Sperling, G. (2001). Motion perception models. In N. J. Smelser & P. B. Baltes (Eds.), *2001 International Encyclopedia of the Social and Behavioral Sciences* (pp. 10093-10099). Oxford, UK: Pergamon.

Motion Perception Models

A motion perception model is a theory or computation that relates visual stimuli or visual scenes that contain motion to the motion perceptions and motion-related actions of observers. We restrict ourselves here to formal mathematical and computational models. We can know about another person's perceptions only by his or her observable responses. Therefore, the input to a motion model is a quantitative description of a stimulus or scene, the output is a predicted response. In a mathematical motion model, a stimulus (such as a moving sinewave grating) is described by an equation. For example, a mathematical theory might make a prediction, for sinewave gratings, of the probability that an observer could correctly discriminate a particular leftwards-moving grating from a rightwards-moving grating.

In a computational theory, the stimulus typically is described in terms of an x, y, t cube (see Fig. 1a). For example, the scene facing a pilot landing an airplane can be described in terms of the luminance at each point (x, y) of the pilot's visual field as a function of time t. The theory might predict the pilot's perceived orientation and velocity relative to a runway. In a psychophysical experiment (in an aviation simulator), the perception would be measured by asking the pilot to set a marker to the spot at which he believes the aircraft is heading at the moment of testing. In a more complex simulation, the theory would predict the manipulation of the controls in a real or simulated aircraft.

Motion theories themselves break down into components: pure motion processing, and subsequent perceptual, decision, and motor components. Even the pure motion processing is now believed to consist of several stages: motion direction is computed first, then velocity, then more complex motion computations that incorporate velocity at many locations to derive velocity gradients—curl, shear, and divergence—and finally, structure-from-motion, a computation in which the 3-D structure of an object or environment is derived from the complex properties of the 2-D motion field.

1. *Historical Background*

The formal psychophysical study of motion perception is usually traced to Exner (1875), who produced a spark at one spatial location x_1 followed after a short interval Δt by a spark at an adjacent location x_2. For a wide range of distances Δx between the locations and of time intervals Δy, observers perceived motion between the two locations. This is often called apparent motion (in contrast to real motion), because there is no instantaneous movement of the stimulus. Exner's observations, elaborated by Wertheimer (1912), ultimately became a cornerstone of Gestalt Psychology as the Phi phenomenon: The whole, which includes a perception of motion, is more than the sum of the parts—two static flashes.

One might have expected that demonstrations of stroboscopic motion in the 1830s, which preceded Exner, and the later advent of motion pictures early in the twentieth century would have demystified apparent motion. It is perfectly obvious that as the number of frames per second in a motion picture increases, its representation of motion more and more closely approximates real motion. Phi is a special case of only two frames of a movie that depicts motion. Although a simple theory that encompasses both sampled (apparent) and continuous (real) motion was published by Fourier (1822), it was not applied to motion until more than 150 years later.

2. *Fourier Analysis of Motion Stimuli*

Fourier analysis is now regarded as the default motion model, or better perhaps, as the default description of motion itself. As noted above, when color is neglected, a stimulus is characterized by the luminance falling at a point x, y at time t in the 2-D visual field. Suppose the x dimension is sampled at n_x points, the y dimension at n_y points and the t dimension at n_t points. Then it requires $N = n_x \times n_y \times n_t$ points to describe the visual stimulus. In computer terminology, one would say the n_x, n_y, n_t stimulus cube consists of N voxels (units of volume). Whereas visual space is three-dimensional (it takes three numbers to describe the location of a point in 3-D space), the dimensionality of the stimulus cube is N, because N numbers are required to describe it.

