Pergamon

0042-6989(95)00025-9

*Vision Res.* Vol. 35, No. 19, pp. 2697–2722, 1995

0042-6989/95 $9.50 + 0.00

# The Functional Architecture of Human Visual Motion Perception

ZHONG-LIN LU,* GEORGE SPERLING*†

*Received 10 August 1994; in revised form 23 January 1995*

**A powerful paradigm (the pedestal-plus-test display) is combined with several subsidiary paradigms (interocular presentation, stimulus superpositions with varying phases, and attentional manipulations) to determine the functional architecture of visual motion perception: i.e. the nature of the various mechanisms of motion perception and their relations to each other. Three systems are isolated: a first-order system that uses a primitive motion energy computation to extract motion from moving luminance modulations; a second-order system that uses motion energy to extract motion from moving texture–contrast modulations; and a third-order system that tracks features. Pedestal displays exclude feature-tracking and thereby yield pure measures of the first- and second-order systems which are found to be exclusively monocular. Interocular displays exclude the first- and second-order systems and thereby to yield pure measures of feature-tracking. Results: both first- and second-order systems are fast (with temporal frequency cutoff at 12 Hz) and sensitive. Feature tracking operates interocularly almost as well as monocularly. It is slower (cutoff frequency is 3 Hz) and it requires much more stimulus contrast than the first- and second-order systems. Feature tracking is both bottom-up (it computes motion from luminance modulation, texture–contrast modulation, depth modulation, motion modulation, flicker modulation, and from other types of stimuli) and top-down—e.g. attentional instructions can determine the direction of perceived motion.**

Visual motion energy  Visual feature-tracking  Attention  Reichardt detectors  Visual second-order processing  Interocular displays  Visual channels  Visual bottom-up/top-down processing

## INTRODUCTION

The experimental study of human visual motion perception begins in the 19th century with Exner (1875). From the beginning, researchers have maintained that motion perception is "a primary sensation" in its own right (Exner, 1875; Wertheimer, 1912), because introspection seems to suggest that it invokes a unique perceptual experience quite different from other experiences. On the other hand, motion also seems to involve an early stage of pattern recognition, because the *same* pattern appears to be located first here and then there (Barlow, 1979).

Also, from the beginning, the explanation of motion has involved multiprocess theories. Early researchers defined different kinds of motion appearance in terms of Greek letters: alpha, phi, etc. (Wertheimer, 1912; Kenkel, 1913). Current dual-process and multi-process motion theories distinguish between short-range vs long-range motion (Braddick, 1974; Pantle & Picciano, 1976; Mather, Cavanagh & Anstis, 1985; Georgeson & Shackleton, 1989; Cavanagh, 1991); motion-energy and Reichardt detectors (van Santen & Sperling, 1984; Adelson & Bergen, 1985) vs zero crossings (Marr & Ullman, 1981) or gradients (Adelson & Bergen, 1986); first-order vs second-order motion (Cavanagh & Mather, 1989; Chubb & Sperling, 1989a), and so on. The problems with motion theories have been two-fold. On the one hand, there is considerable difficulty in adequately discriminating between the algorithm by which motion is computed and the preprocessing of the visual input prior to the motion computation. On the other hand, except perhaps for measurements of first-order motion with very low-contrast sine waves (e.g. Kelly, 1979; Burr & Ross, 1982), experimental isolation of proposed mechanisms has been problematic.

Here, several new approaches are described that, in combination, enable us to infer both the motion algorithms and the image preprocessing prior to motion computations. The procedures involve motion pedestal tests, interocular presentations, relative phase dependence tests, and selective attention instructions. A central concept to all of these is the *elaborated Reichardt detector* as a model for a motion detector. We build the discussion around the elaborated Reichardt detector because it is similar or computationally equivalent to

*Human Information Processing Laboratory, Department of Cognitive Sciences and Institute for Mathematical Behavioral Sciences, University of California, Irvine, CA 92717, U.S.A.
†To whom all correspondence should be addressed.

the other contending motion models, and it has the advantage of having properties that are easier to derive.

*Elaborated Reichardt detectors, motion-energy detectors, standard motion analysis*

Based on Reichardt's model for insect vision (Reichardt, 1957, 1961), van Santen and Sperling (1984) developed a computational theory of human visual motion perception based on *elaborated Reichardt detectors*. An elaborated Reichardt detector consists of two mirror-image subunits (e.g. "left" and "right") tuned to opposite directions of motion (Fig. 1). Subunit R multiplies the output of a spatiotemporal filter at spatial location A with the *delayed* output of another spatiotemporal filter at a rightward adjacent spatial location B. Subunit L multiplies signal (output from the spatiotemporal filter) at spatial location B with the *delayed* signal (output from the spatiotemporal filter) at spatial location A. The output of each subunit is integrated for a period of time and the direction of movement is indicated by the sign of the difference between the subunit outputs (Reichardt, 1957, 1961). The elaborated Reichardt *model*, consisting of a bank of elaborated Reichardt detectors tuned to various spatiotemporal frequencies and motion directions, predicts human performance by combining the outputs of many elaborated Reichardt detectors. van Santen and Sperling (1985) proved that two other motion theories (Watson & Ahumada, 1983; Adelson & Bergen, 1985) of motion perception were computationally equivalent to the elaborated Reichardt detector. Adelson and Bergen (1986) demonstrated the similarity of gradient detection to motion-energy detection. Even when the overall system performance is indiscriminable, different theories make different predictions about how computations might be carried out at the level of neural components (e.g. Emerson, Bergen & Adelson, 1992). However, our emphasis is on system performance. For the observations we make, it will not be necessary to distinguish between elaborated Reichardt detectors and motion-energy detectors. Previously, Chubb and Sperling (1989a) used the term *standard motion analysis* to refer to the detection mechanism whenever it was unnecessary to distinguish between elaborated Reichardt detectors and motion-energy detectors. However, we prefer here to use the terms *motion-energy detection* and *motion-energy analysis* to refer to an input–output computation that can be realized either as an elaborated Reichardt detector or as an Adelson–Bergen directional energy detector.

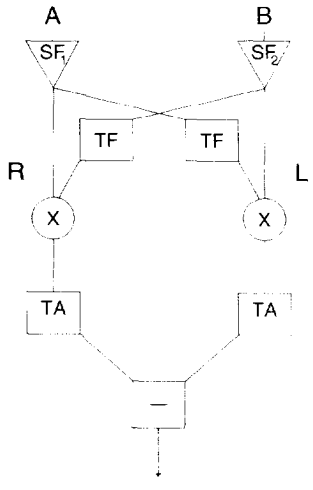


FIGURE 1. Elaborated Reichardt detector. It computes motion direction from two inputs that sample the visual display at two adjacent spatial locations A and B. $SF_1$ and $SF_2$ denotes linear spatial filters (receptive fields) that may be different. In the R ("right") subunit of the detector, the output of SF at B is delayed by the temporal delay filter TF and then multiplied with the direct output of SF at A. The output of the multiplier is temporally averaged over a temporal window (defined by the linear filter TA) to produce the final output of the "R" subunit. In the "L" subunit of the detector, the output of SF at A is delayed by the temporal delay filter TF and then multiplied with the direct output of SF at B. The final output of the "L" subunits is the temporal average of the output of the multiplier. The sign of the difference between the outputs of L and R subunits determines the perceived direction of motion. Outputs $>0$ indicate stimulus motion from B to A; outputs $<0$ indicate stimulus motion from A to B.

*Second- and first-order motion*

The term "second-order motion" was introduced in the literature (Cavanagh & Mather, 1989; Chubb & Sperling, 1989a) with two somewhat different connotations. We use it here to describe motion of broad classes of drift-balanced and microbalanced stimuli (Chubb & Sperling, 1988, 1989b, 1991; see also Ramachandran, Rau & Vidyasagar, 1973; Lelkens & Koenderink, 1984; Turano & Pantle, 1989; Victor & Conte, 1989; Derrington & Badcock, 1985) that are constructed out of drifting modulations of texture–contrast, spatial frequency, texture type, or flicker, and whose motion is not directly accessible to motion-energy analysis. Chubb and Sperling (1989a) proposed that some grossly nonlinear preprocessing (e.g. linear filtering followed by absolute value or square-law rectification) would expose the latent motion in all the above-named second-order stimuli to motion-energy analysis (Fig. 2). But, there has never been a direct test to establish that, after rectification, motion-energy analysis is indeed the ultimate mechanism for perceiving second-order motion. Experiment 2 (below) offers such a test.

*Two properties of the elaborated Reichardt detector (and motion-energy analysis)*

van Santen and Sperling (1984) proved several useful properties of elaborated Reichardt detectors (and equivalent motion-energy systems), two of which will be extremely useful here. (1) *Pseudo-linearity*: 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 the responses to individual inputs. (It is called *pseudo*-linearity because linearity holds only for sine inputs of *different* temporal frequencies.) (2) *Static-displays are ignored*: the output to any sinusoid of zero temporal frequency—a stationary pattern—is zero. From (1) and (2), it follows that adding a stationary sine (temporal frequency is zero and therefore output is zero) to any moving pattern (*moving* means temporal frequency is nonzero) does not change

the output of an elaborated Reichardt detector to the moving stimulus.

*The pedestal test*

*Pedestal immunity. Continuous displays.* Consider a *pedestaled motion stimulus*, i.e. a compound stimulus resulting from linear superposition of a drifting sine wave (the motion stimulus) and a stationary sine wave of the same spatial frequency (the pedestal). A corollary from the properties of pseudo-linearity and the ignoring of static displays is that the output of an elaborated Reichardt detector to a pedestaled stimulus is exactly the same to that of the motion component alone (Figs 3 and 4). If the elaborated Reichardt detector were the algorithm by which the human visual system computed motion direction, subjects' performance would be the same whether the motion stimulus were shown alone or pedestaled. In practice, nonlinearities of human vision before and after motion computation require that, for psychophysical tests, the combined amplitudes of component stimuli be small (e.g. less than about 5 percent modulation depth); within this range, the elaborated Reichardt detector properties are expected to hold exactly. If observers attempted to track the peaks (a kind of *feature tracking*) to discover motion, they would not be able to perceive coherent motion because the peaks merely oscillate back and forth [Figs 3(c, f) and 4(c, f)] without a consistent left–right direction. In fact, the back-and-forth oscillation is not symmetrical, and subjects might be able to learn to use this asymmetry to correctly make direction-of-motion judgments. That is one reason why feedback of the correctness of responses was not offered in these experiments (see also the section General Methods/Trials). Because the pedestal test defeats feature tracking, it offers a powerful way of discriminating between models of motion processing.

*Pedestal immunity. Sampled displays.* There is a caveat. Pseudo-linearity is a property of elaborated Reichardt detectors only insofar as the time constant of

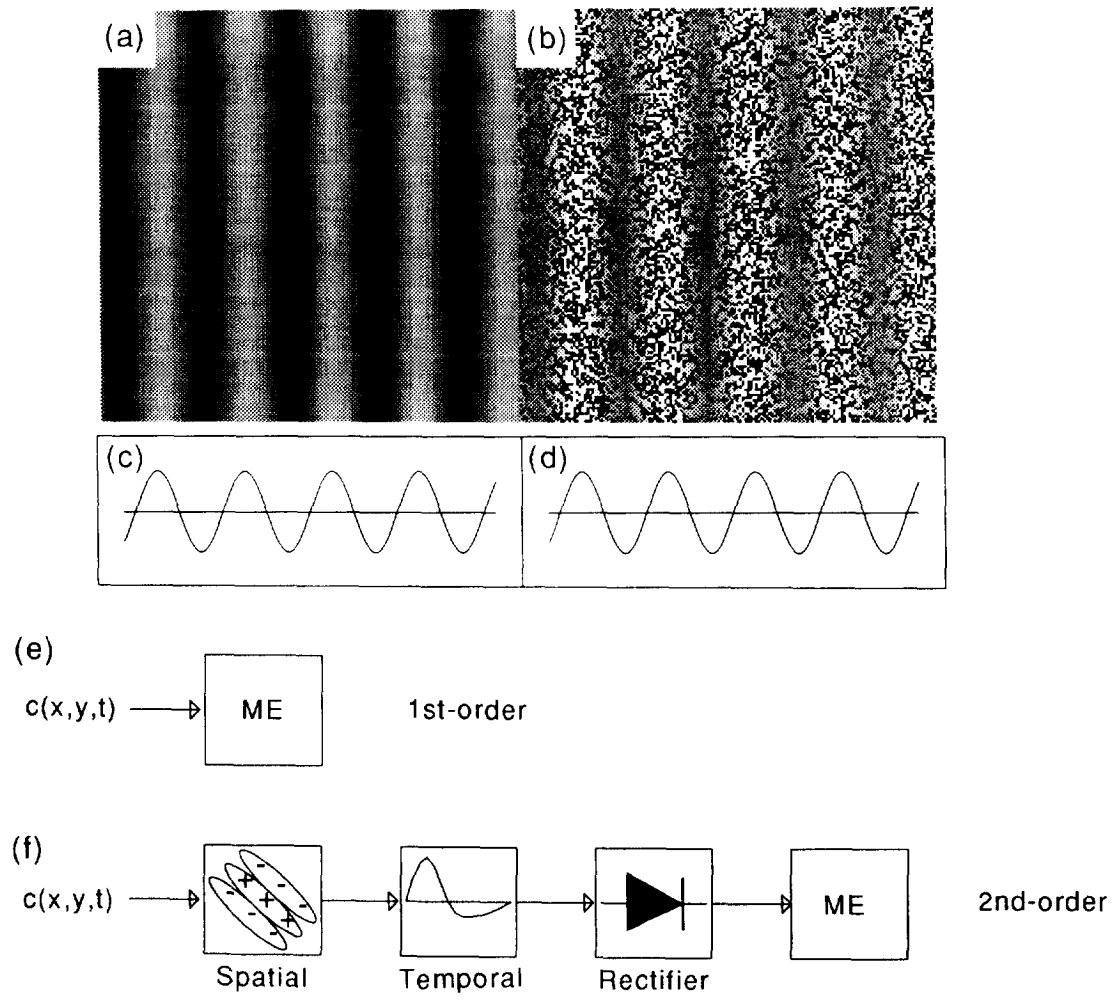


FIGURE 2. First-order and second-order motion stimuli and the mechanisms for detecting them. (a) One frame of a luminance modulation stimulus. (b) One frame of a texture contrast modulation stimulus. (c, d) The amplitude modulated waveforms of (a) and (b). (e) Motion of a luminance modulation stimulus [defined by its contrast $c(x, y, t)$] is extracted directly by a motion energy computation. This is "first-order motion" extraction. (f) A system that can extract motion from a texture–contrast modulated stimulus (b) and other driftbalanced and microbalanced stimuli, including texture quilts. The input signal $c(x, y, t)$ passes through a texture grabber—a spatial filter, a temporal bandpass filter, and a full-wave rectifier—and the texture-motion is then extracted by the motion-energy computation.

the output filter (TA in Fig. 1) is exactly the same as one stimulus cycle, or insofar as it is asymptotically long relative to a stimulus cycle (a property originally assumed by Reichardt, 1961). We can bypass this potential problem by presenting merely one cycle of the stimulus. When the stimulus is sampled in time (vs continuous in time), we can prove that preserving pseudolinearity requires exactly one full cycle plus one extra frame, so that the first and last frames are identical. (Obviously, as the sampling becomes finer, the extra frame becomes negligible, and the frame-enhanced cycle becomes asymptotically equivalent to merely a full cycle.) As long as the time constant TA (Fig. 1) is long enough to encompass this entire stimulus, it can be shown that the responses of a motion-energy mechanism to moving sinusoids and to pedestaled sinusoids are the same. That is, the computation is indifferent as to whether the restriction to a single cycle is caused by the internal time constant TA or by the restricted input stimulus.

