an Isoluminant Red–Green Grating Is Processed by and Only by the Third-Order Motion System

We consider here the movement of a vividly colored red–green grating. The red and green stripes are so carefully matched in luminance that there is no luminance-motion component when this grating moves. Because every neuron that responds to color has slightly different characteristics, isoluminance is a statistical balance that applies only to the particular sample of neurons involved in a particular stimulus condition. Therefore creating isoluminant stimuli requires individual calibration for each observer, for each spatial and temporal frequency of grating, for each grating configuration, for each fixation point, and for each degree of color saturation. And a single set of parameters for isoluminance applies only within a limited spatial region.

There are four lines of evidence that isoluminant chromatic motion (ICM) is computed by the third-order motion system and by no other system.[58,59]

1. The temporal tuning function for isoluminant chromatic motion matches the previously measured tuning function for third-order motion (i.e., low-pass, corner frequency approximately 3–5 Hz).
2. ICM fails the pedestal test (which is characteristic only of third-order motion).
3. ICM is perceived exactly, in interocular as well as in monocular viewing. This is characteristic only of third-order motion; first- and second-order motion normally are not perceived interocularly. (When brief stimuli are perceived by the interocular first-order system, there is a 2× to 3× loss of sensitivity.)
4. Motion standstill (perceptual standstill of physically rapidly moving gratings) can be produced with vivid, high-saturation isoluminant gratings. Motion standstill is produced by careful adjustment of the relative saturation of the (isoluminant) red and green stripes (different for each observer and for each spatial and temporal frequency of the grating). Motion standstill is verified by objective failure to judge motion direction (up versus down) and by subjective speed ratings. Standstill can be produced over the whole range of spatial and temporal parameters that *a priori* would have been believed to be ideal for perceiving color motion.

That motion standstill can be produced with vivid high-saturation, rapidly (or slowly) moving color gratings by merely matching the salience of the red and green stripes (i.e., nulling them for third-order motion) means that no other color system is computing the motion of these stimuli (or they would not appear to stand still). Thereby, the observation of motion standstill excludes the possibility of specialized chromatic motion mechanisms within the color pathways.

F. Motion Standstill

The fact that motion standstill occurs, and occurs in a predictable way, is strong evidence for the assumptions underlying a third-order motion system. The underlying principle of third-order motion is that the salience of different portions of the visual field varies as a function of space and time and that the third-order system computes the salience flow field. In particular, in order for grating motion to be perceived by the third-order motion system, the grating must contain a salience modulation, that is, stripes of higher salience alternating with stripes of lower salience.

For an isoluminant red–green grating on a neutral background, it is reasonable to suppose that the salience of the colored stripes is proportional to their color difference from the background. Isoluminant red and green stripes differ from a neutral background only in color saturation. For a red–green grating on a neutral background, salience is assumed to increase monotonically with color saturation. Therefore, for a given saturation of red, there must be a saturation of green that will make the red and green stripes equally salient. At this saturation ratio, there no longer is any red–green salience modulation in the grating. No salience modulation implies no salience motion. That motion standstill occurs indicates that, although the third-order motion system has failed, the shape and color systems are still able to extract the grating's shape and color.

In normal vision, uncontrolled motions of the eyes, head, and body sweep the images across the retina at speeds and over distances that are huge compared with the fine details in the objects and the diameters of rods and cones. The function of shape, color, and texture analysis systems is to extract invariant shapes, colors, and textures from the space–time-varying retinal image. When these systems succeed, they present subsequent

processes with unchanging representations of shapes, colors, and textures. Only the motion systems report movement.

Motion standstill was originally reported by Julesz and Payne[35] in small, rapid back-and-forth oscillations of a dynamic random-dot, stereo-defined depth grating. Tseng *et al.*[113] demonstrated motion standstill in depth-defined stereo gratings even when the gratings moved continuously in a consistent direction. Lu *et al.*[59] demonstrated motion standstill for rapidly moving isoluminant red–green gratings. It is quite possible that it will be possible to demonstrate motion standstill for most pure third-order motions. This is because third-order motion has a poor response at high temporal frequencies and because shape, color, and texture systems seem to maintain responsiveness at high frequencies. In the critical range of temporal frequencies, there is a window of opportunity for observing motion standstill.