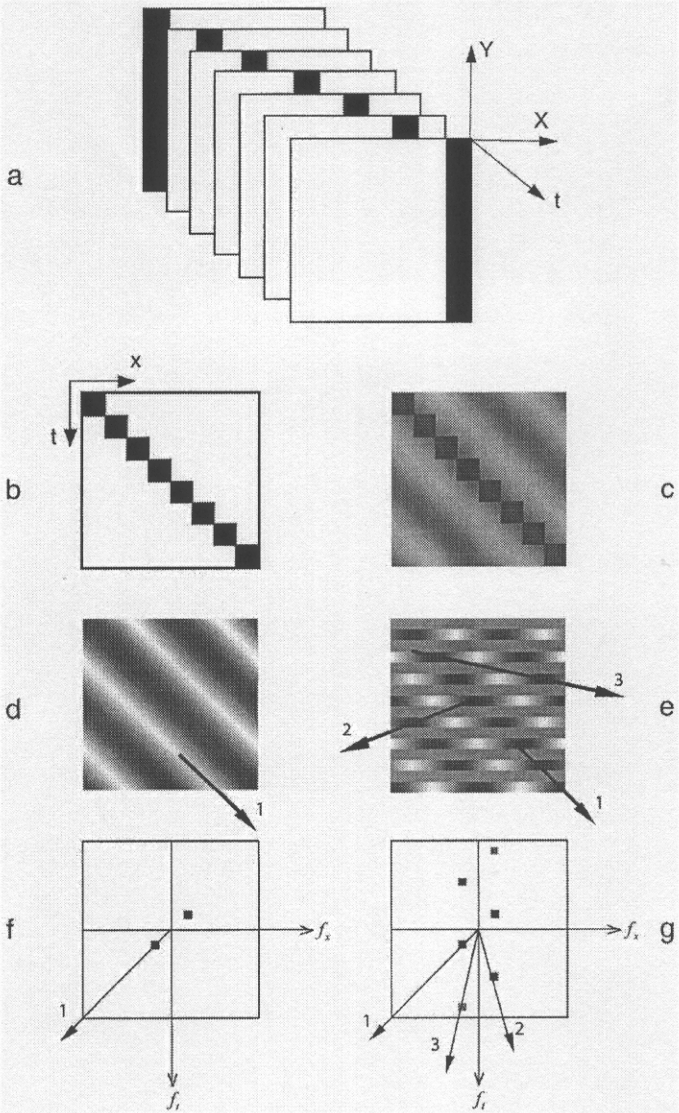

Figure 1
(a) A stimulus cube illustrating eight successive frames in the left-to-right movement of a vertical black bar. (b) An x, t cross section of (a). (c) Same as (b) with a sinewave (the dominant sinewave component in (a) superimposed. (d) x, t cross section of a sinewave grating that is moved continuously. The arrow represents the direction of movement. (e) x, t cross section of a sinewave grating that is sampled every 90 degrees. The arrows represent directions of movement. (f) Fourier amplitude spectrum of (d). The axes represent the frequencies f_x and f_t of the Fourier components. The weight of the points represents the amount of the indicated component. The arrow(s) represent the Fourier component(s) corresponding to the direction(s) of movement as (e). (g) Fourier amplitude spectrum of (e). Axes similar to (f) (after Chubb and Sperling 1985 and 1989)

A discrete Fourier transform describes an n_x, n_y, n_t stimulus cube as the sum of $N = n_x \times n_y \times n_t$ sine and cosine component waves. The Fourier representation is equivalent to a rotation in N-dimensional space. That is, the same space-time stimulus is simply viewed from a different vantage point; all the relational properties within the stimulus remain unchanged. The Fourier representation of a sampled motion stimulus

(e.g., a Phi stimulus or a motion picture) has all the same components as the continuously moving stimulus plus some extra components that represent the sampling function (Fig. 1). Watson et al. (1986) propose a useful rule of thumb to determine whether an observer can discriminate a continuously moving from a sampled stimulus: Discrimination requires that at least one of the extra temporal frequencies introduced by sampling is less than 30 Hz and less than 30 cycles per degree (cpd).

The Fourier representation of the typical Phi apparent-motion stimulus contains many sampling-produced frequencies below 30 Hz and 30 cpd. Therefore, according to the Fourier model, Phi is easily discriminable from a continuously moving stimulus. But it is perceived as moving because it also contains the same motion components as a continuously moving stimulus.