In formal experiments, we determined how accurately subjects could perceive the direction of a motion stimulus in a standard pedestaled test. By *standard*, we mean that the amplitude ratio of pedestal:test was kept at 2:1. At this amplitude ratio, the summation of the pedestal and motion stimulus produces a sine wave that has a back-and-forth phase oscillation equal to one-sixth of the spatial cycle [Figs 3(c) and 4(c)]. In these tests, we first determined each subject's threshold amplitude for direction discrimination of the motion stimulus alone. A pedestal with twice the measured threshold amplitude was added and a subject's accuracy of motion-direction judgments was measured to determine whether or not it was influenced by the pedestal.

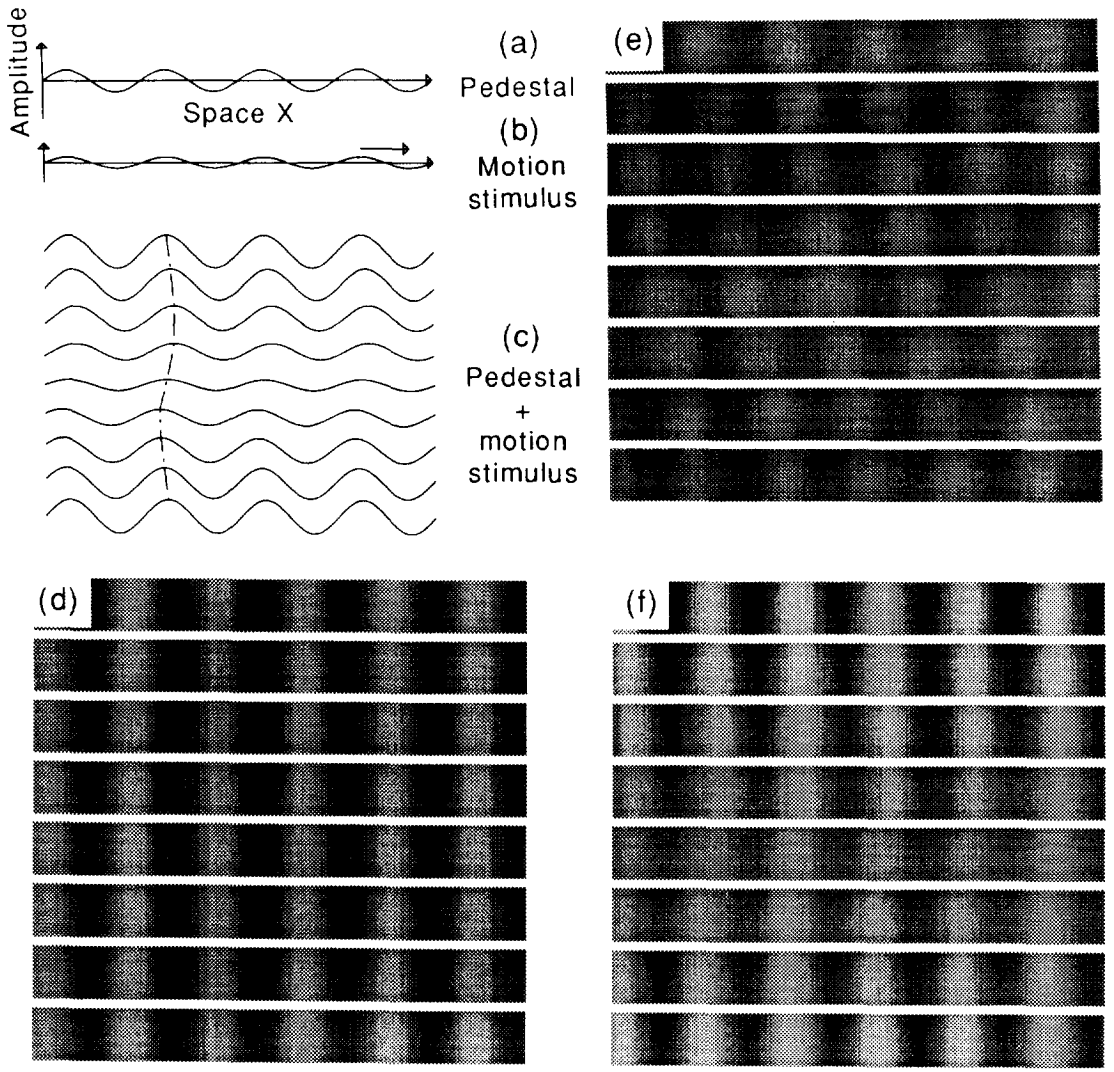


FIGURE 3. Pedestaled luminance modulation stimuli. (a) A stationary sine wave (the pedestal). (b) A moving sine wave (the motion stimulus) with half the amplitude of (a). (c) Eight frames of pedestaled motion: the sum of (a) and (b). From frame to frame, the motion stimulus moves one-eighth of a spatial cycle from left to right. The zigzag movement of a peak of the compound waveform is indicated by the dotted line. Any mechanism that computes motion from stimulus features such as peaks, valleys, or zero-crossings perceives only the zigzag motion. (d, e, f) Space–time plots: the vertical dimension is time, and the horizontal dimension is space. (d) Eight frames of (a), the stationary sine-wave pedestal, as displayed to the subjects. (e) Eight frames of (b), the drifting luminance modulation. Consecutive frames are shifted to the right by 45 deg. (f) Pedestaled motion: the sum of (d) and (e). The eight frames are those shown schematically in (c).

*Other paradigms*

*Interocular motion displays.* By the term *interocular motion display*, we mean a motion display in which the motion stimulus in each eye of an observer is ambiguous, yet perception of coherent motion is possible if the subject can combine information from both left and right eyes. Interocular displays answer the question: does the motion computation occur before or after the site of binocular combination?

*Relative phase dependence test.* Consider two physically independent channels. When a signal $s_1$ is carried only by one channel, and a second signal $s_2$ only by another, the output of the system to the superposition $s_1 + s_2$ of the two signals does not depend on the relative phase of the two signals. Conversely, when the output of the system to the superposition $s_1 + s_2$ does not depend on relative phase, we say they are carried in independent channels. (Strictly speaking, both phase and amplitude—in the extreme, presence vs absence—should be varied to determine channel independence. Because phase variations implicitly contain amplitude variations in the combined signal, a test of phase independence usually is sufficient.) A test of relative phase dependency offers a way to determine whether two stimuli activate the same or different channels. This principle has been widely applied in audition, and we use it here to determine the independence of various motion channels (cf. Graham, 1989).

*Selective attention manipulations.* Verbal instructions to the subject prior to a trial to selectively attend to one of the stimulus features in a complex stimulus influence a subject's perception of *what* appears to move in an

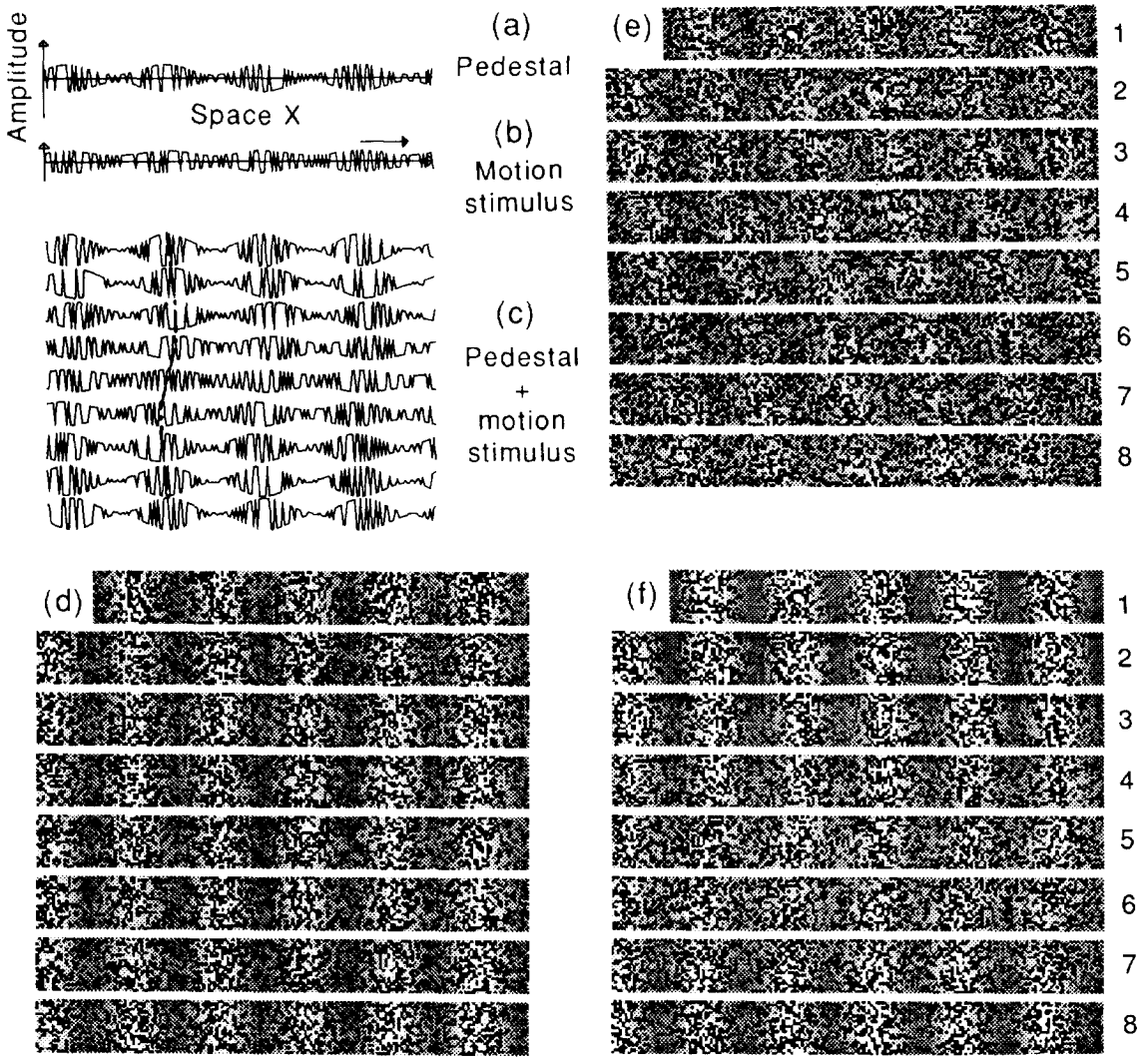


FIGURE 4. Pedestaled texture-contrast modulation stimuli. (a) The pedestal: a stationary sinusoidal modulation of texture-contrast. (b) The motion stimulus: a moving sinusoidal texture-contrast modulation with half the amplitude of (a). (c) Eight frames of pedestaled motion: the sum of (a) and (b). From frame to frame, the motion stimulus moves one-eighth of a spatial cycle from left to right. The zigzag movement of a peak of the compound waveform is indicated by the dotted line. Any mechanism that computes motion from stimulus features such as peaks, valleys, or zero-crossings perceives only the zigzag motion. (d, e, f) Space–time plots: the vertical dimension is time, and the horizontal dimension is space. (d) Eight frames of (a), the stationary sine-wave pedestal. (e) Eight frames of (b), the drifting texture-contrast modulation. Consecutive frames are shifted to the right by 45 deg. (f) Pedestaled motion: the sum of (d) and (e). The eight frames are those shown schematically in (c).

otherwise ambiguous motion display. This manipulation is considered in a companion paper (Sperling & Lu, 1995) but it is brought forward here because it is critical in demonstrating the top-down cognitive influence on feature tracking, the third-order motion process.

## GENERAL METHODS

Except where noted, all the experiments used the following methods.

*Basic motion displays: definitions*

*Stationary carriers $C_k$ and moving modulators $M$.* A stimulus is simply a function that gives the luminance $L(x, y, t)$ of each point in space $x$, $y$ as a function of time $t$. All motion stimuli considered here can be described as the product of a modulation function $1+m$ times a carrier $C_k$, $(1+m)C_k$. Four different kinds of spatial carriers were modulated to produce four kinds of moving modulations: luminance modulations, texture–contrast modulations, depth-modulations, and motion–motion modulations [Figs 3(d), 4(d), 5(b) and 6]. The modulation function is the same for all stimulus types; the subscript $k$ indicates the carrier type ($k$ = luminance, texture, depth, motion–motion).

*The moving modulator $M(x-t)$.* A moving modulation is a function of three variables $x$, $y$, $t$ that describe the direction of a movement in space-time, $M^{(3)}(x, y, t)$. However, all modulation movements considered here were linear movements in one dimension, either horizontal or vertical (usually horizontal), and the modulation was constant in the other dimension. Thus the modulation functions were effectively functions of a single variable $u$. For example, for horizontal movements, we have $m^{(3)}(\alpha u, \quad, \beta u)$, where the sign of $\beta/\alpha$ indicates the direction of movement, the magnitude of $\beta/\alpha$ indicates the speed, and — indicates an irrelevant variable. The simplified notation $M(\alpha x - \beta t) = M^{(3)}(\alpha x, y, -\beta t)$ suffices to describe the modulator; $\alpha$ and $\beta$ are omitted when the parameters of movement need not be indicated.

*The four kinds of moving-modulation stimuli.* A *luminance modulation* stimulus is simply $L(x, y, t) = [1 + M(x-t)]L_0$, where there carrier is a constant $L_0$, the mean luminance of the display, and the moving-modulator is $M(x-t)$.

The first component in the description of a texture–contrast modulation stimulus is the texture itself, given by $h(x, y)$. A moving texture–contrast-modulation stimulus is simply $[1 + M(x-t)]h(x, y)$. Similarly, in depth-modulation stimuli, the modulator $M$ multiplies the binocular disparity of a stereoptically defined depth grating. In motion modulation, the modulator $M$ defines the proportion of random dots that jump in a given direction at a location with spatial coordinates $x$, $y$ (in degrees of visual angle) at a time $t$ (in seconds).

Displays have the following properties. (1) The carrier is defined within a display window which is surrounded by a uniform background. The (expected) mean luminance is the same throughout the entire display. (2) The modulator $M(\alpha x - \beta t)$ is one dimensional and its motion is uniform linear translation. Motion is horizontal except in the case of dichoptic displays, where it is vertical. A horizontally moving modulator is denoted as $M(x-t)$, a vertically moving modulator as $M(y-t)$. (3) In all cases, $M(x-t)$ and $M(y-t)$ are sinusoidal functions of time. Let $x$ and $y$ be measured in degrees of visual angle (deg) and $t$ in sec. A horizontally moving sinusoidal modulation with spatial frequency $\alpha$ c/deg and temporal frequency $f$ Hz (c/sec) is

$$M(2\pi\alpha x - 2\pi f t) = m \sin(2\pi\alpha x - 2\pi f t). \qquad (1)$$

The magnitude of the modulator $|M|$ is $m$. $0 \leqslant m \leqslant 1$.

*Apparatus*

All the stimuli were created off-line using HIPS image-processing software (Landy, Cohen & Sperling, 1984a, b) and displayed using a software package (Runtime Library for Psychology Experiments, 1988) designed to drive an AT-Vista video graphics adapter installed in an IBM 486PC compatible computer. Stimuli were presented on a 60 Hz vertical retrace IKEGAMI DM516A (20 in. diagonal) monochrome graphics monitor with a fast, white P4-type phosphor. While many monitors have pixel interactions so that the intensity of an isolated pair of adjacent intensified horizontal pixels is different from a pair of adjacent vertical pixels, the IKEGAMI DM516A monitor has a sufficiently extended temporal frequency response to reduce such interactions to insignificance. 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. We chose the background luminance to be that value which, when it is assumed by every pixel, produces $0.5 \times (325 + 12.1) = 169$ 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

FIGURE 5 *(facing page).* Pedestaled depth-modulation stimuli as stereograms. (a) The pedestal. To see stereoptic depth, fuse the left and right images. Only a single temporal frame of the original dynamic stimulus is represented. (b) An $x$, $y$ representation of the moving $x$, $t$ depth modulation. The display conventions are similar to Figs 3(e) and 4(e). (c) An $x$, $y$ representation of the pedestaled depth-motion stimulus, the linear addition of depth-modulations (a) and (b).