Motion standstill can occur in the presence of first- and second-order motion. In pedestaled motion, as the temporal frequency of motion increases, the wobble of the pedestal becomes imperceptible. However, with rapid back-and-forth motion, it is difficult or impossible to discriminate simple averaging from standstill—the (unconscious) perceptual selection of only one of many possible frames. In rapid (high-temporal-frequency) interocular motion without a pedestal, the perceptual standstill of the moving sine grating initially masked the first-order motion output that observers learned to see as a wind sweeping over the grating. Can motion standstill occur with other motion stimuli that stimulate only the first- or second-order motion systems? The sensitivity of first- and second-order motion at high temporal frequencies may exceed that of the shape, color, and texture systems. That may explain why neither first-order nor second-order motion standstill has yet been observed.

5. DISCUSSION ISSUES, SUMMARY, AND CONCLUSIONS

A. Discussion

1. Feature Tracking versus Third-Order Motion

The phenomenon of motion standstill clarifies the distinction between feature tracking and third-order motion. Without a functioning motion system, when a feature moves, it is still represented by the shape, color, and texture systems as an invariant feature at an invariant location. At some point, when it has moved sufficiently, it is re-represented. But there is no sensation of motion. Tracking a feature, that is, knowing that it is moving, requires a motion system. According to this interpretation, the third-order motion system is a prerequisite for feature tracking, whether the tracking be attentional or by eye movements. Without a functional motion system one can have feature search but not feature tracking.

2. Algorithm of Third-Order Motion

Van Santen and Sperling[14] derived three properties of Reichardt models: pedestal immunity, strength multiplication (motion amplification), and pseudolinearity. They experimentally verified these properties for the perception of first-order motion direction. The motion energy model[16] and a model based on Hilbert transforms[22] were shown to be equivalent or nearly equivalent to the Reichardt model.[15–17] The cumulative data are quite constraining. Other candidate models for first-order motion (e.g., Baloch *et al.*,[132] Johnston *et al.*[20]) that have not been as critically tested experimentally, even if fundamentally different, are constrained to be Reichardt-like for a wide range of first-order stimuli. Pedestal immunity[41] and motion amplification[124] were shown to hold for second-order as well as first-order motion.

None of the above properties have been established for third-order motion. Is this because the third-order motion system utilizes a fundamentally different algorithm than the first-order system? Or because the third-order motion system uses the same motion-extraction algorithm as first-order motion but the nature of its input is very different—perhaps much more coarsely quantized? Third-order motion fails the pedestal test,[41] exhibits a minimal multiplication property,[133] and has characteristics different from first- and second-order motion in a variety of motion tasks.[69,134] Unfortunately, none of these tasks discriminates a different type of preprocessing from a different type of algorithm. In Fig. 5 the third-order algorithm was listed as "ENERGY?" because it seemed to us that the genetic argument was still the strongest and because no alternative computational algorithm remained as a viable candidate.

B. Summary

Since the publication of a formal three motion-systems theory[41] and the original evidence, there has been an abundance of both new evidence and new criticism. Strong support for the theory was found in the highly selective neurological damage to first- and to second/third-order motion systems, in brain imaging of first- and second-order systems, in selective adaptation of first-, second-, and third-order motion systems, in different gain-control mechanisms for first- and second-order systems, in the differential effects of attention on the third-order motion system, in second-order reverse phi (the forward direction is third order), in the establishment of isoluminant chromatic motion as third-order motion, in third-order motion standstill, in stimuli that utilized binocular competition to selectively stimulate second- and third-order systems, and in motion competition between systems.

Among the criticisms of the theory were various proposals that a smaller number of systems could account for the data. These proposals collapsed when the data set was enlarged—a particular difficulty being the phase procedure. (For two different sinusoidal stimuli of the same spatial and temporal frequencies and the same strength, moving in the same direction within a single motion system, there must be a phase and amplitude relation where motion perception is canceled and another phase where it is enhanced). Reported failures of pedestal immunity were shown to actually produce results that were entirely consistent with the three-systems theory. Supposed first-order artifacts were shown to be inconsequential in properly constructed second-order stimuli. With selective attention, observers were able to perceive the direc-

tion of briefly displayed interocular first-order motion but not in the presence of effective pedestals—an exception to other instances of first-order motion that are immune to pedestals. Data that allegedly implied a non-Reichardt (nonmotion-energy) detection mechanism for third-order motion were more parsimoniously explained by input filtering.

C. Conclusions

The advantage of a more formal theory is that it is very clear when it fails: Failure of first-order theory to explain second-order stimuli required postulation of a second-order motion system. A reiteration of this process resulted in the postulation of a third-order motion system. The major criticisms directed at second- and third-order motion systems have been resolved. On the other hand, alternative theories[20,132] that encompass some of the data considered here cannot be excluded. The many diverse studies in the six years since the initial publication of the three-systems theory of motion perception that are reviewed here provide much welcome new data and the opportunities for significant clarification and extension of the theory.