3. *Reichardt Detector for First-order Motion*

The Fourier model is a descriptive theory. It does not specify the actual processes by which motion is perceived. The first computational motion process theory was proposed by Reichardt (1957, 1961) to account for motion-induced responses of the beetle, *Chlorophanus*. Reichardt's theory was adapted for human vision by van Santen and Sperling (1984, 1985). According to their theory, corresponding to every small neighborhood of the visual field, there is an array of Reichardt detectors of different scales (from small to large) and different orientations (from 0 to 179 degrees) that compute motion direction in this neighborhood. Motion is detected according to a voting rule, that is, a rule for combining the outputs of the detectors. The voting rule that seems to be used for computing motion direction when there are many oppositely directed motions (as is typical in sampled motion), is to choose the direction that has greatest number of Fourier components above threshold, i.e., the greatest number of active Reichardt detectors (Sperling et al. 1989).

Figure 2 shows an elaborated Reichardt detector (ERD). SF_A and SF_B represent spatial filters in adjacent spatial locations. To detect left-to-right motion, the output of SF_A is delayed by filter TF. When an object traveling in the external world from SF_A to SF_B reaches SF_B with the same delay as TF, then the direct signal from SF_B and the delayed signal from SF_A both arrive at the comparator at the same time. Basically, a delay-and-compare computation is at the core of all motion theories; in the Reichardt detector the comparison operation is multiplication (covariance).

It is assumed that the inputs to a Reichardt detector are point contrasts. That is, the contrast of the mean luminance l_0 is taken as zero. A point contrast less than

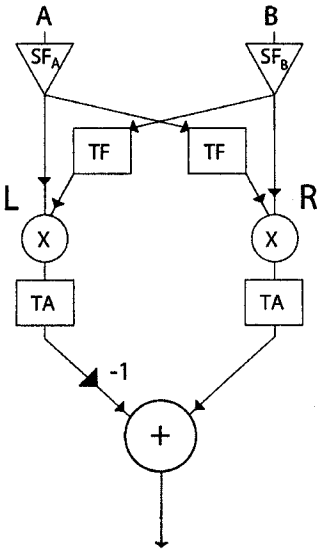

Figure 2
Elaborated Reichardt Detector (ERD). It computes motion direction from two inputs that sample the visual display in two spatial areas A and B. SF_A and SF_B denote the linear spatio-temporal filters (receptive fields) that may have different spatial distributions. In the R ('right') subunit of the detector, the output of SF_A at A is delayed by the temporal delay filter TF and then multiplied ($\times$) with the direct output from SF_B. In the L subunit of the detector, the output of SF_B is delayed by the temporal delay filter TF and then multiplied with the direct output of SF_A. The sign of the difference between the outputs of L and R subunits determines the perceived direction of motion. Outputs greater than zero indicate stimulus motion from A to B; outputs less than zero indicate stimulus motion from B to A (reproduced by permission of the Optical Society of America from Lu and Sperling 2001)

0 means the luminance at that point is $< l_0$, a positive point contrast means the luminance at that point is $> l_0$. If both the straight-through input and the delayed input are positive, or both are negative, then the multiplication produces a positive output indicating motion (in this case) from left to right. If either input is zero, the comparison produces a zero output, and if one input is positive, and the other negative, the ERD signals motion in the reverse direction.

The Reichardt detector has two subunits, one which detects left-to-right motion, the other which detects right-to-left motion. Ultimately, the outputs of these two subunits are subtracted. Thus, a positive output for the Reichardt detector of Fig. 1 indicates left-to-right motion, a negative output indicates right-to-left motion. The subtraction is critical for filtering out certain nonmotion signals that stimulate the subunits.

For example, a simple flickering light (which contains no net motion) stimulates each subunit equally. These subunit outputs are canceled in the final subtraction so that the output of the Reichardt detector is zero for nonmoving stimuli.