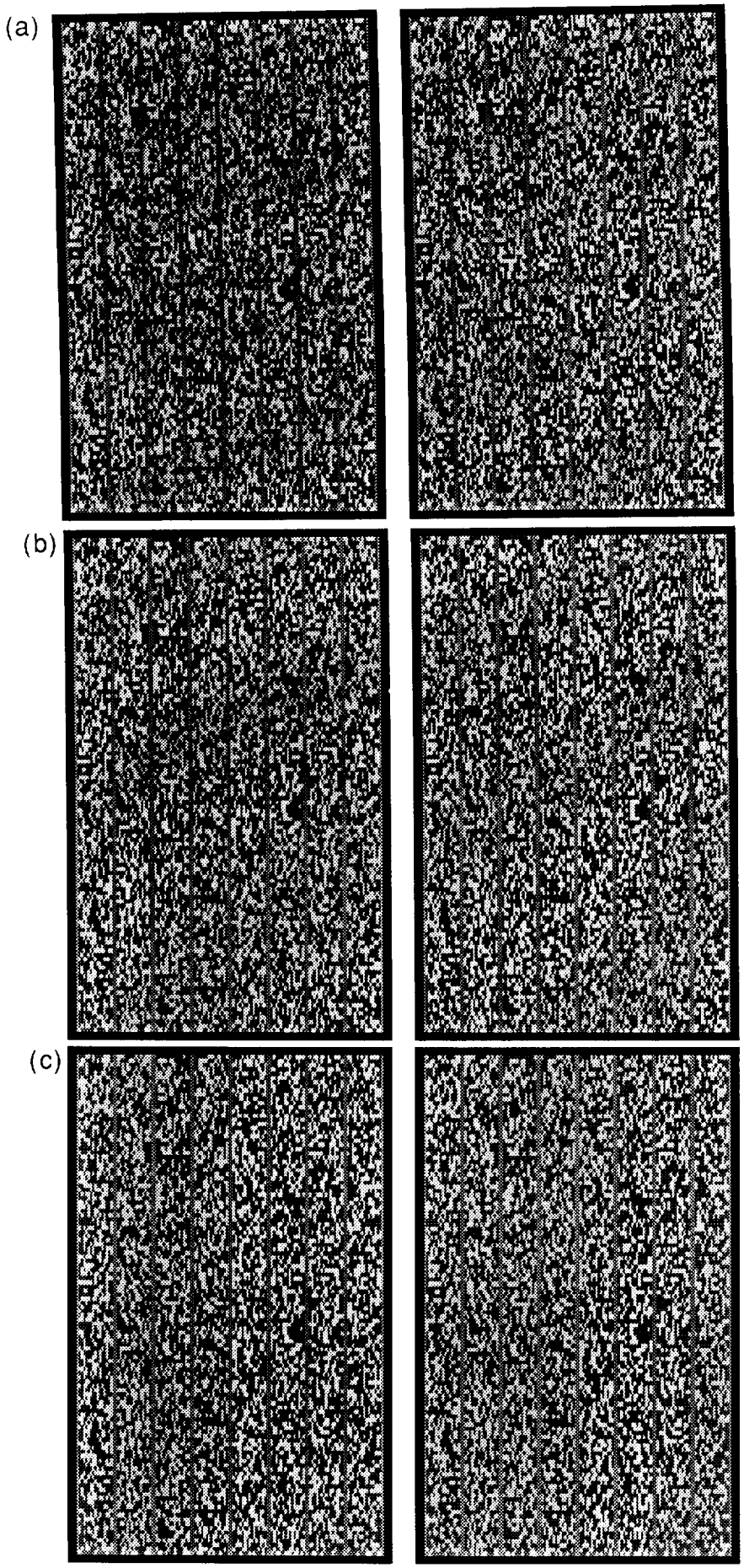


FIGURE 5. *Caption opposite.*

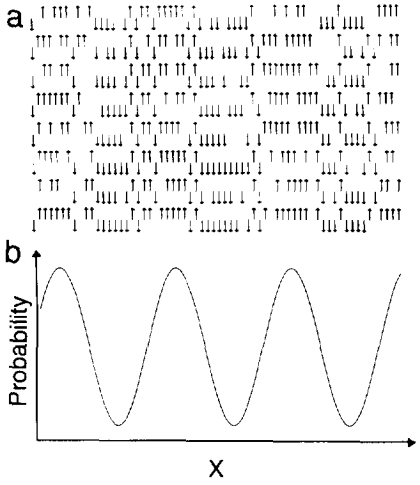


FIGURE 6. The motion-modulation stimulus. (a) Stationary sinusoidal motion modulation. The arrows indicate the directions of motion of random dots in the two-dimensional display. (b) The probability waveform that governs whether a dot at a particular horizontal location moves up or down. In a motion-modulation stimulus, the pattern of motion modulation (up vs down) moves either to the left or to the right in consecutive frames.

background luminance (for contrasts $<1\%$). All the displays were viewed binocularly with natural pupils in a dimly lighted room (the average luminance in the room is about 10 cd/m$^2$) except where noted.

*Trials*

The subject initiated every trial by pushing a button. A fixation point appeared at the center of the display which lasted 0.5 sec before the moving stimulus appeared and stayed on throughout the whole trial. 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 always identical to the first frame. In this way, we removed any positional cue on which subjects could base their judgments.

The subject's primary task was to judge direction of movement. The judgment was made by pushing one of two buttons. The percentage of "correct" (as *a priori* defined by the experimenter) judgments of motion direction was the main dependent variable of the experiments. Subjects also gave a confidence rating of their judgment. The confidence rating ranged from 0 to 5, with 0 meaning "totally uncertain" and 5 meaning "absolutely sure". Confidence judgments were used in pedestal experiments to determine whether the whole range of confidences, not merely the threshold, was the same in isolated and in pedestaled tests.

In the attention experiments, no feedback was given to the subject because we were primarily interested in how attentional instructions influenced perceived motion-direction, not in how feedback might influence button presses. Indeed, with the more exotic stimuli, it seemed important to demonstrate that subjects naturally and immediately perceived motion in the displayed motion direction. For the pedestal experiments, feedback might have enabled subjects to learn to interpret asymmetries in the feature back-and-forth zigzag to correctly infer motion direction. Therefore we decided not to use feedback in the experiments, and we interpreted the confidence judgments as direct indicators of perceived motion. In pilot studies, we did further investigate feedback vs no feedback. We found that, after initial "training" without feedback, subsequent feedback did not further improve subjects' performances for the non-pedestaled stimuli studied here, including the more complex stimuli used in phase-sensitivity tests (see however, Sperling, Landy, Dosher & Perkins, 1989).

Experimental sessions were blocked by different stimulus types (e.g. luminance modulation, texture–contrast modulation, depth modulation, motion modulation). For a given stimulus type, all temporal frequencies were mixed within a block. A block normally consisted of about 200 trials and lasted approx. 15 min. Intermissions between blocks were about 5 m. Subjects normally were given a 2 min dark adaptation period if they entered the test room from day light. A session lasted approx. 2 hr.

*Subjects*

A UCI graduate student (EB), naive to the purposes of the experiments, and the first author served as subjects in all the experiments. Both have corrected-to-normal vision.

## EXPERIMENTS

*Experiment 1. Temporal Frequency-sensitivity Functions*

The purpose of this experiment was to determine the modulation threshold as a function of temporal frequency for the four types of moving stimuli: luminance modulations, texture–contrast modulations, depth modulations, and motion modulations. Of particular interest were the highest frequencies for which accurate motion-direction judgments were possible.

*Stimuli*

To define a moving stimulus we make the following definitions: the horizontal spatial coordinate is $x$. The vertical spatial coordinate is $y$. The units of $x$ and $y$ are degrees of visual angle. The mean luminance is $L_0$ ($L_0 = 169$ cd/m$^2$). The spatial frequency of a translating modulation of stimulus type $s$ is $\alpha_s$, $f_j$ is the temporal frequency, $f_j = [0.94, 1.88, 3.75, 7.50, 15.0$ Hz$]$. $\beta = \pm 1$ indicates the direction of motion ($\beta = +1$ for leftward motion, and $\beta = -1$ for rightward motion); and $m(s, f_j)$ is the modulation depth of a stimulus of type $s$ and of temporal frequency $f_j$.

*(1) Luminance modulation [Fig. 3(e)].* The luminance-modulation stimulus is a rigidly translating sine-wave grating, the type of first-order stimulus from which much motion psychophysics has evolved. For moving luminance modulations: type $s = 1$; the spatial frequency

$\alpha_1 = 2.55$ c/deg; and the temporal frequencies $f_j$ are 0.94, 1.88, 3.75, 7.50 and 15.0 Hz

$$L(x, y, t, \beta, j) = L_0[1.0 + m(1, f_j)\sin(2\pi(\alpha_1 x + \beta f_j t))]. \quad (2)$$

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.

*(2) Texture–contrast modulation [Fig. 4(e)].* A texture–contrast modulation is the second type of stimulus ($s = 2$), and it is defined in terms of its carrier and modulation frequencies: $\alpha_2$ is the modulation frequency ($\alpha_2 = 1.28$ c/deg), and $R(x, y)$ is the carrier descriptor. $R(x, y)$ is a random variable that assumes values $+1$ and $-1$ with equal probability, independently at each spatial location $x, y$. $R(x, y)$ produces a uniform-amplitude carrier spectrum extending from 0.17 to 10.8 c/deg corresponding to wavelengths of 128–2 pixels. The temporal frequencies studied were the same as for luminance modulations: $f_j = [0.94, 1.88, 3.75, 7.50, 15.0$ Hz]. The texture–contrast modulation stimuli have the same size (3.13 × 1.57 deg as the luminance modulation stimuli.

$$C(x, y, t, \beta, j) = L_0[1.0 + R(x, y) \times (0.5 + m(2, f_j)\sin(2\pi(\alpha_2 x + \beta f_j t)))]. \quad (3)$$

The texture–contrast modulation is a pure second-order stimulus: its expected luminance is the same everywhere; its motion cannot be determined by motion-energy detection because the fundamental Fourier motion components are uninformative. However, a nonlinearity such as full-wave rectification (e.g. absolute value or squaring) could expose the motion of the texture–contrast modulation to motion-energy detection (Fig. 2).

*(3) Depth modulation [Fig. 5(b)].* A dynamic stereo depth-modulation is created from stereo views of left- and right-half images. It appears in depth as a surface whose apparent distance from the observer varies sinusoidally, as illustrated. The depth modulation exists only as a space-varying correlation between the pixels in the left- and right-eye images; each monocular image is completely homogeneous without any hint of a modulation, and successive frames are uncorrelated. Figure 5(a) illustrates a single frame of the depth-modulation stimulus.

The depth modulations were made of white (193 cd/m$^2$) random dots of dimensions 1.46 × 2.92 min arc on a gray background (153 cd/m$^2$). The spatial frequency of the modulation was $\alpha_3 = 1.28$ c/deg, and the temporal frequencies were $f_j = [0.94, 1.88, 3.75$ Hz], the useful range within which subjects were able to make motion-direction judgments. The motion of the modulator was sampled at 45 deg intervals (8 frames per cycle), and a new random-dot configuration was displayed after each movement.* The horizontal disparity between left and right eyes was:

$$D(x, y, t, \beta, j) = 1.46 \min \mathrm{Int}[(m(3, f_j)\sin(2\pi(\alpha_3 y + \beta f_j t))]. \quad (4)$$

Int( ) is a function (integer) to represent the fact that pixels are discrete and have a width of 1.46 min arc. Int( ) takes real numbers as input and rounds them to the nearest integer. The stimuli for each eye extended 5.88 deg vertically and 2.94 deg horizontally. In each frame, there was a 40% probability for a dot to be white; no correlation existed between successive frames. The left- and right-eye images were displayed adjacent to each other on the CRT, and a system of mirrors [a modified Helioth–Wheatstone stereoscope (Wheatstone, 1838; Dudley, 1951)] directed each image to the appropriate eye. To assist in producing good binocular fusion, central fixation points and surrounding black frames were provided in each eye's image. Subjects were instructed to begin a new trial only after they had achieved stable fusion. Figure 5(b) follows the conventions of Figs 3 and 4 to illustrate the moving depth modulation; Fig. 5(c) illustrates the pedestaled depth modulation.

*(4) Motion modulation ("motion–motion", Fig. 6).* The motion modulation consists of dots that make step jumps in successive frames (Zanker, 1993).† Within a column, all dots move in a consistent direction. The proportion of upward vs downward moving columns varies sinusoidally from left-to-right and there are 32 columns per cycle of the modulator. Because all columns have the same amount (although not the same direction) of movement, the expected amount of activity (i.e. dot changes) is the same everywhere. Therefore that the motion modulation could not be perceived by a mechanism that merely detected "activity" (e.g. Werkhoven, Sperling & Chubb, 1993). Perceiving the movement of the motion modulation requires (1) computing the direction of motion of the dots, and (2) noting that the sinusoidal pattern of dot-motion moves with time. The concept "motion of motion modulation" seems to suggest a hierarchical organization of motion detectors.

Like the depth stimuli, the motion modulation stimuli were composed of white dots (215.6 cd/m$^2$, 5.84 × 5.84 min arc) on gray background (169 cd/m$^2$). The probability that a given column would move up $M_u$ or down $M_d$ was modulated as a sine-wave function of $x, t$:

$$M_u(x, y, t, \beta, j) = 0.5 + m(4, f_j)\sin(2\pi(\alpha_4 x + \beta f_j t)) \quad (5a)$$

$$M_d(x, y, t) = 1.0 - M_u. \quad (5b)$$

*In such dynamic depth-motion stimuli, the rate of new images is 30 Hz at the highest temporal frequency ($f_j = 3.75$ Hz) and decreases four-fold for $f_j = 0.94$ Hz. A control experiment showed that there was no change in accuracy of motion-direction judgments when the rate of new images at $f_j = 0.94$ Hz was increased to 30 Hz. That is, the rate at which new instantiations of the random-dot stereograms occur is unimportant for motion-direction judgments within the range tested (7.5–30 Hz).

†This kind of motion–motion modulation was called *theta motion* by Zanker (1993). The original report confounded motion direction and "quantity of motion"; this is corrected in our stimuli and in a subsequent report (Zanker, 1994).

The spatial frequency of the motion modulation $\alpha_4 = 0.64$ c/deg; the temporal frequencies studied were $f_j = [0.94, 1.88, 3.75$ Hz]. There was a 40% probability for a given dot to be white, and a white dot moved 5.84 min arc (1 pixel) from one frame to the next. The displays extended 7.04 deg horizontally and 4.69 deg vertically.

*Procedure*

The temporal frequency characteristic is a graph of the smallest visible amplitude of modulation of a stimulus vs temporal frequency. The frequency characteristic is a kind of signature for dynamic systems. Where we find the same frequency characteristic, there is a presumption that similar or identical mechanisms are involved. The aim of Expt 1 is to determine the temporal frequency characteristic for each of the four types of stimuli.

Determining the frequency characteristic requires that a threshold be measured at each of the frequencies to be tested. To determine such a threshold, we use the method of constant stimuli (Woodworth & Scholsberg, 1954) to generate a psychometric function, and we designate the 75% correct point as the threshold. Psychometric functions consisting of five points were obtained for the four types of motion stimuli, for each of the temporal frequencies tested, and for each subject. At least 100 observations were made by each subject at every point on the psychometric functions.

For a given motion stimulus type $s$ and temporal frequency $f_j$, we defined subject's threshold as amplitude $m_{75}(s, f_j)$, corresponding to the 75% correct point on the psychometric function. The temporal sensitivity functions were generated by plotting $\log_{10}[1/m_{75}(s, f_j)]$ as a function of $\log_{10}(f_j)$ for different stimulus types.

*Results*

With above threshold stimuli, both our primary subjects and four other observers who viewed these motion displays reported vivid motion perception from all four types of motion stimuli at temporal frequencies from 0.94 to 3.75 Hz. Vivid motion perception for luminance and texture–contrast stimuli was possible up to the highest frequency tested, 15.0 Hz. While there were obvious differences in the appearance of the stimuli, introspection did not suggest any consistent differences of a qualitative nature between the motion percepts induced by the various stimulus types.