ACKNOWLEDGMENT

This research was supported by the U.S. Air Force Office of Scientific Research, Life Science Directorate, Visual Information Processing Program.

Address correspondence including requests for reprints, to Zhong-Lin Lu, Department of Psychology, SGM 501, University of Southern California, Los Angeles, California 90089-1061. Phone; 213-740-2282; fax, 213-746-9082; e-mail, zhonglin@rcf.usc.edu or George Sperling, Department of Cognitive Science, SSPA-3, University of California, Irvine, Irvine, California 92697-5100. Phone; 949-824-6879; fax, 949-824-2517; e-mail, sperling@uci.edu.

REFERENCES AND NOTES

1. E. G. Boring, *Sensation and Perception in the History of Experimental Psychology: A History of Experimental Psychology*, 2nd ed. (Appleton-Century-Crofts, New York, 1942).
2. F. Kenkel, "Untersuchungen ueber Zusammenhang zwischen Erscheinungsgross und Erscheinungsbewegung beim einigen sogenannten optischen Tauschungen," Z. Psychol. **61**, 358–449 (1913).
3. A. Korte, "Kinematoskopische Untersuchungen," Z. Psychol. **72**, 193–206 (1915).
4. M. Wertheimer, "Ueber das Sehen von Scheinbewegunen und Scheinkorpern," Z. Psychol. **61**, 161–265 (1912).
5. O. Braddick, "A short-range process in apparent motion," Vision Res. **14**, 519–529 (1974).
6. A. Pantle and L. Picciano, "A multistable movement display: evidence for two separate motion systems in human vision," Science **193**, 500–502 (1976).
7. G. Mather, P. Cavanagh, and A. M. Anstis, "A moving display which opposes short-range and long-range signals," Perception **14**, 163–166 (1985).
8. M. A. Georgeson and T. M. Shackleton, "Monocular motion sensing, binocular motion perception," Vision Res. **29**, 1511–1523 (1989).
9. P. Cavanagh and G. Mather, "Motion: the long and the short of it," Spatial Vision **4**, 103–129 (1989).
10. P. Cavanagh, "Short-range vs long-range motion: not a valid distinction," Spatial Vision **5**, 303–309 (1991).
11. C. Chubb and G. Sperling, "Two motion perception mechanisms revealed by distance driven reversal of apparent motion," Proc. Natl. Acad. Sci. USA **86**, 2985–2989 (1989).
12. W. Reichardt, "Autokorrelationsauswertung als funktionsprinzip des zentralnervensystems," Z. Naturforsch. **12b**, 447–457 (1957).
13. W. Reichardt, "Autocorrelation, a principle for the evaluation of sensory information by the central nervous system," in *Sensory Communication*, W. A. Rosenblith, ed. (Wiley, New York, 1961).
14. J. P. H. van Santen and G. Sperling, "Temporal covariance model of human motion perception," J. Opt. Soc. Am. A **1**, 451–473 (1984).
15. J. P. H. van Santen and G. Sperling, "Elaborated Reichardt detectors," J. Opt. Soc. Am. A **2**, 300–321 (1985).
16. E. H. Adelson and J. R. Bergen, "Spatio-temporal energy models for the perception of apparent motion," J. Opt. Soc. Am. A **2**, 284–299 (1985).
17. E. H. Adelson and J. R. Bergen, "The extraction of spatio-temporal energy in human and machine vision," in *Motion: Representation and Analysis (IEEE Workshop Proceedings)*, (IEEE Computer Society Press, Washington D.C., 1986) pp. 151–155.
18. D. Marr and S. Ullman, "Directional selectivity and its use in early visual processing," Proc. R. Soc. London Ser. B **211**, 151–180 (1981).
19. C. L. Fennema and W. B. Thompson, "Velocity determination in scenes containing several moving objects," Comput. Graph. **9**, 301–315 (1979).
20. A. Johnston, P. W. McOwan, and H. Buxton, "A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells," Proc. R. Soc. London Ser. B **250**, 297–306 (1992).
21. B. Hassenstein and W. Reichardt, "Systemtheoretische analyse der zeit-, reihenfolgen- and vorzeichenauswertung bei der bewegungsperzeption des russelkafers chlorophanus," Z. Naturforsch. **11b**, 513–524 (1956).
22. A. B. Watson and A. J. Ahumada, Jr., "A look at motion in the frequency domain," in *Motion: Perception and Representation*, J. K. Tsotsos, ed. (Association for Computing Machinery, New York, 1983) pp. 1–10.
23. D. J. Heeger, "A model for the extraction of image flow," J. Opt. Soc. Am. A **4**, 1455–1471 (1987).
24. H. R. Wilson, V. P. Ferrera, and C. Yo, "A psychophysically motivated model for two-dimensional motion perception," Visual Neurosci. **9**, 79–97 (1992).
25. S. J. Nowlan and T. J. Sejnowski, "Filter selection model for motion segmentation and velocity integration," J. Opt. Soc. Am. A **11**, 3177–3200 (1994).
26. V. S. Ramachandran, M. V. Rau, and T. R. Vidyasagar, "Apparent movement with subjective contours," Vision Res. **13**, 1399–1401 (1973).
27. G. Sperling, "Movement perception in computer-driven visual displays," Behav. Res. Methods Instrum. **8**, 144–151 (1976).
28. A. M. M. Lelkens and J. J. Koenderink, "Illusory motion in visual displays," Vision Res. **24**, 293–300 (1984).
29. A. M. Derrington and D. R. Badcock, "Separate detectors for simple and complex grating patterns?" Vision Res. **25**, 1869–1878 (1985).
30. C. Chubb and G. Sperling, "Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception," J. Opt. Soc. Am. A **5**, 1986–2006 (1988).
31. K. Turano and A. Pantle, "On the mechanism that encodes the movement of contrast variations—I: velocity discrimination," Vision Res. **29**, 207–221 (1989).
32. J. D. Victor and M. M. Conte, "Motion mechanisms have only limited access to form information," Vision Res. **30**, 289–301 (1989).
33. C. Chubb and G. Sperling, "Texture quilts: basic tools for studying motion-from-texture," J. Math. Psychol. **35**, 411–442 (1991).