In the visual system, prior to motion detection, there is very powerful contrast gain control (Lu and Sperling 1996). To observe the properties of the Reichardt detector in human psychophysics requires using stimuli with sufficiently low contrast (e.g., less than a few percent) so that they are relatively undistorted by the gain control. With such stimuli, several very surprising, counterintuitive predictions of the Reichardt detector have been verified (van Santen and Sperling 1984). For example, adding a uniform-field flicker to a moving stimulus has absolutely no effect on its visibility unless the flicker happens to be of the same temporal frequency. On the other hand, when the flicker has the same temporal frequency as the motion stimulus, depending on the relative phases, the flicker can either enhance or reverse the direction of apparent motion. The successful prediction of a number of such psychophysical results has established the Reichardt computation as the algorithm of human motion-direction detection.

In addition to the Reichardt detector, two apparently different theories of motion have been proposed: motion energy detection (Adelson and Bergen 1985) Hilbert transforms (Watson and Ahumada 1985). These have been shown to be computationally equivalent to the Reichardt detector (Adelson and Bergen 1985, van Santen and Sperling 1985). Consequently, there is only one theory of human motion-direction detection, although it can be framed in different ways.

4. *Velocity Detection*

The output of the Reichardt detector gives only a measure of motion-strength in a particular direction, not a measure of velocity. Although velocity cannot be derived from an individual Reichardt detector, it can be derived from the outputs of an array Reichardt (or equivalent) detectors. Each individual Reichardt detector is optimally stimulated by a particular velocity; an appropriate voting rule enables an array of Reichardt detectors to signal the most likely velocity.

5. *Second-order Motion*

It had long been suspected that motion detection may involve different systems. Following the description of Phi motion, introspectionists proposed other categories (α, β, and γ: Kenkel 1913; Δ: Korte 1915), in addition to Phi to describe experiential aspects of motion perception (for a review, see Boring 1942).

Figure 3
Space-time representations of drift-balanced and microbalanced stimuli that selectively stimulate the second-order motion system. The overlay shows Hubel-Wiesel receptive fields oriented at +45 degrees and −45 degrees and illustrates that both have exactly the same expected outputs. Detection of left-to-right motion (or upper-left to lower-right slant) requires second-order motion (or second-order texture processing). (Adapted from Chubb and Sperling 1989.)

Late in the twentieth century, a number of dual process theories of motion were proposed, the most influential being the short- and long-range motion systems proposed by Braddick (1974). However, it was not possible to evaluate such theories because no explicit motion computation had been proposed, merely incidental properties, such as sensitivity to binocular versus monocular presentation.

With the establishment of the Reichardt model for motion perception, it quickly became obvious that observers perceived strong apparent motion in some stimuli that were completely ambiguous to Reichardt detectors and which (equivalently) contained no useful Fourier motion components. Chubb and Sperling (1988) originally characterized such stimuli as having two properties, drift-balanced and microbalanced. Originally, these stimuli were said to activate a non-Fourier motion system. Later the terminology of Cavanagh and Mather (1989) was adopted. 'First-order motion' refers to motion that can be detected by Reichardt detectors; 'second-order motion' refers to apparent motion that is invisible to Reichardt detectors.

In a drift-balanced stimulus, for every Fourier component that represents motion in one direction, there is another Fourier component, with the same expected amplitude, that represents motion in the opposite direction. In other words, a drift-balanced stimulus cannot convey Fourier motion. A microbalanced stimulus is a stimulus that remains drift-balanced when viewed through any space-time separable window. In other words, every neighborhood or combination of neighborhoods of a microbalanced stimulus is itself drift-balanced. An immediate corollary is that a microbalanced stimulus is ambiguous for every Reichardt detector or combination of detectors. Fig. 3 shows two examples of microbalanced stimuli in

which apparent motion is easily perceived. Indeed, observers cannot discriminate when they perceive motion whether the motion computation is a first-order or a second-order computation.

The model for the detection of second-order motion is the same as the Reichardt model for the detection of first-order motion with one exception. Prior to motion detection, the stimulus is processed by a texture grabber—a linear filter plus a rectifier—that produces a positive output that is proportional to the amount of texture present in a neighborhood. Thus the first-order motion system computes the motion conveyed by photons, the second-order motion system computes the motion conveyed by features (the unit of texture).