Figure 7 shows the temporal frequency response functions for all the motion stimulus types. The data for the two subjects are quite similar. Within our temporal frequency range (0.94–15.0 Hz), all the temporal frequency characteristics have typical low-pass filter shapes, i.e. sensitivity decreases monotonically with increasing temporal frequency. The temporal frequency characteristics fall naturally into two groups. The first group contains the luminance modulation and texture–contrast modulation (upper curves, Fig. 7). The second group consists of the dynamic stereo-depth modulation and motion modulation (lower curves, Fig. 7). Within each group, the shape of the temporal response characteristics are remarkably similar. Following common engineering practice, we define cutoff frequency as the frequency at which the sensitivity has dropped to one-half of the maximum sensitivity ($-0.3 \log_{10}$). The temporal sensitivity functions for luminance modulations

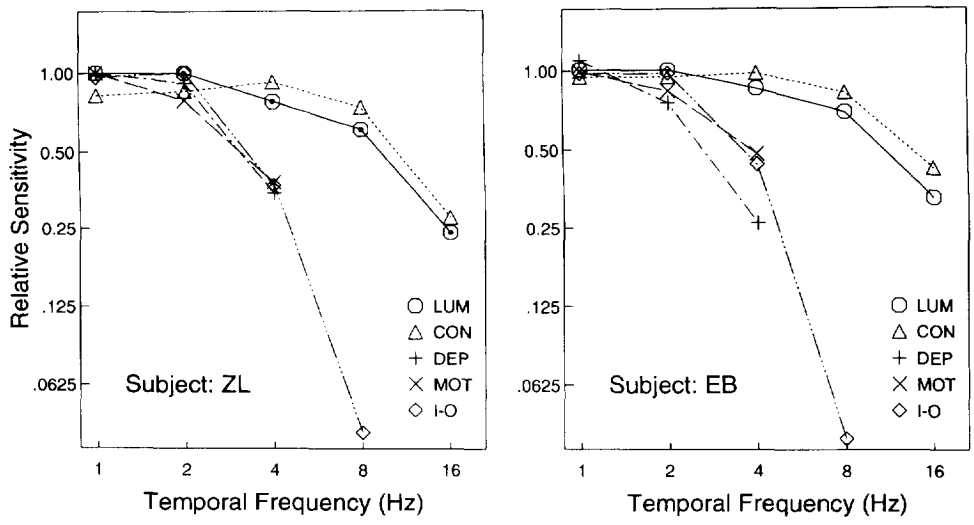


FIGURE 7. Temporal frequency characteristics. Ordinate: threshold amplitudes of modulations required for 75% correct left–right motion discrimination. Abscissa: temporal frequency of a moving sinusoidal modulation. The axes are logarithmic scales. Data are shown for two subjects. ○ Luminance-modulation motion (LUM) for pedestaled and for nonpedestaled stimuli; △ texture-contrast modulation motion (CON) for pedestaled and nonpedestaled stimuli (thresholds are identical); + nonpedestaled depth-modulation motion (DEP); × nonpedestaled motion-modulated motion (MOT); ◇ nonpedestaled interocular (I-O) luminance-modulation motion. The curves have been vertically translated to expose their similarity in shape. The scale value 1.00 on the ordinate represents the following modulation amplitudes for the two subjects ZL and EB respectively: LUM 0.0014, 0.0022; CON 0.027, 0.033; DEP 0.40 min arc, 0.40 min arc; MOT 0.11, 0.16; I-O 0.023, 0.018.

and texture–contrast modulations have a cutoff frequency of 12 Hz; the temporal sensitivity functions for depth and motion modulations have a cutoff frequency of 3 Hz.

*Discussion*

The luminance modulation data are quite similar in overall sensitivity and in frequency response to the previous studies of luminance modulation (e.g. Burr & Ross, 1982). The remarkable similarity of temporal sensitivity functions for texture contrast modulation (second-order stimulus) to the luminance modulation (first-order stimulus) was quite surprising. Derrington, Badcock and Henning (1993) had reported that the temporal frequency characteristic of the texture–contrast modulation (second-order) motion system was much "slower" than that of the luminance-modulation (first-order) motion system. However, our analysis of their stimulus indicates that there were two directions of first-order motion, each of which contained an implicit second-order component, plus their intended second-order motion. With five competing motion components simultaneously present in their stimuli, the detection situation is more complex than they envisioned, and it does not necessarily support their claim.

The clear bimodal grouping of the temporal response characteristics suggests there are two kinds of motion-detection systems, one a fast one that can detect drifting luminance modulation and texture contrast modulation, and a second slower kind that can detect drifting depth modulation and motion modulation. Indeed, if the different frequency characteristics were to derive from different motion-detection systems, then we would expect to find other properties that distinguish them. In Expts 2 and 3, we examine one such property for the four types of stimuli: the ability of motion perception to survive pedestals.

### *Experiment 2. Pedestal Experiments for Luminance Modulations*

The main aim of this experiment is to determine the resistance of motion perception to stationary pedestals. We also pursue several subsidiary aims: (1) to replicate van Santen and Sperling's original pedestal results (1984); (2) to explore early saturating nonlinearities of the first-order motion system; (3) to evaluate the effect of possible quantization errors in the stimulus presentation. In order to evaluate nonlinearities, we explore a range of pedestal amplitudes that includes very large pedestals. To evaluate quantization errors, which are equivalent to adding a particular kind of signal-dependent noise to the stimulus, we explore the effects of intentionally added noise—e.g. a "noise pedestal"—on subjects' ability to detect motion.

*Method*

The motion stimuli are described by equation (2). For each of the five different temporal frequencies ($f_t = 0.94$, 1.88, 3.75, 7.50, and 15.0 Hz), the amplitude of the motion stimulus was fixed around the value of the threshold estimated in Expt 1, $m_{75}(1, f_t)$. Five different static spatial patterns were added to these motion stimuli: (1) nothing; (2) a pedestal to the same amplitude as the motion stimulus, $m_{75}(1, f_t)$ (replicating van Santen & Sperling, 1984); (3) a pedestal twice its amplitude $2m_{75}(1, f_t)$, which is our standard pedestal test; (4) a pedestal 7 times its amplitude $7m_{75}(1, f_t)$; (5) stationary binary random white noise with amplitude equal to 7 times the test amplitude $7m_{75}(1, f_t)$. The added noise, in every pixel of the display, was equal to $\pm$the peak amplitude of the 7× pedestal. If there were a pointwise saturating nonlinearity, it would be expressed much more powerfully in the noise than in the 7× pedestal.

The starting phase of the motion stimulus was chosen randomly on each trial. At each temporal frequency, the amplitude of the motion stimulus was set to a value that would yield approximately 85–90% correct responses in the zero pedestal condition and pedestals were twice this value. (This was not possible for subject EB at 15 Hz because the pedestaled amplitude would have exceeded the available range.) For each temporal frequency, we measured the accuracy of motion direction judgments under all five stationary spatial pattern conditions. All the 25 conditions (temporal frequency × stationary pattern) were mixed with equal frequency in the experimental design. One-hundred trials were conducted for each condition for each subject.

*Results*

*Failure of 1× and 2× pedestals to impair motion-direction judgments.* The results of Expt 2 are the percent of correct motion-direction judgments in each of the five conditions; these are summarized in Fig. 8. Data of the two subjects are very similar. The main result is that the subjects' performance was approximately the same for four of the five conditions (including zero pedestal), and was significantly impaired for the 7× pedestal condition. The different heights of the curves that represent different pedestal frequencies in Fig. 8 are due only to our imperfect preselection of the amplitudes for the moving modulations.

(1) The 1× pedestal condition is a replication of van Santen and Sperling's (1984) procedure, and the results (no effect of pedestal) replicates their finding. In other words, the accuracy of motion direction judgments of drifting sine waves is unaffected by the addition of a stationary pedestal of *the same spatial frequency* and amplitude as the motion stimulus.

(2) Motion direction judgments with the standard 2× pedestal are no different from those with a zero or 1× pedestal. The immunity of motion-direction judgments to these pedestals verifies the strong prediction derived from motion-energy detection (e.g. Reichardt models, motion-energy models). The absence of any significant performance loss with the 2× pedestal for either subject over a wide range of temporal frequencies suggests that luminance-modulation motion perception is served *exclusively* by motion-energy detection mechanisms.

(3) The poor performance of subjects with the 7× pedestal suggests that first-order motion analysis

ultimately suffers from nonlinear saturation. Were the system completely linear in this range (pedestal contrasts around 3%), then a 7× pedestal would have 7 times the effect of 1× pedestal, or 3.5 times the effect of a 2× pedestal, both of which were zero. The fact the 7× pedestal produces a large effect means the system is nonlinear, and the most plausible nonlinearity to account for it is amplitude saturation. It is noteworthy that saturation occurs with amplitudes of only 3% contrast.

(4) The early nonlinear saturation in the luminance-modulation motion system is spatial frequency selective. If saturation were not spatial frequency specific, there would have been more saturation for the large-amplitude noise stimulus, which at every single pixel equaled or exceed the sinusoidal pedestal. In fact, there was no effect of the added noise.

(5) These results permit some estimates of the effects that might have been caused by imperfect intensity representations of the displays. For example, we infer that intensity quantization error in our inherently digital display has a negligible affect on performance. If quantization did have an effect, then adding the 7× noise pedestal, which has an order of magnitude higher amplitude than any conceivable quantization noise, would have an enormously greater effect. But the noise pedestal had essentially no effect on motion thresholds. Hence, the much smaller intensity quantization noise has no effect.

*Measurements of the visibility of the pedestals.* It might be argued that the failure of the pedestals to mask is that they are invisible or barely visible. To exclude this possibility we ran a control experiment in which we used a two-alternative forced-choice task to measure subjects' thresholds for detecting pedestals. The spatial frequency, temporal frequency, display duration and phase randomization were exactly the same as those used in the luminance-modulation pedestal experiments. The motion stimulus was removed and the amplitude of the pedestal was zero in one of the intervals. Pedestal amplitude in the other (signal) interval was varied from trial to trial.

In all cases, the thresholds for detecting the presence or absence of pedestals were somewhat smaller than those for judging motion directions. The ratio of the two thresholds (detecting presence/absence of stationary pedestals vs judging motion direction) decreases from 1:1.15 at a temporal frequency of 0.94 Hz to 1:1.8 at the highest temporal frequency, 15.0 Hz. We define the threshold in the detection experiment as the just noticeable difference (jnd) between two pedestals. By this measure, all the pedestals in our standard 2× pedestal tests were at least 2.3 jnds above their threshold of visibility, some were 3.6 jnds above; all were clearly visible. The masking effect (zero) of the 1× and 2× pedestals was independent of their number of jnds above the background.

*Conclusion.* The failure of 2× pedestals to exert any masking effect on motion direction thresholds is strong confirmation of the prediction from motion-energy analysis. Thus, the data are consistent with the hypothesis that the perception of drifting luminance

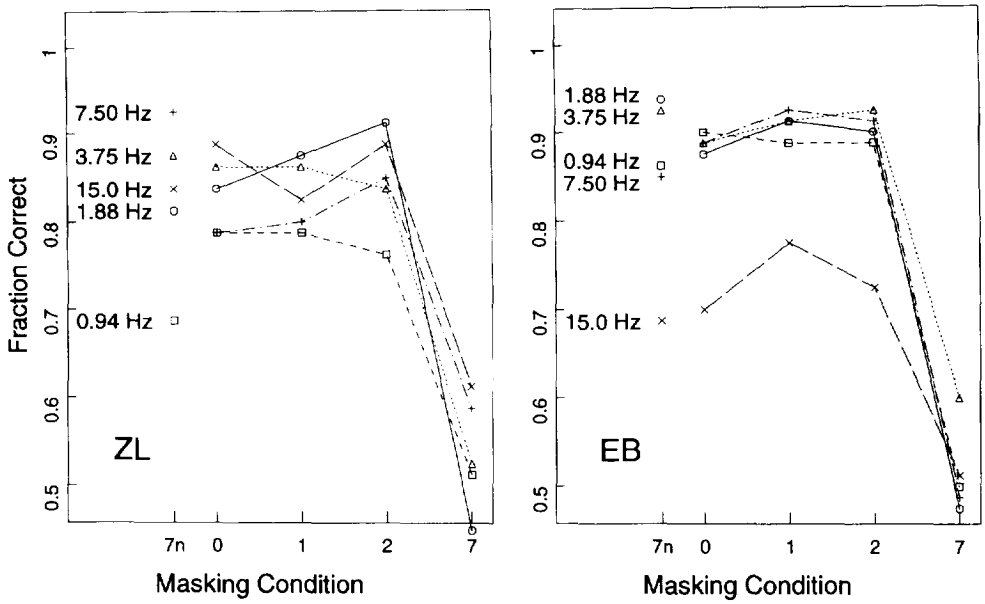


FIGURE 8. Pedestal masking of luminance-modulation motion stimuli for two subjects, ZL and EB. The ordinates denote percent correct in motion-direction discrimination (chance is 0.5). The abscissa denote the various masking conditions: "0", "1", "2", and "7" denotes the relative amplitude of the pedestal to the motion stimulus—0×, 1×, 2× and 7×. Specifically, 0 indicates no pedestal; "n" denotes added binary noise with 7× amplitude of the motion stimulus. Amplitudes of the sinusoidal motion stimuli for the five temporal frequencies were 0.94 Hz, 0.0016 and 0.0024; 1.88 Hz, 0.0016 and 0.0024; 3.75 Hz, 0.0020 and 0.0028; 7.50 Hz, 0.0024 and 0.0036; 15.0 Hz, 0.0060 and 0.0072, respectively, for subjects ZL and EB. Equivalent performance with 0×, 1× and 2× pedestals confirms a motion-energy (Reichardt) detection process. Good performance with 7 n binary noise coupled with poor performance with 7× sinusoidal pedestals indicates frequency specific saturating amplitude nonlinearity (contrast gain control) prior to motion processing.

modulations (first-order motion) is served by motion-energy detection.

### *Experiment 3. Pedestal Tests for Texture–Contrast, Depth, and Motion Modulations*

#### *Method*

In this experiment, the standard 2× pedestal test is applied to the three other kinds of motion stimuli. The motion stimuli are described in the Method section in Expt 1 [equations (3), (4) and (5)]. All the temporal frequencies described there were considered. For every stimulus type and temporal frequency, we fixed the amplitude of the motion stimulus at the nearest digital approximation to $m_{75}(s, f_j)$, where $s$ represents stimulus type and $j$ indexes different temporal frequencies. We observed subjects' performance in two stimulus conditions: (1) motion stimulus alone [Figs 4(e), 5(b) and (6)]; (2) motion stimulus plus a pedestal twice its amplitude [standard pedestal test, Figs 4(f) and 5(c)]. On each trial, the modulation amplitude of the pedestal was randomly set either at 0 or at $2 \times m_{75}(s, f_j)$. Within a session, stimuli of only one type were presented, but all temporal frequencies and pedestal amplitudes were randomly mixed. Thereby, data are available to compare every subject's performance at each temporal frequency, with and without the pedestal, all in the same sessions.

#### *Results*

*Texture–contrast modulation.* For the temporal frequencies tested, which ranged from 0.94 to 15.0 Hz, the presence of a 2× pedestal had absolutely no effect on subjects' performances (Table 1).

*Depth- and motion-modulation motion.* For the temporal frequencies tested, which ranged from 0.94 to 3.75 Hz, the presence of a 2× pedestal reduced subjects' performance to mere chance-guessing levels, although subjects' performance level remained at the previous 75% correct level without the pedestal. For depth- and motion-modulation stimuli with a 2× pedestal, subjects reported that they perceived only back-and-forth motion, and could not judge the direction of (apparently invisible) coherent motion.