34. G. Sperling, "Visual form and motion perception: psychophysics, computation, and neural networks," presented at Fourier and Non-Fourier Perception of Motion and Orientation, meeting dedicated to the memory of the late Kvetoslav Prazdny, Boston University, Boston, Mass., March 5, 1988.
35. B. Julesz and R. Payne, "Difference between monocular and binocular stroboscopic motion perception," Vision Res. **8**, 433–444 (1968).
36. J. T. Petersik, K. I. Hicks, and A. J. Pantle, "Apparent movement of successively generated subjective figures," Perception **7**, 371–383 (1978).
37. M. Shadlen and T. Carney, "Mechanism of human motion revealed by new cyclopean illusion," Science **232**, 95–97 (1986).
38. P. Cavanagh, M. Arguin, and M. von Grunau, "Interattribute apparent motion," Vision Res. **29**, 1379–1386 (1989).
39. J. M. Zanker, "Theta motion: a paradoxical stimulus to explore higher order motion extraction," Vision Res. **33**, 553–569 (1993).
40. Z.-L. Lu and G. Sperling, "Attention-generated apparent motion," Nature **377**, 237–239 (1995).
41. Z.-L. Lu and G. Sperling, "The functional architecture of human visual motion perception," Vision Res. **35**, 2697–2722 (1995).
42. R. Patterson, "Stereoscopic (cyclopean) motion sensing," Vision Res. **39**, 3329–3345 (1999).
43. Z.-L. Lu and G. Sperling, "Three systems for visual motion perception," Curr. Dir. Psychol. Sci. **5**, 44–53 (1996).
44. G. Sperling and Z.-L. Lu, "A systems analysis of human visual motion perception," in *High-Level Motion Processing*, T. Watanabe, ed. (MIT Press, Cambridge, Mass., 1998), pp. 153–183.
45. W. C. Shipley, F. A. Kenney, and M. E. King, "Beta apparent movement under binocular, monocular and interocular stimulation," Am. J. Psychol. **58**, 545–549 (1945).
46. O. Braddick, "Low-level and high-level processes in apparent motion," Philos. Trans. R. Soc. London, Ser. B **290**, 137–151 (1980).
47. Y.-X. Zhou and C. L. B. Baker, Jr, "A processing stream in mammalian visual cortex neurons for non-Fourier responses," Science **261**, 98–101 (1993).
48. T. D. Albright, "Form-cue invariant motion processing in primate visual cortex," Science **255**, 1141–1143 (1992).
49. T. Ledgeway and A. T. Smith, "Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision," Vision Res. **34**, 2727–2740 (1994).
50. Y.-X. Zhou and C. L. Baker, Jr, "Spatial properties of envelope-responsive cells in area 17 and 18 neurons of the cat," J. Neurophysiol. **75**, 1038–1050 (1996).
51. M. W. Greenlee and A. T. Smith, "Detection and discrimination of first- and second-order motion in patients with unilateral brain damage," J. Neurosci. **17**, 804–818 (1997).
52. S. Nishida, T. Ledgeway, and M. D. Edwards, "Multiple-scale processing for motion in the human visual system," Vision Res. **37**, 2685–2698 (1997).
53. C. E. Ho, "Letter recognition reveals pathways of second-order and third-order motion," Proc. Natl. Acad. Sci. USA **95**, 400–404 (1998).
54. L. P. O'Keefe and J. A. Movshon, "Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey," Visual Neurosci. **15**, 305–317 (1998).
55. A. T. Smith, M. W. Greenlee, K. D. Singh, F. M. Kraemer, and J. Hennig, "The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI)," J. Neurosci. **18**, 3816–3830 (1998).
56. E. Blaser, G. Sperling, and Z.-L. Lu, "Measuring the amplification and the spatial resolution of visual attention," Proc. Natl. Acad. Sci. USA **96**, 11681–11686 (1999).
57. C. E. Ho and G. Sperling, "Selecting second and third-order motion pathways," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **40**, S425 (1999).
58. Z.-L. Lu, L. Lesmes, and G. Sperling, "Mechanisms of isoluminant chromatic motion perception," Proc. Natl. Acad. Sci. USA **96**, 8289–8294 (1999).