6. *Third-order Motion*

There is now a large class of exotic stimuli that are invisible to first- and second-order motion computation in which observers clearly see motion (Lu and Sperling 1995b). The basis for this motion computation appears to be figure-ground. The clearest example is a sequence of frames, each of which contains a square that is perceived as figure and in which the square translates in a consistent direction (Fig. 4). In frame 1, the figure is a red square on a green background, in frame 2, a high contrast texture on a low contrast background, in frame 3 a stereoptically produced square that appears closer to the observer than the background, and so on. The only common element linking these frames to permit a motion computation is that the area perceived as figure moves in a consistent direction.

It has been assumed that the actual motion algorithm, i.e., a Reichardt computation, is the same for all motion systems, only the preprocessing of the input differentiates them. The first-order motion system computes motion more-or-less directly from the raw representation of the stimulus (the luminance cube, $I(x, y, t)$). Second-order motion operates on an input that represents the 'amount of' features $f(x, y, t)$ at each location x, y, t. Third-order motion computes motion from a salience field $s(x, y, t)$ in which positive values of $s(x, y, t)$ denote 'figure' and zero denotes 'ground.'

7. *Properties of the Three Motion Systems*

Most natural motion, as well as most stimuli used in motion experiments, stimulates all three motion systems. Although it can be difficult to stimulate only one system, such experiments have revealed the properties of each of the three systems. First- and second-order motion seem to be primarily monocular (computed separately for stimuli arriving in each eye) and quite fast with a 'corner' frequency of 10–12 Hz. The third-order system is relatively slower, (corner frequency typically 4 Hz) and exclusively binocular, and is indifferent to the eye of origin. When a monocular image contains a third-order stimulus, but the binocular image does not, the motion cannot be processed by the third-order system (Solomon and Morgan 1999). On the other hand, only the third-order system can decode isoluminant chromatic motion (Lu et al. 1999a, 1999b), and motion in dynamic random-dot stereograms. Unlike first- and second-order motion, third-order is influenced strongly by attention (Lu and Sperling 1995a). Brain trauma can selectively destroy first-order (but not second-order) motion perception in one hemifield (Vaina et al. 1998) or second-order (but not first-order motion) perception (Vaina and Cowey 1996).

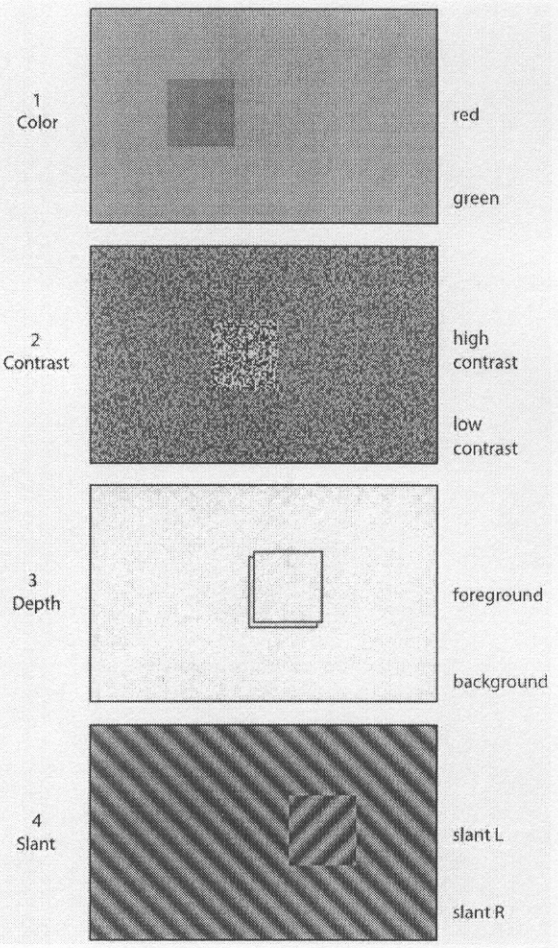

Figure 4
A sequence of frames for stimulating only the third-order motion system (after Lu and Sperling 2001)