*No threshold or saturation artifacts.* One might argue that the subjects' inability to perform the depth- and motion-modulation tasks with pedestals is due to an early compressive nonlinearity in the motion extraction system, similar to subjects' inability to perform with the 7× luminance pedestal. Two further brief experiments were conducted to determine the possible role of saturation. The first procedure used motion-modulation stimuli without pedestals. As the amplitude of the up–down modulation was increased to levels exceeding 3 times threshold, accuracy of motion-direction judgments continued to increase within the available range. This argues against a completely compressive saturation.

*Coarse quantization of low-amplitude stimuli?* In a second procedure, the amplitude of the entire 2:1 pedestal-plus-test stimulus was increased. If there were very poor discrimination of modulation depth at threshold, this procedure would move the modulations into a midrange. It did not yield improved performance, however, suggesting that the difficulty with pedestals is intrinsic to the pedestal manipulation, not to the modulation range in which it is tested.

For drifting texture–contrast modulations, the fact that subjects' performances were completely unaffected by the presence of a 2× pedestal supports the hypothesis that their motion is detected by motion-energy detection. Pedestals remove trackable features, and these are not needed to judge the motion of the texture–contrast modulations. On the other hand, the inability of subjects to judge motion direction of pedestaled depth and pedestaled motion modulations suggests that motion extraction from these stimuli does indeed depend on feature tracking.

### *Conclusions from Expts 1, 2 and 3*

The results in Expts 1, 2 and 3 clearly indicate that there are two qualitatively different kinds of motion extraction processes. One process computes motion energy and has a cutoff of 12 Hz in its temporal

TABLE 1. Pedestal test results: accuracy of motion-direction judgments with and without pedestals

| | | Subject | | | |
|---|---|---|---|---|---|
| | Temporal frequency | ZL | | EB | |
| Stimulus type | (Hz) | 0× pedestal | 2× pedestal | 0× pedestal | 2× pedestal |
| Contrast modulation | 0.94 | 0.77 | 0.81 | 0.72 | 0.72 |
| | 1.88 | 0.73 | 0.75 | 0.77 | 0.78 |
| | 3.75 | 0.73 | 0.75 | 0.73 | 0.74 |
| | 7.50 | 0.75 | 0.76 | 0.74 | 0.70 |
| | 15.0 | 0.79 | 0.78 | 0.88 | 0.87 |
| Depth modulation | 1.88 | 0.82 | 0.51 | 0.78 | 0.52 |
| Motion modulation | 0.94 | 0.68 | 0.52 | 0.66 | 0.51 |
| | 1.88 | 0.80 | 0.47 | 0.73 | 0.48 |

Chance performance level is 0.50. Each data entry is based on 100 observations. 2× indicates twice the threshold modulation amplitude; 0× indicates no pedestal.

$t$-Test: $P \leqslant 0.0011$, 2× vs 0× pedestal condition in depth and motion modulation.

$t$-Test: $P \geqslant 0.55$, 2× pedestal vs 0.50 in depth and motion modulation.

$t$-Test: $P \geqslant 0.33$, 2× vs 0× pedestal condition in contrast-modulation condition.

sensitivity characteristics. This kind of process subserves both luminance modulation motion (first-order) and texture–contrast modulation motion (second-order) stimuli. The second kind of process utilizes stimulus feature tracking. It is slower (cutoff frequency of 3 Hz), but it can detect kinds of motion that are invisible to the first process.

### *Experiment 4. Relative Phase Dependence Test of First- and Second-order Motion System*

Experiments 2 and 3 showed that motion perception of both luminance modulations and texture–contrast modulations utilizes a motion energy computation. We now ask: are the luminance modulation (first-order) and texture–contrast modulation (second-order) motion computations carried out in the same or in separate motion-energy channels?

*Motion transparency test.* There is a considerable literature dealing with methods for demonstrating channel independence (e.g. Graham, 1989). In the case of motion perception, the *motion transparency test* may provide the best approach. By definition, when two independent motion channels carry motion signals, there should be transparency—both motions should be simultaneously visible. For example, superimposing oppositely-directed equal-strength luminance and texture contrast modulation stimuli should produce motion transparency, not cancellation, if the motion were extracted in independent channels.

Solomon and Sperling (1995) systematically studied interactions between superimposed, oppositely moving, Fourier and full-wave stimuli. They found that: (1) there was a significant range of contrasts where their subjects could correctly report the direction of either stimulus (i.e. there was motion transparency); (2) for stimuli near threshold but sufficiently intense to be clearly visible, there was relatively little cross masking; (3) subjects could attend simultaneously to motions of both luminance modulations and texture contrast modulations and report both directions without any loss (relative to reporting the motion direction of only one stimulus). Solomon and Sperling (1995) noted that, while their paradigm demonstrates the existence of two independent channels, it does not identify the channels. Furthermore, their stimuli were sufficiently above threshold that they might have involved the feature tracking system. Because Solomon and Sperling (1995) had no manipulation to restrict motion analysis to a motion energy computation, the two motion-direction judgments could have been based entirely on a motion energy computation or on feature tracking, or on a combination of both.

*Pedestaled motion transparency test (opposite directions of motion).* In order to demonstrate motion energy channels in super-threshold stimuli, it is necessary for the component stimuli to be presented on pedestals because that is, so far, the secure method of eliminating feature tracking. In our *pedestaled motion transparency test*, we superimpose oppositely-directed equal-strength luminance modulation (first-order) and texture–contrast modulation (second-order) stimuli, each on its own pedestal. These stimuli are sufficiently intense that we might expect to see transparency. Whenever motion transparency is observed, we conclude that the luminance modulation and texture–contrast modulation motion are computed in separate motion channels and the outputs are fed into separate higher level units. However, when we fail to observe motion transparency, we cannot distinguish two possibilities: (1) motion directions of the luminance modulations and texture–contrast modulations are computed in separate motion channels but the outputs are combined in a higher level unit. (2) Motion directions of the luminance modulations and texture–contrast modulations are computed in the same motion channel.

*Relative phase dependence test (same directions of motion).* As will be detailed below, we failed to observe motion transparency in the *pedestaled motion transparency test*. When we added opposite-directed equal-strength pedestaled luminance modulations and texture contrast modulations, they simply cancelled. To distinguish between the same-channel and different-channel origin of this transparency failure, we use a *relative phase dependence test*. In this test, a luminance modulation and a texture–contrast modulation are presented together drifting in the *same* direction. In one case, both modulations have the same spatial frequency and temporal frequency [Fig. 9(a, b)]; in the other case, the parameters of the texture–contrast modulation remain the same, but the spatial and the temporal frequency of the luminance modulation are only half that of the first case [Fig. 9(c, d)].

The reason for halving the spatial frequency of the luminance modulation stimulus is that, if this stimulus were to be full-wave rectified by the (second-order) texture–contrast modulation system [Fig. 9(c)], it would appear to have doubled its frequency. Halving the spatial frequency, *a priori*, produces a full-wave fundamental component of the luminance-modulation stimulus and a texture modulation with precisely the same spatial and temporal frequencies [Fig. 9(c, d)].

Because the stimuli in this experiment are near threshold for the motion-energy system, it is not necessary to use pedestals—as will be shown in Expts 6 and 7, they are more than 10 × below threshold for the feature-tracking system. The percent of correct direction-of-movement judgments (of both modulations together) is determined for eight relative spatial phases (every 45 deg defined regarding to the luminance modulation). It is critical that in both cases, the luminance modulation and texture–contrast modulation move in the *same* direction. If the two kinds of stimuli were combined prior to the motion computation, the *magnitude* of the combination would depend on the relative phase of the components. Indeed, stimuli of the same frequency moving in the same direction but with opposite phases could add up to a constant and thereby cancel all perceived motion, which depends on the maximum minus the minimum of the waveform (in order to compute a correlation). This

predicted effect of phase is shown schematically in the graphs of Fig. 9.

The absence of any phase dependence would mean that luminance modulation and texture–contrast modulation motion strengths are first computed separately; then, the two motion strengths are combined [Fig. 9(e, f)].

*Method: pedestaled motion transparency*

*Stimuli.* We superimposed oppositely-moving luminance modulations and texture contrast modulations by putting them in alternative pixel rows of our display. Although our display was calibrated for linearity, this extra precaution insures absolutely linear addition of the two modulations. The mathematical descriptions of the stimuli can be found in equations (2) and (3).

At a viewing distance of 1.20 m the height of a pixel row was 1.466 min arc. Both modulations had the same spatial frequency of 2.4 c/deg, and the whole display extended 7.04 deg horizontally and 4.69 deg vertically. The amplitudes of both the luminance modulation and texture–contrast modulation are 2–3 times above their threshold amplitude. A luminance pedestal with twice the amplitude of the luminance modulation motion, and a texture–contrast pedestal with twice the amplitude of the texture–contrast modulation motion were added to the display. The relative phase of the two pedestals was randomized from trial-to-trial. The temporal frequencies of both motion modulations were the same within a trial and varied, from trial-to-trial, from 0.94 to 15 Hz.

*Motion cancellation procedure.* The aim of the experiment is to determine whether there is a point where an oppositely moving luminance modulation and a texture–contrast modulation perfectly cancel. The sensitivity of the second-order system declines with increasing spatial frequency much more quickly than for the first-order system (in the range under study). Therefore, a simple way to modulate the relative strength of the (second-order) texture–contrast motion system is for the observer to move closer to or further away from the display. The subjects were instructed to adjust their distance from the display and to report whether they could find a distance that provided perfect motion cancellation.

*Control procedure to test for masking.* Once a subject had found a distance that provided perfect motion cancellation, we wanted to verify that this was indeed motion cancellation and not a form of nonmotion masking of one stimulus by the other. To discriminate motion cancellation from nonmotion masking, we substituted for one of the moving stimuli a counterphase grating (of the same amplitude and spatial and temporal frequency) and determined whether the motion of the other (moving) stimulus was visible. In this procedure,

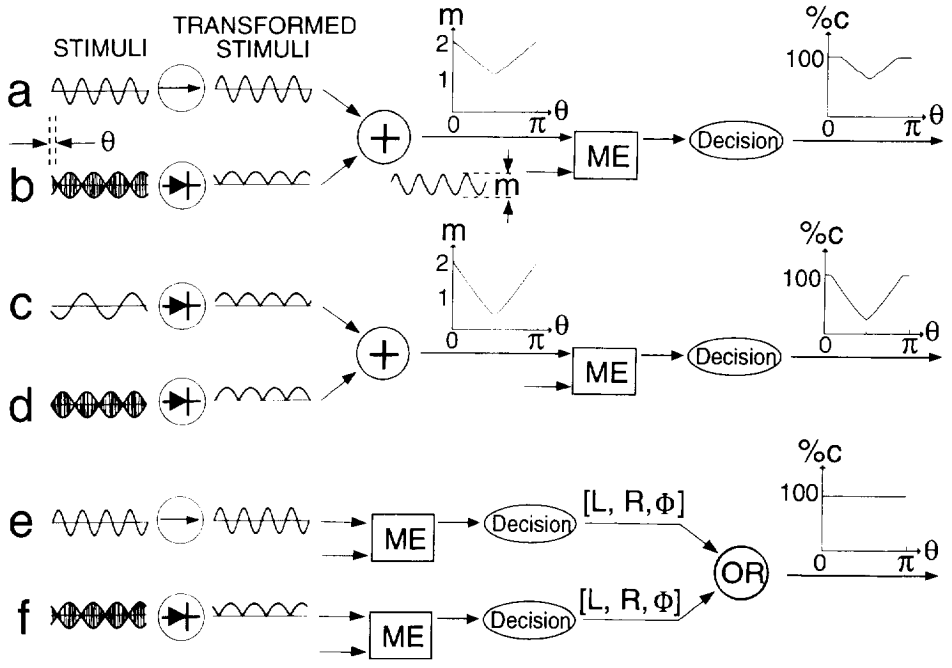


FIGURE 9. Phase dependence: stimuli, models, and predictions. (a) A drifting luminance sine wave and (b) a drifting texture contrast modulation in a first-order channel. The arrow indicates transformation of (a) by a linear filter. The → indicates full-wave rectification of (b). + indicates linear addition early within or prior to a motion-energy detector (ME). The lower input arrow to ME indicates a spatially displaced second input of ME (e.g. Fig. 1, "B"). $\theta$ is the phase difference; $m$ indicates the maximum minus the minimum of the output after +. The graph shows $m$ vs $\theta$. The graph at the far right illustrates the percent of correct motion direction judgments vs $\theta$. (c) A luminance sine wave with half the frequency of the texture modulation (d). The luminance modulation is rectified, motion energy is computed from the combined luminance and contrast signal (second-order computation). (e, f) Independent processing of luminance (e) and contrast (f) modulations and probability summation. The decision component signals either L (leftward motion) or R (rightward) if motion detection occurs, otherwise it outputs a null $\phi$. A motion judgment (L, R, $\phi$) is output by the logical OR component. When the OR component outputs a null, a random guess (L or R) is generated (not shown).

both pedestals were retained unchanged, and the subjects remained at the same locations where they had found perfect motion cancellation.

On the whole, the counterphase grating has the same physical parameters as the motion stimulus, and we would expect it to have the same masking power. If the absence of apparent motion were due to nonmotion masking, we would expect the subjects to also report absence of motion in the counterphase control experiment. If the absence of motion were due to motion cancellation, we would expect the subjects to perceive the motion of the residual motion stimulus almost as well as if it were presented alone. (We say "almost" because there might be some reduction in visibility due to nonmotion masking.) Four subjects served in this experiment.

*Method: phase dependence*

*Two alternative hypotheses.* Two different relative spatio-temporal frequencies are used to test for two possibilities. (1) Only a luminance-modulation channel: motion of the texture–contrast modulation is extracted by the same channel that extracts motion of luminance modulations [Fig. 9(a, b)]. (2) Only a contrast-modulation channel: motion of the luminance modulation is extracted by the same channel that extracts motion of texture–contrast modulations [Fig. 9(c, d)].

*Stimuli: test for exclusive luminance-modulation motion channel.* To test for the possibility of a luminance-modulation channel only, we used texture contrast modulations and luminance modulations that had the same spatial frequency and the same temporal frequence [Fig. 9(a, b)]. The texture–contrast modulation stimulus can be described as:

$$L_s(x, y, t, \beta, p) = L_0\{1.0 + R(x, y)m(5, f)\sin[2\pi(\alpha_s x + \beta f t) + p\pi/8]\}. \quad (6)$$

In equation (6), the mean luminance $L_0 = 169$ cd m$^{-2}$; the spatial frequency $\alpha_s = 1.2$ c/deg; the temporal frequency $f = 1.5$ Hz; the relative phase is indexed by $p = [0, 1, 2, 3, 4, 5, 6, 7]$; the direction of motion is determined by $\beta = +1$ or $-1$; and the carrier texture is described by $R(x, y)$ which is a random variable of only spatial location $(x, y)$ and which assumes value $+1$ and $-1$ with equal probability. Full-wave rectification of the texture contrast modulation would produce a stimulus with a temporal frequency of 3 Hz temporal and a spatial frequency of 2.4 c/deg. The stimulus extends 3.13 deg horizontally and 1.57 deg vertically.

The luminance modulation stimulus is described by equation (2) with spatial frequency $\alpha_l = 2.4$ c/deg and temporal frequency $f = 3.0$ Hz.

*Stimuli: test for an exclusive channel for motion of texture-contrast modulations.* To test for the possibility that a single texture contrast modulation channel analyses both luminance and texture–contrast modulations, we used texture contrast modulations and luminance modulations with relative spatial and temporal frequency ratio 2:1 [Fig. 9(c, d)]. The full-wave-rectified component of the luminance modulation has the same temporal and spatial frequency as the texture–contrast modulation, and is designed to optimally stimulate the same motion channel as the texture–contrast modulation.

Specifically, the texture–contrast modulation can be described using equation (6) with $L_0 = 169$ cd/m$^2$, $\alpha_s = 1.2$ c/deg, $f = 1.5$ Hz, $p = 0, 1, 2, 3, 4, 5, 6, 7$, $\beta = +1$ or $-1$ depending on the motion direction. And the luminance modulation can be described using equation (2) with $\alpha_l = 1.2$ c/deg, $f = 1.5$ Hz.