59. Z.-L. Lu, L. Lesmes, and G. Sperling, "Perceptual motion standstill from rapidly moving chromatic displays," Proc. Natl. Acad. Sci. USA **96**, 15374–15379 (1999).
60. I. Mareschal and C. L. Baker, Jr., "Cortical processing of second-order motion," Visual Neurosci. **16**, 1–14 (1999).
61. N. E. Scott-Samuel and M. A. Georgeson, "Does early non-linearity account for second-order motion?" Vision Res. **39**, 2853–2865 (1999).
62. J. A. Solomon and M. J. Morgan, "Dichoptically canceled motion," Vision Res. **39**, 2293–2297 (1999).
63. S. Nishida and H. Ashida, "A hierarchical structure of motion system revealed by interocular transfer of flicker motion aftereffects," Vision Res. **40**, 265–278 (2000).
64. A. J. Schofield and M. A. Georgeson, "The temporal properties of first- and second-order vision," Vision Res. **40**, 2475–2487 (2000).
65. G. Sperling, T.-S. Kim, and Z.-L. Lu, "Direction-reversal VEP's reveal signature of first-and second-order motion," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **41**, S334 (2000).
66. G. Sperling and C. E. Ho, "Third-order versus first-order and second-order motion in ambiguous stimuli: competition reveals temporal tuning functions, monocularity/binocularity, and the role of attention," Perception **29**, 83 (2000).
67. E. Taub, J. D. Victor, and M. M. Conte, "Nonlinear preprocessing in short-range motion," Vision Res. **37**, 1459–1477 (1997).
68. N. M. Grzywacz, S. N. J. Watamaniuk, and S. P. McKee, "Temporal coherence theory for the detection and measurement of visual motion," Vision Res. **35**, 3183–3203 (1995).
69. A. E. Seiffert and P. Cavanagh, "Position displacement, not velocity, is the cue to motion detection of second-order stimuli," Vision Res. **38**, 3569–3582 (1998).
70. T. Carney, "Evidence for an early motion system which integrates information from the two eyes," Vision Res. **37**, 2361–2368 (1997).
71. L. Zemany, C. F. Stromeyer III, A. Chaparro, and R. E. Kronauer, "Motion detection on stationary, flashed pedestal gratings: evidence for an opponent-motion mechanism," Vision Res. **38**, 795–812 (1998).
72. A. T. Smith and T. Ledgeway, "Separate detection of moving luminance and contrast modulations: fact or artifact?" Vision Res. **37**, 45–62 (1997).
73. J. A. Solomon and G. Sperling, "Full-wave and half-wave rectification in 2nd-order motion perception," Vision Res. **34**, 2239–2257 (1994).
74. C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," Hum. Neurobiol. **4**, 219–227 (1985).
75. P. Burt, "Attention mechanisms for vision in a dynamic world," in *Proceedings of the Ninth International Conference on Pattern Recognition, Rome, Italy* (IEEE Computer Society Press, Washington, D.C., 1988), pp. 977–987.
76. M. Mozer, *The Perception of Multiple Objects: a Connectionist Approach*, (MIT Press, Cambridge, Mass., 1991).
77. S. Ahmad and S. Omohundro, "Efficient visual search: a connectionist solution," International Computer Science Institute Technical Report tr-91-040 (University of California, Berkeley, Calif., 1991).
78. J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," Artif. Intell. **78**, 507–545 (1995).
79. G. Sperling, A. Reeves, E. Blaser, Z.-L. Lu, and E. Weichselgartner, "Two computational models of attention," in *Attention*, C. Koch and J. Braun, eds. (MIT Press, Cambridge, Mass., 2001), pp. 177–214.
80. P. Werkhoven, G. Sperling, and C. Chubb, "Motion perception between dissimilar gratings: a single channel theory," Vision Res. **33**, 463–485 (1993).
81. G. Sperling and Z. L. Lu, "Update on the three-motion-systems theory," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **39**, S461 (1998).