8. Structure from Motion, KDE

When a twisted wire is rotated in front of a projection lamp so that it casts a shadow on a screen, the 2-D shadow appears to be a dynamically rotating 3-D object. Originally labeled the kinetic depth effect (KDE) by its discoverers (Wallach and O'Connell 1953), the process of the recovery of the 3-D shape is now known as 'structure from motion.' The shadow demonstration shows that cues such as texture, shading, and object familiarity are not necessary for the recovery of 3-D shape. Ullman (1979) provided the first formal theory for the recovery of 3-D structure from 2-D motion. He was able to prove that three independent views of four identifiable noncoplanar points were sufficient to recover the 3-D structure of a rigid, rotating object. Subsequently, numerous algebraic algorithms have been proposed for the recovery of a rigid 3-D structure from n views of m points. In a technical sense, these algorithms do not involve motion, merely different views. In a practical sense, they fail to recover even slightly nonrigid motion or rigid motion from imperfect data.

An alternative algorithm for deriving 3-D structure from 2-D motion uses a 'hill climbing' optimization procedure. The x, y coordinates of the structure are given in the image, only the z coordinate (the distance from the observer) is unknown. In frame 1, z values are randomly assigned to each point. In general, frame 2 will be inconsistent with a rigid rotation of the object defined by frame one. Therefore, the z-values in frame 2 are perturbed in such a direction as would have made the rotation from frame 1 to frame 2 more rigid. By successive improvements, such an algorithm can converge to a rigid object when one is presented, and can also accurately represent a slowly changing nonrigid object (Ullman 1984).

Imagine black spots painted on a perfectly transparent surface, such as a mound. When such a mound is stationary and viewed from above, it is perceived as a flat, spotted plane. As soon as it starts to move, the 3-D shape of the surface is instantly apparent. This is another example of the KDE. However, suppose that as the surface moves, in each new frame, half the dots are replaced, so that dots have a lifetime of only two frames (Sperling et al. 1989). The two-frame lifetime would cause most algorithms that depend on frame-to-frame dot correspondence to fail. On the other hand, a motion flow-field can be calculated perfectly well in such a display. And observers perfectly well perceive the 3D depth. Indeed, complex properties of the motion flow-field (such as shear) are sufficient to enable (computational) recovery of latent 3D object structure (Koenderink and van Doorn 1986).

9. Heading and Self-motion

When we move in a textured environment, the self-movement generates a motion flow-field on the retina. For the special case in which we fixate on the point towards which we are moving, the flow-field is one of divergence from a focus of expansion, and the focus itself is the point towards which we are moving. However, if we happen to be looking elsewhere, matters are much more complicated. These complications can be mathematically resolved, and it can be shown that such flow-fields contain sufficient information to determine the heading direction (Longuet-Higgins and Prazdny 1980) and other useful quantities, such as time to contact (Lee 1980).

The computation of structure from motion (and presumably also heading direction) has been shown to rely almost entirely on the first-order motion system (Dosher et al. 1989). This is probably a matter of spatial resolution. The computation of complex properties of a flow-field (such as shear and divergence) requires accurate comparison of velocity in adjacent neighborhoods. Computing small velocity differences between large velocities demands more resolution that is available in the second- and third-order motion systems.

See also: Fourier Analysis; Motion Perception: Psychological and Neural Aspects; Sensation and Perception: Direct Scaling; Statistical Pattern Recognition; Vision for Action: Neural Mechanisms; Vision, High-level Theory of; Vision, Low-level Theory of; Vision, Psychology of; Visual Space, Geometry of

Bibliography

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: Optics and Image Science* **2**: 284–99

Boring E G 1942 *Sensation and Perception in the History of Experimental Psychology*. Appleton-Century-Crofts, New York

Braddick O 1974 A short-range process in apparent motion. *Vision Research* **14**: 519–29

Cavanaugh P, Mather G 1989 Motion: The long and the short of it. *Spatial Vision* **4**: 103–29