In both cases, the stimuli were shown for a duration that equals a full temporal cycle plus one frame of the relevant luminance modulation.

*Procedure.* The relative phase dependence test is designed to study system interactions near threshold. This requires measuring threshold for all stimuli prior to their combination. The method of constant stimuli was used to generate psychometric functions for this purpose. Because, in the subsequent super-position tests, the luminance modulation and the contrast modulation occupied only alternating pixels rows (at width 1.466 min arc), we interleaved blank rows (at background luminance levels) in the stimuli while measuring psychometric functions.

In subsequent phase dependence tests, we measured the dependence of subjects' performance on the relative phase of the luminance modulation and texture–contrast modulation. The two moving modulations were superimposed spatially in alternating pixel rows in one of eight phase relations. Modulation amplitudes were held at their previously measured thresholds. Ten stimulus conditions occurred randomly within a block: luminance modulation at its threshold, texture–contrast modulation at its threshold, super-position of luminance and texture–contrast modulations at their respective thresholds in eight different relative spatial phases.

*Results: pedestaled motion transparency*

For each temporal frequency tested, each subject was able to find a viewing distance at which there was perfect motion cancellation of the superimposed, pedestaled, oppositely-directed luminance and contrast modulations. The distances at which motion cancellation was observed were relatively similar for different subjects and temporal frequencies.

In control sessions, when one motion stimulus was changed to equal amplitude counterphase flicker, all subjects easily perceived the motion of the other motion stimulus. The results of the control experiment demonstrates that the motion cancellation is not due to some kind of nonmotion masking.

Perfect motion cancellation indicates that luminance motion and texture–contrast ultimately share a common pathway. But with these results alone, we cannot determine the level of this interaction.

*Results: phase dependence*

*Test for exclusive luminance-modulation motion channel.* The results for luminance modulations and

texture–contrast modulations of the same spatio-temporal frequencies are summarized in Fig. 10(a, b). (This is the case where, if there were only one channel, it would be a "luminance-modulation channel".) It is obvious that there is no motion cancellation nor any appreciable phase dependence. There is no interaction between the modulations; at least two channels are required to account for these data.

The performance at any or all of the relative spatial phases can be well predicted by a simple probability summation model [Fig. 9(e, f)]. The model assumes that each type of modulation (luminance, texture–contrast) is processed by a corresponding type of channel. For each of the two channels, there are two possible outcomes on a trial with compound modulations. Either there is a "true" detection of the motion directed at that channel, or there is not. If a true detection occurs in either channel, the subject reports that as the direction of motion of the compound stimulus. If neither channel has a true detection, the subject guesses a random direction.

To formalize the model, let the $P_j^t$ represent the probability of a true motion detection in channel $j$; $j$ has the values lum for the luminance modulation channel, con for the texture–contrast modulation channel, and U (union) for a detection in either or both channels. Ultimately, $P_{\text{lum}}^t$ and $P_{\text{con}}^t$ are estimated from single-modulation presentations, and used to predict performance in dual modulation presentations.

Let the $P_j^0$ represent the experimentally observed probability of a correct motion response for a stimulus of type $j$. The only way to fail to make a "true" detection with a compound stimulus is to fail on both channels:

$$1 - P_{\text{U}}^t = (1 - P_{\text{lum}}^t) \times (1 - P_{\text{con}}^t). \tag{7}$$

Given random guessing in the absence of a "true" detection, the "true" percentage correct $P^t$ is estimated

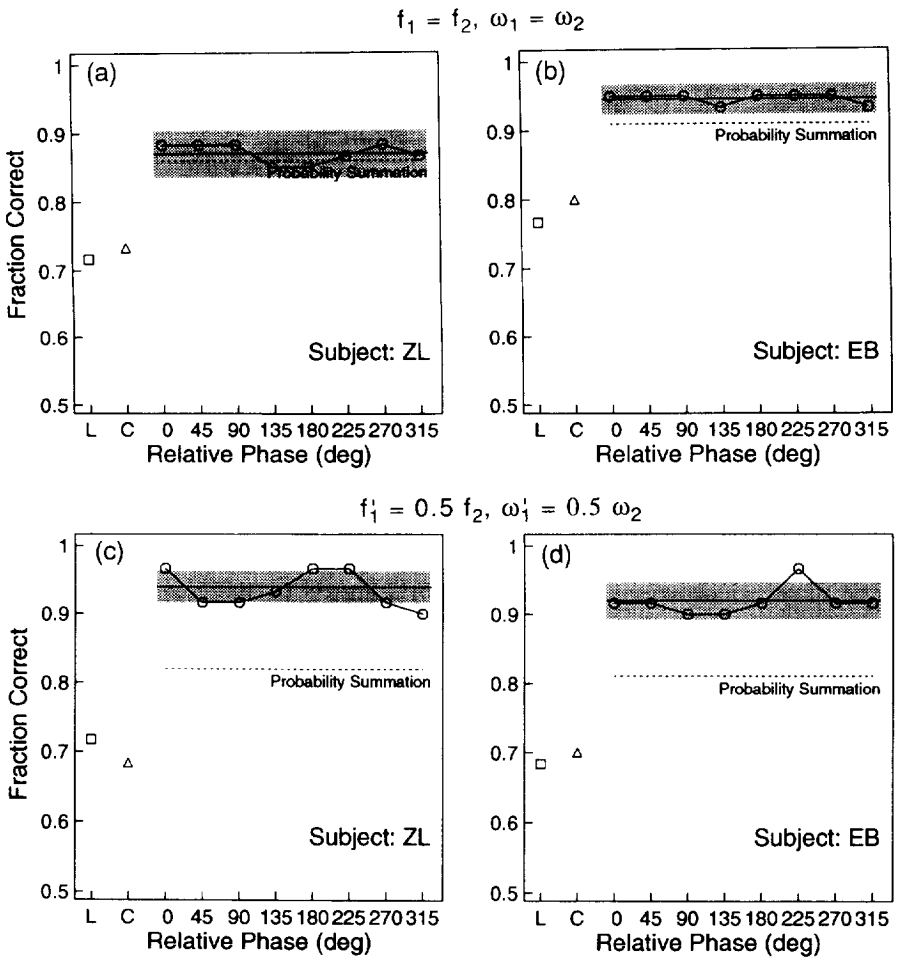


FIGURE 10. Results of phase dependence tests. (a, b) Data for two subjects performing phase dependence tests when the luminance modulations and texture–contrast modulations of the stimuli have same spatial frequency ($f_1 = f_2$) and the same temporal frequency ($\omega_1 = \omega_2$), and travel in the same direction. The abscissas denote different conditions: L is luminance modulation stimulus alone; C is texture–contrast modulation stimulus alone; 0, 45, 90, 135, 180, 225, 270, 315 denote the different relative phases with which the luminance modulations and the contrast modulations were linearly added. The ordinate is the fraction of correct motion direction judgments. The dotted line labeled "probability summation" indicates the predictions of the probability summation model for compound stimuli (L + C) based on the L and C conditions. The solid line indicates the averaged performance level in all the eight different relative phases. The shadow indicates ±2 SEs of the averaged data (0.95 confidence limits). (c, d) Data for phase dependence tests when the luminance modulation has half the spatial ($f_1' = 0.5 f_2$) and temporal frequency ($\omega_1' = \omega_2$) of that of the texture–contrast modulation.

from the observed percentage correct $P_j^0$ by

$$P_j^1 = \frac{(1 - P_j^0)}{(1 - 0.5)}. \quad (8)$$

Equations (7) and (8) completely specify the probability summation model in this application. The values of $P_U^0$ predicted from the probability summation model are illustrated in Fig. 10(a, b). It is obvious that probability summation gives an excellent account of the data for the combined modulations.

Figure 10 shows $\pm 2$ standard errors around the eight observed points based on luminance-plus-contrast stimuli. The probability summation predictions are based on only two points, the probabilities of correct motion direction judgments for pure luminance- and pure contrast-modulation stimuli. Therefore, the standard error of the prediction (now shown) is somewhat greater than the error around the observed data, and easily encompasses the observations.

*Test for exclusive texture–contrast modulation motion channel.* When the luminance modulation has half the spatial frequency and half the temporal frequency of that of the texture–contrast modulation, the full-wave rectified luminance modulation has the same spatial and temporal fundamental frequency as the texture–contrast modulation. In this case, if both modulations were analyzed by a single channel, it would be a "texture–contrast modulation" channel. Figure 10(c, d) shows there is no motion cancellation for any phase nor, indeed, any evidence of strong phase-dependent interactions between the two modulations. Therefore, we again conclude that at least two independent channels process the two kinds of modulations.

Given that there are two (or more) channels at work, it is clear from Fig. 10(c, d) that they combine their outputs in a way that far exceeds the expectation of probability summation. We would say that the superposition of these two modulations results in *supersummation.*

*Luminance modulations leak into the texture–contrast modulation channel.* In addition to overall supersummation, the data of Fig. 10(c, d) show a slight but reliable dependence on the motion-direction judgments on the relative phase of the two modulations. This weak phase dependence is understandable in terms of the stimulus construction. The luminance modulation is inherently a compound stimulus which contains both luminance modulation (first-order) and texture–contrast modulation (second-order) information. For example, in the luminance modulation stimulus, the peaks and valleys represent areas of high contrast, while the zero crossings are in areas of low contrast. When the luminance modulation is superimposed on the texture–contrast modulation so that the areas of high and low contrast in the two modulations coincide exactly, they are said to be "in phase". For modulation pairs that are exactly in phase, or exactly out of phase, the overall amplitude of texture–contrast modulation is simply the sum of the amplitudes of contrast modulation of the two components:

$$m_{\text{lum + con}} = m_{\text{lum}} + m_{\text{con}}. \quad (9)$$

The subscript indicates the nominal stimulus source of the modulation. All modulations are directed to the texture–contrast modulation channel. From the psychometric function for texture–contrast modulations, we can estimate that the contribution of the luminance modulation to the texture–contrast modulation channel (relative to the texture–contrast modulation itself) is 0.10 for subject ZL and 0.11 for subject EB. That is $m_{\text{lum}}/m_{\text{con}} \approx 0.1$, a relatively small amount of stimulus input into a nonoptimal channel. There is no sign of any leakage of texture–contrast modulation into the luminance modulation channel.

*Conclusions*

Superimposing a pedestaled luminance modulation and a texture–contrast modulation that move in opposite directions results in complete motion standstill—perfect motion cancellation. When the two modulations move in the same direction, to a good first approximation, motion-direction judgments show no dependence on the relative phase. From the lack of phase dependence, we conclude independent motion-energy channels extract the motion of luminance modulations and of texture–contrast modulations. From perfect motion cancellation, we conclude that the outputs of these channels are combined at a higher level to arrive at a single-valued representation of motion direction. Finally, sine-wave luminance stimuli inevitably will leak into second-order (contrast-motion) channels, but (properly constructed) second-order stimuli have no effect on first-order channels.

### *Experiment 5. Interocular Pedestal Tests: Psychophysical Localization of the Motion Energy System*

In interocular motion presentation mode, successive frames of a five-frame motion stimulus are alternately directed to the left and right eyes (Fig. 11). Successive frames are separated by 90 deg. Within an eye, therefore, successive frames are separated by a phase of $\pm 180$ deg, and the motion signal is inherently ambiguous. In viewing these interocular stimuli with only one eye, coherent motion perception is impossible. However, coherent motion perception would be possible if information from the two eyes could be combined. When subjects can perceive coherent motion in the interocular presentations, it means that the motion computation is performed after the signals from the two eyes combine.

Shadlen and Carney (1986) and Georgeson and Shackleton (1989) demonstrated that subjects can indeed perceive interocular motion in luminance modulation stimuli. Here, we use pedestals to determine whether the interocular motion mechanism utilizes motion energy, and we compare monocular to interocular sensitivity. If subjects were to fail to perceive coherent interocular pedestal motion, it would mean that the luminance motion energy computation is inherently monocular.

Because sensitivity and temporal frequency response are signatures of processing systems, such measurements will be a critical tool for elucidating and discriminating between the mechanisms involved in monocular and interocular motion perception—a theme which will be further continued in Expt 6.

*Method*

*Stimuli: monocular and interocular displays.* Two types of drifting modulations (luminance and texture–contrast) with only five frames per cycle were created, successive frames being separated by 90 deg (Fig. 11). They were presented to the subjects in four different display modes: left eye only, right eye only and two interocular modes (left eye first, right eye first). The mode was chosen randomly on each trial.

The stimuli were composed of horizontal modulations that moved either up or down. The spatial frequency of the modulations was 1.28 c/deg and each frame of the modulation extended 2.94 deg horizontally and 5.88 deg vertically. (The horizontally narrow configuration permitted the left- and right-eye stimuli to be presented on the same monitor.) The spatial phase of the modulations consistently shifted upward 90 deg with every new frame or, on other trials, the phase shifted 90 deg downward. In monocular viewing conditions, the modulations were shown only to one eye. The other eye was shown the mean-luminance background. In interocular presentations (Fig. 11), the motion stimuli were shown to the left and the right eye alternately. Whenever one eye was shown the motion stimuli, the other eye was shown the background. In this way, the stimuli within each eye contained flicker (180 deg phase shifts) but no consistent motion information. Both luminance modulations and texture–contrast modulations were tested, each with and without its own pedestal.

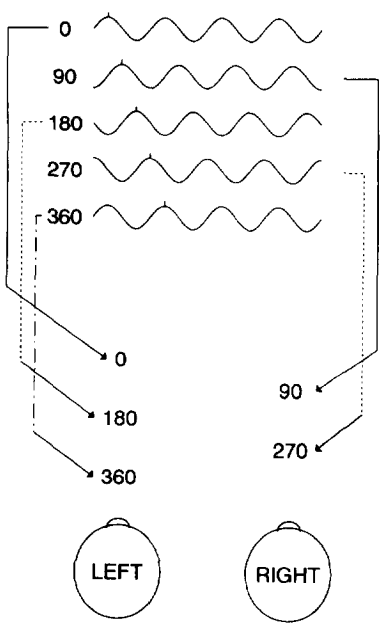


FIGURE 11. Interocular-presentation paradigm. Alternative frames of a five-frame display are directed to the left or right eye. From frame-to-frame, there is a 90 deg phase shift to the right. Successive frames in each individual eye have a 180 deg phase shift and therefore contain no motion-direction information.

Subjects viewed all the stimuli through a mirror stereoscope. To make sure that our subjects maintained good binocular fusion, every display was embedded in a surrounding black frame and contained a central fixation point. The subjects were instructed to initiate a new trial only after they fused both eyes' views (of the fixation point and black frame).

*Procedure.* In a preliminary procedure, the method of constant stimuli was used to determine the psychometric functions for both kinds of stimuli (luminance modulations, texture–contrast modulations) and the four (monocular, interocular) viewing conditions. Only one temporal frequency (3.0 Hz) was used. Trials were blocked by stimulus types. Within a block all four viewing conditions were randomly mixed. The outcome of this procedure was the threshold modulation $m_{75}(s, v)$ for each kind of stimulus $s$ and viewing condition $v$.

The threshold values $m_{75}(s, v)$ were carried forward to the main experiment which used the standard 2:1 pedestal test for both stimulus types and for the four different viewing conditions. For each stimulus type and viewing condition, the amplitude of the motion stimulus was presented at each subject's estimated threshold. The subject made motion-direction judgments when the motion stimulus was shown alone or with a pedestal twice its amplitude. (Obviously, the pedestal was always shown in the same eye and together with the moving modulation.) The four viewing conditions and the two pedestal conditions (pedestal, no pedestal) were randomly mixed within a block; different stimulus types were tested in separate blocks.