82. In actual practice, because the response of the early stages of visual processing before motion detection is a linear function of—i.e., faithfully represents—the dark–light difference only when the difference is less than approximately 5%,[83,84] 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.
83. K. Nakayama and G. H. Silverman, "Detection and discrimination of sinusoidal grating displacements," J. Opt. Soc. Am. A **2**, 267–274 (1985).
84. Z.-L. Lu and G. Sperling, "Contrast gain control in first- and second-order motion perception," J. Opt. Soc. Am. A **13**, 2305–2318 (1996).
85. W. Prinzmetal, H. Amiri, K. Allen, and T. Edwards, "Phenomenology of attention: I. Color, location, orientation, and spatial frequency," J. Exp. Psychol. Hum. Percep. Perform. **24**, 261–282 (1998).
86. C. Tseng, J. L. Gobell, and G. Sperling, "Sensitization to color: induced by search, measured by motion," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **41**, S40, Abstr nr. 207 (2000).
87. S. M. Anstis, "Phi movement as a subtraction process," Vision Res. **10**, 1411–1430 (1970).
88. S. M. Anstis and B. J. Rogers, "Illusory reversal of visual depth and movement during changes of contrast," Vision Res. **15**, 957–961 (1975).
89. Second-order reversed phi was first reported by S. Nishida, "Spatiotemporal properties of motion perception for random-check contrast modulations," Vision Res. **33**, 633–645 (1993).
90. Z.-L. Lu and G. Sperling, "Second-order reversed phi," Percept. Psychophys. **61**, 1075–1088 (1999).
91. S. J. Anderson, D. C. Burr, and M. C. Morrone, "Two-dimensional spatial and spatial-frequency selectivity of motion-sensitive mechanisms in human vision," J. Opt. Soc. Am. A **8**, 1340–1351 (1991).
92. Z.-L. Lu, G. Sperling, and J. Beck, "Selective adaptation of three motion systems," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **38**, 237 (1997).
93. L. M. Vaina, N. Makris, D. Kennedy, and A. Cowey, "The selective impairment of the perception of first-order motion by unilateral cortical brain damage," Visual Neurosci. **15**, 333–348 (1998).
94. G. T. Plant, K. D. Laxer, N. M. Barbaro, J. S. Schiffman, and K. Nakayama, "Impaired visual motion perception in the contralateral hemifield following unilateral posterior cerebral lesions," Brain **116**, 1337–1353 (1993).
95. L. M. Vaina, M. Le May, and N. M. Grzywacz, "Deficits of non-Fourier motion perception in a patient with normal performance on short-range motion tasks," Soc. Neurosci. Abstract **19**, 1284 (1993).
96. L. M. Vaina and A. Cowey, "Impairment of the perception of second order motion but not first order motion in a patient with unilateral focal brain damage," Proc. R. Soc. London Ser. B **263**, 1225–1232 (1996).
97. J. Rademacher, A. M. Galaburda, D. N. Kennedy, P. A. Pilipek, and V. S. Caviness, "Human cerebral cortex: localization, parcellation, and morphometry with magnetic resonance imaging," J. Cog. Neurosci. **4**, 352–374 (1992).
98. L. M. Vaina, A. Cowey, and D. Kennedy, "Perception of first- and second-order motion: separable neurological mechanisms?" Hum. Brain Mapping **7**, 67–77 (1999).
99. C. L. Baker, Jr., "Central neural mechanisms for detecting second-order motion," Curr. Opin. Neurobiol. **9**, 461–466 (1999).
100. C. Chubb and G. Sperling, "Second-order motion perception: space–time separable mechanisms," in *Proceedings: Workshop on Visual Motion* (IEEE Computer Society Press, Washington, D.C., 1989), pp. 126–138.