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: Optics and Image Science* **5**: 1986–2006

Dosher B A, Landy M S, Sperling G 1989 Kinetic depth effect and optic flow: 1. 3D shape from Fourier motion. *Vision Research* **29**: 1789–813

Exner S 1875 Experimentelle Untersuchung der einfachsten psychischen Processe. *Pflugers Archiv für Gesamte Physiologie des Menschen und der Tiere* **11**: 403–32

Fourier J B J 1822 *Théorie analytique de la chaleur*. Chez Firmin Didot, Paris

Kenkel F 1913 Untersuchungen über Zusammenhang zwischen erscheinungsgross und Erscheinumngsbewegung beim einigen sogenannten optischen Tauschungen. *Zeitschrift Psychologie* **61**: 358–449

Koenderink J J, van Doorn A J 1986 Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America A* **3**: 242–9

Korte A 1915 Kinematoskopische Untersuchungen. *Zeitschrift Psychologie* **72**: 193–206

Lee D M 1980 Visuo-motor coordination in space-time. In: Stelmach G E, Requin J (eds.) *Tutorials in Motor Behavior*. North-Holland, Amsterdam, pp. 281–95

Longuet-Higgins H C, Prazdny K 1980 The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B* **208**: 385–97

Lu Z-L, Lesmes L A, Sperling G 1999 The mechanism of isoluminant chromatic motion perception. *Proceedings of the National Academy of Science* **96**: 8289–94

Lu Z-L, Lesmes L A, Sperling G 1999 Perceptual motion standstill in rapidly moving chromatic displays. *Proceedings of the National Academy of Science* **96**: 15374–9

Lu Z-L, Sperling G 1995a Attention-generated apparent motion. *Nature* **377**: 237–9

Lu Z-L, Sperling G 1995b The functional architecture of human visual motion perception. *Vision Research* **35**: 2697–722

Lu Z-L, Sperling G 1996 Contrast gain control in first- and second-order motion perception. *Journal of the Optical Society of America A: Optics and Image Science* **13**: 2305–18

Reichardt W 1957 Autokorrelationsauswertung als funktionsprinzip des zentralnervensystems. *Zeitschrift für Naturforschung* **12b**: 447–57

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

van Santen J P H, Sperling G 1984 Temporal covariance model of human motion perception. *Journal of the Optical Society of America A: Optics and Image Science* **1**: 451–73

van Santen J P H, Sperling G 1985 Elaborated Reichardt detectors. *Journal of the Optical Society of America A: Optics and Image Science* **2**: 300–21

Solomon J A, Morgan M J 1999 Dichoptically canceled motion. *Vision Research* **39**: 2293–7

Sperling G, Landy M, 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**: 426–40

Ullman S 1979 *The Interpretation of Visual Motion*. MIT Press, Cambridge, MA

Ullman S 1984 Maximizing rigidity: The incremental recovery of 3-D structure from rigid and non-rigid motion. *Perception* **13**: 225–74

Vaina L M, Cowey A 1996 Impairment of the perception of second order motion but not first order motion in a patient with unilateral focal brain damage. *Proceedings of Royal Society of London B* **263**: 1225–32

Vaina L M, Makris N, Kennedy D, Cowey A 1998 The selective impairment of the perception of first-order motion by unilateral cortical brain damage. *Visual Neuroscience* **15**: 333–48

Wallach H, O'Connell D N 1953 The kinetic depth effect. *Journal of Experimental Psychology* **45**: 205–217

Watson A B, Ahumada A J 1985 Model of human visual-motion sensing. *Journal of the Optical Society of America A* **1**: 322–42

Watson A B, Ahumada A J, Farrell J E 1986 The window of visibility: A psychophysical theory of fidelity in time-sampled motion displays. *Journal of the Optical Society of America A* **3**: 300–7

Wertheimer M 1912 Über das Sehen von Scheinbewegunen und Scheinkorpern. *Zeitschrift Psychologie* **45**: 205–217

G. Sperling

International Encyclopedia of the Social & Behavioral Sciences

ISBN: 0-08-043076-7