*Results*

*Monocular viewing.* Data for the two monocular viewing conditions (left eye, right eye) are lumped together because there were no discernible differences. Table 2 summarizes two subjects' performance in all the experimental conditions. In monocular viewing conditions, for both types of modulations (luminance and texture–contrast) and for both subjects, levels of performance as indexed by percentage correct in judging motion direction are the same whether the 2× pedestal was present or absent. However, in interocular presentation, the presence of the 2× pedestal reduced the subjects' performance to chance guessing level.

Again $t$-tests indicate that there is no difference between 2× and 0× pedestal conditions in monocular viewing modes ($P \geq 0.48$), i.e. there is no difference between pedestal and nonpedestal performance.

*Interocular viewing.* Data for the two interocular viewing conditions (left eye first, right eye first) are lumped together because there were no discernible differences. In interocular viewing, subjects were able to make accurate motion-direction judgments of modulations without pedestals. This confirms the observations of Shadlen and Carney (1986) and Georgeson and Shackleton (1989). With pedestals, however, subjects' performance was statistically equivalent to chance

TABLE 2. Interocular and monocular pedestal tests: accuracy of motion direction judgments

| Viewing mode | Stimulus type and threshold amplitude (ZL/EB) | Subject: ZL, 0× pedestal | Subject: ZL, 2× pedestal | Subject: EB, 0× pedestal | Subject: EB, 2× pedestal |
|---|---|---|---|---|---|
| Interocular | Luminance modulation (0.050/0.050) | 0.84 | 0.52 | 0.75 | 0.48 |
| | Contrast modulation (0.092/0.130) | 0.80 | 0.54 | 0.79 | 0.55 |
| Monocular | Luminance modulation (0.0055/0.0065) | 0.79 | 0.80 | 0.76 | 0.74 |
| | Contrast modulation (0.045/0.056) | 0.82 | 0.84 | 0.78 | 0.75 |

Chance performance level is 0.50. Each data entry is based on 100 observations. Entries of the form (0.050/0.050) indicate the threshold modulation amplitudes for subjects (ZL/EB). 2× indicates twice the threshold modulation amplitude, 0× indicates no pedestal.

$t$-Test: $P \leq 1.95 \times 10^{-7}$, 2× vs 0× pedestal conditions in interocular displays.

$t$-Test: $P \geq 0.32$, 2× pedestal vs 0.50 in interocular displays.

$t$-Test: $P \geq 0.48$, 2× vs 0× pedestal in monocular displays.

guessing (50% correct). No pedestal cell (subject × stimulus type) even remotely approaches a statistically significant difference from 50%. However, the statistical significance for the difference between the 2× and 0× pedestal conditions is $1.95 \times 10^{-7}$.

*Interocular quality loss.* Here we consider only stimuli without pedestals. Table 2 shows the threshold contrasts for each subject for the two nonpedestal stimuli (luminance modulation, texture–contrast modulation). To achieve threshold, luminance modulations require, on the average, 10 times more contrast for interocular presentations than for monocular presentations. For texture–contrast modulations, interocular stimuli require 3 times more contrast. Presumably these ratios reflect the insensitivity of the feature-tracking system relative to each kind of monocular motion-energy system.

### *Conclusion*

We interpret the failure of subjects to perceive pedestaled interocular motion for either drifting luminance or texture–contrast modulations to mean that the motion-energy systems are exclusively monocular. We conclude that the motion-energy systems are monocular and they reside at neural sites before binocular combination.

Given that the motion-energy systems fail to perceive the pedestaled interocular displays, interocular motion can still be extracted by the feature-tracking mechanism. However, relative to the monocular motion-energy systems, feature tracking extracts the motion of interocular nonpedestal stimuli at a cost. Relative to monocular detection, interocular motion-direction detection thresholds are elevated three-fold for texture–contrast modulations and 10-fold for luminance modulations. We will show in Expt 6 that the inefficiency of perceiving interocular luminance modulations is due almost entirely to the inherent inefficiency of feature-tracking mechanism for these displays, not to the interocular mode of presentation.

## *Experiment 6. Temporal Frequency Characteristics: Interocular Motion of Luminance Modulations and of Motion Modulations*

In Expt 5, we established that the motion-energy system fails to extract motion from interocular drifting luminance modulation. It seems likely that when interocular motion is perceived, it is extracted by the feature-tracking system. In Expt 1, we found two temporal frequency characteristics: (1) luminance modulation and texture–contrast modulation motions had a 12 Hz cutoff frequency; (2) depth- and motion-modulation motion had a 3 Hz cutoff. These were identified, with a motion-energy system and a feature-tracking system respectively. To gain further evidence for the hypothesis that interocular motion is perceived by the feature tracking system, we wish to determine whether the temporal frequency characteristic of interocular motion extraction matches feature-tracking's previously observed 3 Hz cutoff. To determine the temporal frequency characteristic, we measure subjects' motion-direction thresholds for interocular drifting luminance modulations at various temporal frequencies.

### *Method*

*Frequency response characteristics.* Five-frame, 90 deg, luminance modulations were presented interocularly as described in Expt 5. There were no pedestals. Stimuli were presented at four temporal frequencies: 0.94, 1.88, 3.75 and 7.50 Hz. The spatial frequency of the stimuli was 1.28 c/deg, and each frame of the stimuli extended 2.94 deg horizontally and 5.88 deg vertically. Motion was either up or down. At each temporal frequency, the method of constant stimuli was used to estimate a 75% motion-direction threshold for the two subjects.

*Monocular–interocular motion-modulation comparison.* A control experiment was conducted to compare sensitivity in monocular and interocular displays for the motion-modulation stimulus. Session were composed of a random mixture of monocular and

interocular trials, so that the thresholds for these two kinds of presentations could be directly compared at the same temporal frequencies (0.94 Hz) and spatial frequencies (0.64 c/deg).

*Results*

*Frequency response characteristics.* For luminance-modulation motion, compared with binocular presentation, interocular presentation increases the contrast threshold (at low temporal frequencies) by a factor of 16 for subject ZL (from 0.14% to 2.3%), and by a factor of 8 for subject EB (from 0.22% to 1.8%). An even more significant change is the decrease in cutoff frequency from 12 to 3 Hz. For each of the subjects, the temporal frequency characteristics of interocular luminance-modulation motion are exactly like those of the depth modulations and the motion modulations (Fig. 7).

*Monocular–interocular motion-modulation comparison.* For the motion modulations, thresholds in judging motion direction in the monocular 0.94 Hz five-frame displays are considerably higher than in 0.94 Hz binocular continuous displays: 0.40 vs 0.11 for ZL; 0.44 vs 0.16 for EB. Interocular presentation of five-frame (with no pedestal) vs monocular presentation did not produce any significant difference in threshold for motion-direction discrimination for either subject.

*Discussion and Conclusions for Expts 5 and 6*

*Monocular motion energy computations vs binocular feature tracking.* The temporal frequency characteristics we have measured for interocular luminance modulations are the same as for depth motion and motion–motion. Depth motion is inherently binocular; motion–motion had the same thresholds monocularly and interocularly. This suggests all motions that can be computed interocularly are extracted by a single mechanism which, because it fails all pedestal tests, is designated as a feature-tracking mechanism. The feature-tracking mechanism apparently exhibits exactly the same frequency cutoff when it detects drifting sinusoidal luminance modulations, texture–contrast modulations, depth modulations, and motion modulations.

The motion-modulation movement can only be perceived by the feature-tracking mechanism. The fact that we observed only small threshold changes when it was converted from monocular to interocular presentation, indicates that the feature-tracking mechanism is inherently indifferent to the eye of origin.

We conclude that luminance motion-energy and contrast-modulation motion-energy detection are monocular processes and occur at neural sites before binocular combination. The feature-tracking system is almost equally sensitive to monocular and interocular displays and therefore resides beyond the neural sites of binocular combination.

*A superthreshold luminance sine wave activates all systems.* In Expt 4, luminance modulations and double-frequency texture–contrast modulations were moved together in the same direction. Motion-direction judgments depended slightly but consistently on their relative phase, an effect that was interpreted as a 10–11% leakage of a sinusoidal luminance modulation into a $2f$ texture–contrast modulation channel. A consequence of all the above is that the motion of a drifting luminance sinusoid, an apparently simple stimulus, is computed by all three systems: the monocular luminance-modulation system, which is fast and sensitive; the binocular feature-tracking system which is slow and more than 10 times less sensitive; and at a double frequency, the monocular full-wave-contrast system, which is also about 10 times less sensitive to this kind of stimulus. The drifting sinusoidal grating, which long has been regarded as a universal tool for visual psychophysics, turns out to be not a particularly useful tool for discriminating between motion mechanisms.

*Experiment 7. Spatial Contrast Sensitivity Functions of Motion-Energy Detection and Feature Tracking*

Experiment 5 established that energy detection of luminance modulation motion and of contrast-modulation motion was entirely monocular. In interocular presentations, where combined information from both eyes is required to solve the motion problem, the perceptual system relies on feature-tracking mechanisms. We exploit these facts to measure the modulation transfer functions of the motion-energy and feature-tracking mechanisms.

*Method*

Only luminance modulations are used. The four spatial frequencies are 0.6, 1.2, 2.4 and 4.8 c/deg. There are four display modes: monocular, left eye, right eye; interocular, left-eye first, right-eye first. Spatial frequency was varied by changing the viewing distance while keeping the display the same. Other display conditions were as in Expt 5 (Fig. 11).

The method of constant stimuli was used to measure psychometric functions for modulation contrast for two subjects under all the stimulus conditions. All display modes were mixed within a block. Because of the method of producing different spatial frequencies (the subject walks between viewpoints), only one spatial frequency was tested within a block. A Latin Square balanced design was used to balance all the spatial frequency conditions.

*Results*

At the highest contrast tested, performance was perfect or almost perfect for all spatial frequencies. This indicates that interocular misalignment is not a limiting factor. We define the threshold $t_{75}(\alpha)$ as the amplitude of the stimulus at a given spatial frequency $\alpha$, which, when presented to the subject, produced 75% correct response level in motion direction judgments. The spatial contrast sensitivity function is plotted in Fig. 12 using $\log_{10}[1/t_{75}(\alpha)]$ as the vertical coordinate and $\log_{10}(\alpha)$ as the horizontal coordinate. There is no difference between the left and right eye in the monocular viewing conditions, so the combined data are shown in the curve labeled "mono". Similarly, there was no discernible

response difference between interocular left-first and right-first presentations, so these are combined in the curve labeled "interocular". The monocular contrast sensitivity function is virtually a horizontal line in the spatial frequency range of our investigation (from 0.6 to 4.8 c/deg). This result confirms van Santen and Sperling's earlier observations (van Santen & Sperling, 1984).

The modulation transfer function for the interocular displays is low-pass in the spatial frequency range under investigation with a continuous decline of sensitivity from 0.6 to 4.8 c/deg. For subject ZL there is about a $1.2 \log_{10}$ ($16\times$) loss of sensitivity as spatial frequency increases from 0.6 to 4.8 c/deg ($0.9 \log_{10}$). The average slope on log–log coordinates is about $-0.75$ (Fig. 12). At the highest spatial frequency subject EB could not perform the task even with 100% luminance modulation; his sensitivity loss over the range of frequencies tested is $>16\times$, with a correspondingly larger negative slope.

From the spatial contrast sensitivity functions, we conclude that the luminance motion-energy mechanism is equally sensitive to all the spatial frequencies in the range tested. Sensitivity of feature-tracking mechanism declines (thresholds increase) in direct proportion to spatial frequency.

## GENERAL DISCUSSION

The results of all the seven experiments are embodied in the flowchart of Fig. 13 that represents the functional architecture of human visual motion perception. The left and right sides of the chart represent the two parallel systems: the fast, monocular, motion-energy system on the left and the slower, binocular, feature-tracking system on the right.

### *The motion-energy system*

Signals from both eyes feed both systems. The monocular motion-energy system contains four components (channels): for each eye there is a motion-energy first-order channel to process motion of luminance modulations and second-order motion-energy channel to process motions of texture–contrast modulations. The motion-energy systems combine motion information from both eyes and from first-order and second-order stimuli only *after* the motion computation has been completed in each channel.

With respect to the combination rules, for one-dimensional motions investigated in Expt 4, there was no transparency: the combined output is expressed as a single number representing the motion strength in a given direction (plus or minus). How that number is computed depends on the frequencies of the component stimuli: probability summation is the rule for same-frequency motion of luminance and contrast modulations; supersummation is the rule when the contrast modulation has twice the frequency of the luminance modulation. This sparse description of the motion-combination algorithms reminds us that there remains a great deal to be learned about motion combination and especially about velocity perception, which has not been considered here.

The motion-energy system is fast (cutoff frequency is 12 Hz), sensitive (threshold modulation is about 0.002) and equally sensitive to a wide range of spatial frequencies. It is immune to stationary pedestals but fails to survive interocular presentations.

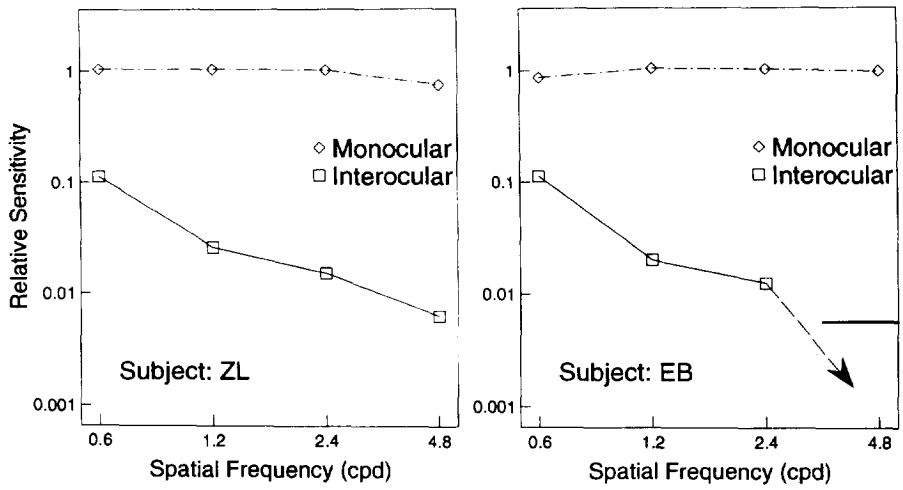


FIGURE 12. Spatial contrast sensitivity characteristics of the visual system for detecting motion of monocular and interocular luminance-modulation stimuli. The ordinate, relative sensitivity, is computed by dividing the modulation threshold for each spatial frequency and type of stimulus into the average threshold of the three lowest spatial frequency luminance-modulation stimuli. The abscissa is spatial frequency. The data graphs are "normalized spatial frequency characteristics" for monocular and interocular luminance-modulation motion-direction judgments. Data are shown for two subjects. The flat monocular characteristics indicate that the monocular motion energy computation has both high spatial resolution and high sensitivity throughout this spatial frequency range. The sloping interocular characteristics indicate that, in this spatial frequency range, the binocular feature-tracking computation—at its best—is 10 times less sensitive than the energy computation, and it has poorer spatial frequency resolution. The bar in the right margin (subject EB) indicates the maximum available contrast; the dotted line and arrow indicate that this was still below EB's threshold.