101. For a class of motion models consisting of a pointwise transformation $T[f(x,y)]$ of the contrast of the input stimulus followed by a Reichardt (or equivalent motion energy) computation, this set of stimuli could potentially be used to provide the coefficients of a Taylor series expansion $T[\cdot] = f_K(x,y)$, $(K = 1,\ldots,\infty)$ of an early pointwise nonlinearity, and thereby to exactly define the nonlinearity if $f_K(x,y)$ (K = 1, 2, 3, or 4) activates only the Kth term in the Taylor series expansion of T. This scheme is awkward because, generally, the TVC higher-order stimuli activate more than one term in T. This is probably why Taub *et al.*[67] discarded the Taylor expansion approach. On the other hand, the different terms in the Taylor expansion generate different spatial and temporal frequencies, and this could have been used to isolate and define T.
102. Taub *et al.*[67] use the term "motion amount" instead of the usual "motion energy" to indicate that motion is computed between pairs of points in immediately consecutive images in the stimulus, rather than, more commonly, between nearby or overlapping regions in space–time. To be consistent with the TVC model, TVC's definition of MS_{TVC} is used throughout this section.
103. The TVC motion strength computation is not exactly a Reichardt computation, because the forward motion and the reverse motion are computed between different pairs of points rather than the same pairs of points, as in a Reichardt computation. However, because motion between all pairs of points is summed in the computations under consideration here, this subtle difference is insignificant. Further, the Taub *et al.*[67] motion computation [Eq. (3)] is a pixel-by-pixel computation; all averaging occurs in summing all the pixel-by-pixel motions. A more realistic model would incorporate some averaging (spatial filtering) before each nonlinearity (intensity compression and motion-direction extraction). Our explorations with their model showed that such enhancements, while desirable, do not fundamentally change the character of the predictions. Jonathan D. Victor, Department of Neurology and Neuroscience, Cornell University Medical College, 1300 York Avenue, New York, N.Y. 10021 (personal communication, July 31, 1996).
104. A. Johnston and C. W. G. Clifford, "A unified account of three apparent motion illusions," Vision Res. **35**, 1109–1123 (1995).
105. A. Johnston and C. W. G. Clifford, "Perceived motion of contrast-modulated gratings: predictions of the multichannel gradient model and the role of full-wave rectification," Vision Res. **35**, 1771–1783 (1995).
106. A. Johnston, C. P. Benton, and P. W. McOwan, "Induced motion at texture-defined motion boundaries," Proc. R. Soc. London Ser. B **266**, 2441–2450 (1999).
107. K. Nakayama and C. W. Tyler, "Psychophysical isolation of movement sensitivity by removal of familiar position cues," Vision Res. **21**, 427–433 (1981).
108. D. H. Kelly, "Motion and vision: II. Stabilized spatio-temporal threshold surface," J. Opt. Soc. Am. **69**, 1340–1349 (1979).
109. S. J. Anderson and D. C. Burr, "Spatial and temporal selectivity of the human motion detection system," Vision Res. **25**, 1147–1154 (1985).
110. G. Sperling, H.-J. Kim, and Z.-L. Lu. (2001). "Is there interocular first-order motion?" Talk at annual meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale, Fla., May 2, 2001.
111. Why did these observers in the interocular motion experiment overlook the first-order perceptual wind when they had previously very successfully reported the motion of the wind in pedestaled luminance-modulation motion and pedestaled texture-contrast motion?[41] Probably because of the context of the other trials. In pedestaled motion experiments, half of the stimuli have a pedestal, so observers are alerted to look for wind motion. In the interocular motion experiment, there was no pedestal. Most of the stimuli consist of a clearly moving grating. Only at the highest frequencies is there an illusory pedestal (which is created by motion standstill in the third-order system).[59] In this context, the observers attempted to interpret the output of the third-order motion system, and they overlooked the first-order wind.