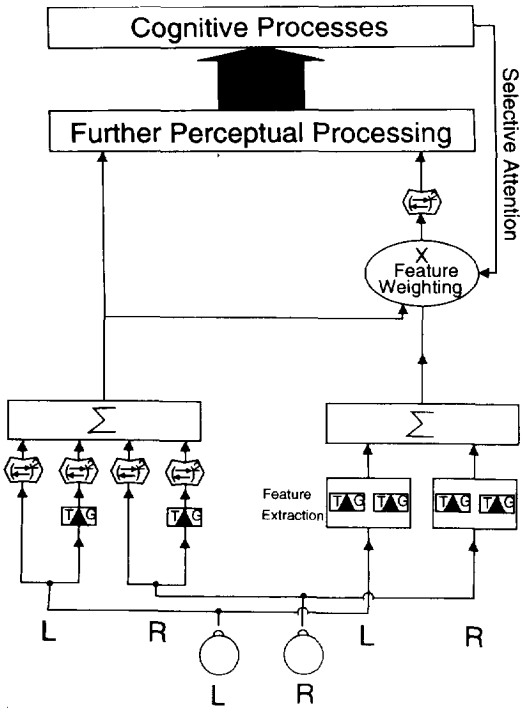


FIGURE 13. Functional architecture of the visual motion system. The left half represents the fast monocular systems; the right half represents the binocular feature-tracking system. L and R indicate left and right eye signals respectively; $(\rightleftarrows)^2$ indicates a motion-energy (Reichardt) detector; TG indicates texture grabber (a spatial filter followed by full-wave rectification); Σ indicates (possibly complex) summation. Feature extraction represents processes similar to but more extensive than monocular texture grabbing; feature extraction occurs both before and after binocular combination as indicated by the two feature-extraction boxes; × represents multiplication—the differential weighting by selective attention of the features extracted from the binocular (cyclopean) image *and* from the monocular motion-extraction processes (represented by the central horizontal arrow).

*The feature tracking system*

The feature tracking system seems to combine inputs without regard to the eye of origin. Neurally, it resides beyond the locus of interocular signal combination. It computes motion in a more restricted temporal range (cutoff frequency about 3 Hz), has lower spatial resolution and lower sensitivity than the motion-energy system. Feature strength as a function of spatial location, the input to the feature tracking system, is based on a computation that has both bottom-up and top-down components. The basis for asserting a top-down action for feature tracking is a companion experiment (Sperling & Lu, 1995) in which subjects viewed ambiguous motion displays in which alternate frames contained completely different kinds of features (e.g. two different depths and two different textures). The apparent direction of movement depended on which feature values were attended, demonstrating strong top-down influences in the feature-tracking motion system.

The ultimate algorithm by which motion computation is done in the feature tracking system has yet to be determined. We suspect that motion-energy analysis (elaborated Reichardt detector) is still a plausible candidate. The failures of our subjects to perceive coherent motion in pedestaled interocular luminance modulations, pedestaled depth modulations, and in pedestaled motion modulations have two possible explanations. (1) Perhaps these stimuli are represented in the feature system with a too-coarse quantization of stimulus values to resolve the pedestal stimuli. This might well be an inherent property of a feature computation—features are represented only as present or absent, 0 or 1. In that case, the motion algorithms itself might be a motion-energy computation but it would fail because of too-coarse input quantization. (2) On the other hand, the motion extraction algorithm might be conceptually different from the motion-energy computation.

Shadlen and Carney (1986) studied stimuli that contained a mixture of sine waves moving at different velocities, which tends to eliminate the correspondence of features between eyes in interocular displays. Carney and Shadlen (1993) used random textures in which a similar interocular feature-correspondence-failure was introduced. Nevertheless, subjects were able to perceive interocular apparent motion in both kinds of displays. On the other hand, Georgeson and Shackleton (1989, 1992) studied interocular motion in translations of square waves that have a missing fundamental spatial frequency. They theorize that the apparent direction of interocular motion of these stimuli is mediated only by feature visibility, not by a motion energy computation.

The resolution to the apparently contradictory claims for the role of features in interocular displays is that both are partially true. We have experimented with several definitions of features (none were provided by the authors cited above) and applied them to the interocular displays cited above. A motion energy system, operating on the outputs of these feature detectors, does extract the latent motion in all the displays in which humans have been able to perceive interocular motion. This suggests that the third-order system (motion feature-tracking), utilizes a motion energy computation even though it normally is driven by feature inputs whose amplitude is coarsely-quantized. Thus interocular motion fails for pedestaled sine waves because the intensity quantization in the feature representation is too coarse. If the features that mark the wobbling peaks and valleys of a pedestaled sine wave could represent subtle variations in amplitude (which they cannot), then the interocular linear motion of the sine-wave grating would be perceived.

*Implications*

The methodologies developed in this study provides useful means for distinguishing between classes of motion extraction mechanisms. We anticipate their application in other stimulus domains, such as equal-luminance color motion perception, drifting flicker, and other types of "second-order" stimuli.

We were only concerned with motion direction discrimination in our study. However, it has been proposed that the outputs from the first-order motion-energy computation provide the inputs to the motion-velocity system (e.g. Heeger, 1987) and to the three-dimensional structure-from-motion computations (Sperling *et al.*,

1989; Dosher, Landy & Sperling, 1989). We also anticipate applications of variants of our methodologies to test these and related issues.

## SUMMARY AND CONCLUSIONS

Four types of moving stimuli were used in the experiments: a luminance modulation [Fig. 3(d)], a texture–contrast modulation [Fig. 4(d)], a depth modulation [Fig. 5(a)], and a motion modulation (Fig. 6).

*Experiment 1*

Subjects' thresholds for 75% accuracy in judging motion direction of the four kinds of stimuli were determined in temporal frequencies from 1 to 16 Hz. These data define the temporal frequency-sensitivity characteristics. All the temporal frequency-sensitivity characteristics are lowpass: detections of motions of luminance modulations and of texture–contrast modulations have an identical cutoff frequency of 12 Hz; detections of motions of depth and of motion modulations have an identical cutoff frequency of 3 Hz.

*Experiment 2*

Fixing the amplitudes of the motion stimuli at their threshold level, luminance modulations were subjected to pedestal tests. The following amplitude ratios of pedestal to motion stimulis were tested: 1:1, 2:1 and 7:1. The 1:1 ratio test replicated van Santen and Sperling's (1984) pedestal test. The 2:1 ratio test was the standard pedestal test used in the rest of this whole project. The 7:1 ratio test was used to explore early nonlinearities of the motion system. In addition to sine pedestals, we also investigated motion thresholds in the presence of added stationary spatial binary white noise with 7 times the motion-threshold amplitude. The aim was to determine whether the early nonlinearities in the motion system are spatial frequency specific. Subjects' complete immunity to 1:1 and 2:1 pedestals indicated that a motion-energy computation was the mechanism of first-order motion-direction detection of low amplitude stimuli. Immunity to 7× binary noise but vulnerability to 7× sine pedestals indicated an early saturating nonlinearity of the luminance-modulation motion system that was spatial-frequency specific.

*Experiment 3*

Immunity of texture contrast modulations (second-order stimuli) to pedestals indicated motion-energy detection. The percept of apparent motion from depth-modulation movements and from motion-modulation movements is abolished by pedestals. Motion extraction from these stimuli is accomplished by a feature-tracking mechanism.

*Experiment 4*

Given that motion of drifting luminance modulations and of texture-contrast modulations is computed by a motion-energy mechanism, we used the *relative phase dependence* test to examine whether luminance and texture–contrast modulations activated separate motion channels. When superimposed luminance and texture–contrast modulations of the same temporal and spatial frequencies move in the same direction, there is no phase dependence in motion-direction judgments. The probability of correct motion detection in the combined stimulus follows probability summation. When the texture–contrast modulation has twice the frequency of the luminance modulation, there is a highly significant supersummation but only a slight phase dependence. When a pedestaled luminance modulation is superimposed on an equal strength, oppositely-moving, pedestaled contrast modulation of the same temporal and spatial frequencies, there is perfect motion cancellation. These results indicate the directions of apparent motion in drifting luminance modulations and contrast modulations are first computed in independent channels and only combined thereafter. The one instance of slight phase dependence was explained as being caused by 10% leakage of a $1f$ luminance modulation sinusoid into a $2f$ texture–contrast modulation channel.

*Experiment 5*

Thresholds of drifting luminance modulations and contrast modulations were measured in three different viewing conditions: monocular left eye, monocular right eye, and interocular displays, all at an intermediate temporal frequency (3 Hz). Then, performance was measured for pedestaled tests in the same three viewing conditions. Subjects' performance was unaffected by the presence of pedestals (their performance was precisely the same with and without pedestals) in the monocular conditions but, with pedestals, they were unable to judge motion direction in interocular displays. This indicates that interocular stimuli activate primarily a feature-tracking mechanism.

*Experiment 6*

The interocular temporal frequency characteristic was used to characterize different modes of motion analysis. Subjects' thresholds for drifting luminance modulations were measured at various temporal frequencies using interocular displays. The cutoff frequency for these stimuli is 12 Hz with monocular displays, but interocular luminance modulations produce exactly the same 3 Hz cutoff frequency as was found earlier for depth and motion modulations. This suggests that interocular presentations, even of luminance modulations, are detected by the feature-tracking motion system; its defining characteristic is the cutoff frequency of 3 Hz.

*Experiment 7*

The accuracy of motion-direction judgments was measured with drifting luminance modulations at various spatial frequencies in monocular and interocular viewing conditions. In monocular viewing, subjects had equal sensitivity to a range of spatial frequencies from 0.6 to 4.8 c/deg. With interocular viewing, there was a 10 dB sensitivity decline at 0.6 c/deg and a further 15 dB decline of sensitivity as spatial frequency increased to

4.8 c/deg. The first-order (luminance motion-energy) system is equally sensitive to a wide range of spatial frequencies, whereas the third-order (feature tracking) system's sensitivity declines roughly in proportion to spatial frequency.

*Functional organization of the human visual motion perception*

Based on the seven experiments, a functional control chart is proposed. A first-order luminance system and a second-order texture–contrast system use independent motion-energy detectors, operate in parallel, and combine their outputs at an early stage. A third-order (feature-tracking) system receives inputs (features) from texture grabbers and from the lower-order motion systems. The strength of feature inputs to the third-order motion system is subject to top-down control—attention to particular features influences their strength and thereby the perceived direction of motion.

*Note added in proof.* Since the preparation of this article, Carney [(1995) *Investigative Ophthalmology & Visual Science, 36/4,* 52] has reported finding observers who can perceive interocular pedestaled motion. The critical difference is his use of very long duration displays to give even very weak processes every advantage in reaching threshold. Using displays similar to Carney's (personal communication), we found that one of our two most sensitive observers could indeed perceive interocular pedestaled luminance-modulation motion. His interocular sensitivity was about a factor of 10 lower than his monocular motion sensitivity. He also could perceive interocular pedestaled contrast-modulation motion slightly above chance, but failed to reach a 75% threshold within the range of physically available contrasts. The other observer failed to perform above chance levels with these interocular pedestaled displays. Therefore, with respect to mechanisms, where the present paper says "exclusively monocular" or "monocular", the reader should understand "primarily monocular" or "greater than 90% monocular". In terms of psychophysics, under ordinary binocular viewing, the interocular component of luminance- or contrast-modulation motion perception is insignificant. In terms of physiology, the existence of even this very small amount of interocular crosstalk in the primarily monocular mechanisms is critical because it places these motion mechanisms beyond the point at which the signals from the two eyes combine.

## REFERENCES


Adelson, E. H. & Bergen, J. K. (1985). Spatio-temporal energy models for the perception of apparent motion. *Journal of the Optical Society of America A, 2,* 284–299.

Adelson, E. H. & Bergen, J. R. (1986). The extraction of spatio-temporal energy in human and machine vision. In *Motion: Representation and analysis (IEEE Workshop Proceedings)* pp. 151–155.

Barlow, H. B. (1979). Reconstructing the retina image in space and time. *Nature, 279,* 189–190.

Braddick, O. (1974). A short-range process in apparent motion. *Vision Research, 14,* 519–529.

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

Carney, T. & Shadlen, M. N. (1993). Dichoptic activation of early motion system. *Vision Research, 33,* 1977–1995.

Cavanagh, P. (1991). Short-range vs long-range motion: Not a valid distinction. *Spatial Vision, 5,* 303–309.

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

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

Chubb, C. & Sperling, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A, 5,* 1986–2006.

Chubb, C. & Sperling, G. (1989a). Two motion perception mechanisms revealed by distance driven reversal of apparent motion. *Proceedings of the National Academy of Sciences, U.S.A., 86,* 2985–2989.

Chubb, C. & Sperling, G. (1989b). Second-order motion perception: Space–time separable mechanisms. *Proceedings: Workshop on Visual Motion (March 20–22, 1989, Irvine, CA)* pp. 126–138. Washington, D.C.: IEEE Computer Society Press.

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

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

Derrington, A. M., Badcock, D. R. & Henning, G. B. (1993). Discriminating the direction of second-order motion at short stimulus durations. *Vision Research, 33,* 1785–1794.

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

Dudley, L. P. (1951). *Stereoptics: An introduction.* London: MacDonald.

Emerson, R. C., Bergen, J. R. & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research, 32,* 203–218.

Exner, S. (1875). Experimentelle Untersuchung der einfachsten psychishen. Processe. *Archiv für die Gesamte Physiologie des Menschen und der Tiere, 11,* 403–432.

Georgeson, M. A. & Shackleton, T. M. (1989). Monocular motion sensing, binocular motion perception. *Vision Research, 29,* 1511–1523.

Georgeson, M. A. & Shackleton, T. M. (1992). No evidence for dichoptic motion setting: A reply to Carney and Shadlen. *Vision Research, 32,* 193–198.

Graham, N. V. S. (1989). *Visual pattern analyzers* (Oxford Psychology Series, No. 16). New York: Oxford University Press.

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

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

Kenkel, F. (1913). Untersuchungen ueber zusammenhang zwischen erscheinungsgross und erscheinumngsbewegung beim einigen sogenannten optischen Tauschungen. *Zietschrift für Psychologie, 61,* 358–449.

Landy, M. S., Cohen, Y. & Sperling, G. (1984a). HIPS: A Unix-based image processing system. *Computer Vision, Graphics and Image Processing, 25,* 331–347.

Landy, M. S., Cohen, Y. & Sperling, G. (1984b). HIPS: Image processing under UNIX-software and applications. *Behavior Research Methods, Instruments and Computers, 16,* 199–216.

Lelkins, A. M. M. & Koenderink, J. J. (1984). Illusory motion in visual displays. *Vision Research, 24,* 293–300.

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

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

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

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

Reeves, A. & Sperling, G. (1986). Attention grating in short-term visual memory. *Psychological Review, 93,* 180–206.

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

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Rosenblith, W. A. (Ed.), *Sensory communication*. New York: Wiley.

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

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

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

Solomon, J. A. & Sperling, G. (1995). Full-wave and half-wave rectification in 2nd-order motion perception. *Vision Research, 34*, 2239–2257.

Sperling, G. & Lu, Z.-L. (1995). Attention affects the perceived direction of visual motion. *Investigative Ophthalmology & Visual Science (Suppl.), 36*, 854.

Sperling, G., Dosher, B. A. & Landy, M. S. (1990). How to study the kinetic depth experimentally. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 445–450.

Sperling, G., Landy, M. S., Dosher, B. A. & Perkins, M. E. (1989). The kinetic depth effect and the identification of shape. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 426–440.

Turano, K. & Pantle, A. (1989). On the mechanism that encodes the movement of contrast variations—I: velocity discrimination. *Vision Research, 29*, 297–221.

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

Watson, A. B. & Ahumada, A. J. Jr. (1983). A look at motion in the frequency domain. In Tosotsos, J. K. (Ed.), *Motion: Perception and representation* (pp. 1–10). New York: Association for Computing Machinery.

Werkhoven, P., Sperling, G. & Chubb, C. (1993). Motion perception between dissimilar gratings: A single channel theory. *Vision Research, 33*, 463–485.

Wertheimer, M. (1912). Ueber des sehen von scheinbewegunen und scheinkorpern. *Zietschrift für Psychologie, 61*, 161–265.

Wheatstone, C. (1838). On some remarkable, and hitherto unresolved, phenomena of binocular vision. *Royal Society of London, Philosophical Transactions*, 371–394.

Woodworth, R. S. & Schlosberg, H. (1954). *Experimental psychology* (revised edn). New York: Holt, Rinehart & Winston.

Zanker, J. M. (1993). Theta motion: A paradoxical stimulus to explore higher order motion extraction. *Vision Research, 33*, 553–569.

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

*Acknowledgements*—This research was supported by AFOSR, Life Science Directorate, Visual Information Processing Program. The authors thank Quest International, Inc. for their help in locating and setting up the IKEGAMI monitor and Erik Blaser for participating in the experiments.