112. H. J. Kim, Z.-L. Lu, and G. Sperling, "Rivalry motion versus depth motion," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **42**, S736 (2001).
113. C. Tseng, H. Kim, J. L. Gobell, Z.-L. Lu, and G. Sperling, "Motion standstill in rapidly moving stereoptic depth displays," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **42**, S504, Abstract nr. 2720 (2001).
114. Zemany *et al.*'s[71] use of 28 frames was stimulated by a misstatement by Lu and Sperling[41]: "... preserving pseudo-linearity requires exactly one full cycle plus one extra frame..." This is correct only for four-frame cycles. See the more general formulation in the text below.
115. B. A. Dosher, M. S. Landy, and G. Sperling, "Kinetic depth effect and optic flow: 1. 3D shape from Fourier motion," Vision Res. **29**, 1789–1813 (1989).
116. The MathWorks, Natick, Mass., 1997.
117. A. T. Smith and T. Ledgeway, "Sensitivity to second-order motion as a function of temporal frequency and eccentricity," Vision Res. **38**, 403–410 (1998).
118. Stimuli were created with HIPS image-processing software[119,120] and displayed by using a software package (Runtime Library for Psychology Experiments, 1988) designed to drive an AT-Vista video graphics adapter installed in an IBM 486 PC-compatible computer. Stimuli were presented on a 60-Hz vertical retrace IKEGAMI DM516A (20-inch diagonal) monochrome graphics monitor with a fast, white P4-type phosphor. A special circuit that combines two output channels produces 4096 distinct gray levels (12 bits). The luminance of the monitor was 12.1 cd/m^2 when every pixel was assigned the lowest gray level and 325 cd/m^2 when every pixel was given the greatest gray level. The background luminance was set at $0.5*(325 + 12.1) = 169$ cd/m^2. A lookup table was generated by means of a psychophysical procedure that linearly divided the whole luminance range into 256 gray levels. When extremely low contrasts were required by the experiment, a simpler lookup table was generated by linearly interpolating luminance levels around the background luminance (for contrasts less than 1%).
119. M. S. Landy, Y. Cohen, and G. Sperling, "HIPS: a Unix-based image processing system," Comput. Vis. Graph. Image Process. **25**, 331–347 (1984a).
120. M. S. Landy, Y. Cohen, and G. Sperling, "HIPS: image processing under UNIX software and applications," Behav. Res. Methods Instrum. **16**, 199–216 (1984b).
121. R. S. Woodworth and H. Schlosberg, *Experimental Psychology* (Rev. ed.). (Holt, Rinehart & Winston, New York, 1954).
122. H. R. Wilson, "Spatiotemporal characterization of a transient mechanism in the human visual system," Vision Res. **20**, 443–452 (1980).
123. M. B. Mandler and W. Makous, "A three channel model of temporal frequency perception," Vision Res. **24**, 1881–1887 (1984).
124. Z.-L. Lu and G. Sperling, "Sensitive calibration and measurement procedures based on the amplification principle in motion perception," Vision Res. **41**, 2355–2374 (2001).
125. Z.-L. Lu and G. Sperling, "Black-white asymmetry in visual perception," Mathematical Behavioral Sciences Rep. MBS 01-14 (University of California, Irvine, Calif., 2001).
126. *Apparatus*. The stimuli were presented on an achromatic 19″ Nanao FlexScan 6600 monitor, driven by the internal video card in a 7500/100 Power PC Macintosh at 120 frames/sec using a C++ version of VideoToolbox.[127] A special circuit was used to combine two 8-bit output channels of the video card to produce 6144 distinct voltage levels (12.6 bits). A psychophysical procedure was used to generate a linear lookup table that evenly divides the entire dynamic range of the monitor (from 1 cd/m^2 to 53 cd/m^2) into 256 levels.
127. D. G. Pelli and L. Zhang, "Accurate control of contrast on microcomputer displays," Vision Res. **31**, 1337–1350 (1991).
128. O. I. Ukkonen and A. M. Derrington, "Motion of contrast-modulated gratings is analyzed by different mechanisms at low and at high contrasts," Vision Res. **40**, 3359–3371 (2000).
129. "Feature tracking" is Ukkonen and Derrington's[128] term. However, a third-order motion system is required in order to do feature tracking. The terms are not synonymous. See Section 5.A.1.
130. F. C. Kolb and J. Braun, "Blindsight in normal observers," Nature **377**, 336–338 (1995).
131. P. Burt and G. Sperling, "Time, distance, and feature trade-offs in visual apparent motion," Psychol. Rev. **88**, 171–195 (1981).
132. A. Baloch, S. Grossberg, E. Mingolla, and C. A. M. Nogueira, "Neural model of first-order and second-order motion perception and magnocellular dynamics," J. Opt. Soc. Am. A **16**, 953–978 (1999).
133. M. J. Morgan and C. Chubb, "Contrast facilitation in motion detection: evidence for a Reichardt detector in human vision," Vision Res. **39**, 4217–4231 (1999).
134. J. Krauskopf and X. Li, "Effect of contrast on detection of motion of chromatic and luminance targets: retina-relative and object-relative movement," Vision Res. **39**, 3346–3350 (1999).