Spatial Vision, Vol. 1, No. 1, pp. 47–56 (1985)

Three theories of stroboscopic motion detection

GEORGE SPERLING,* JAN P. H. VAN SANTEN† and PETER J. BURT‡
Psychology Department, New York University, 6 Washington Place, New York, NY 10003, USA

Received 15 December 1984; accepted in revised form 8 March 1985

Abstract—The three theories derive from three different paradigms. Suprathreshold judgements of perceived quality of motion in multi-flash displays are modelled by space–time Fourier analysis of the motion stimulus. Stroboscopic motion is perceived as being different from real motion to the extent that the additional Fourier components in stroboscopic motion are detectable. Stroboscopic motion of dots along conflicting paths leads to perceptual competition. The theory to describe perceptual resolution derives and proves the uniqueness of strength functions computed only from the time and from the distance between successive points on each path. Time-strength and motion-strength add to determine path-strength; only the strongest path is perceived. Motion-direction detection in continuously drifting two-flash combinations of sinusoidal gratings is described by elaborated Reichardt detectors (ERDs) that compute the covariance of temporal events in two adjacent locations. Other, apparently different, detectors that account for direction-detection data are shown to be equivalent to ERDs.

COMPARISONS OF TWO- AND MULTI-STIMULUS MOTION DISPLAYS

The psychological study of stroboscopic motion perception, as it is taught in psychology courses and described in our texts, begins with Exner (1875). Working in Helmoltz's laboratory, he published experiments using the two-stimulus method for apparent motion that Gestalt psychologists (e.g. Wertheimer, 1912) later made a cornerstone of their system. In the two-stimulus display, a point is flashed briefly at Location 1 and, after a short interval is flashed at Location 2. The time interval between flashes is Δt, the spatial separation is Δx. Although the two-stimulus experiment has dominated psychologists' thinking about motion perception, a brief survey of the practical applications of motion principles in motion pictures, television, computer displays, video telephones and the like, indicates that all these systems use many- not two-flash presentations.

Remarkably, in spite of the extensive psychophysics of two-stimulus motion, no comparable data existed for multi-stimulus motion. Therefore, Miriam Kaplan, a graduate student at NYU, and G. Sperling set out upon a simple set of experiments in which subjects were asked to rate the quality of perceived motion of a point of light in two-stimulus and multi-stimulus displays. Subjects used a rating scale of zero (for no apparent motion) to ten (for apparent motion indistinguishable from real motion). They were presented with a variety of computer generated displays in which Δx and Δt were varied in all combinations, both in two- and in multi-stimulus presentations (see Fig. 1). Luminous points were flashed briefly on a dim background,

*The whom all correspondence should be addressed.
†Now at AT & T Bell Laboratories, Summit, New Jersey.
‡Now at RCA Sarnoff Research Laboratories, Princeton, New Jersey.

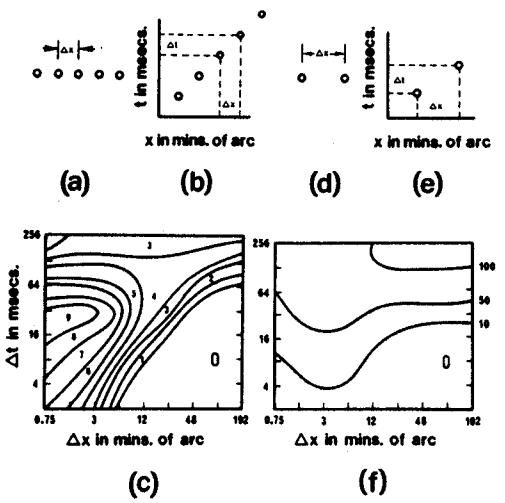

Figure 1. Judged quality of stroboscopic dot motion as function of the time Δt and distance Δx between successive points along the motion path. (a–c) Multi-dot presentations: (a) spatial arrangement; (b) spatio-temporal arrangement; (c) data for one, typical, subject. The contour lines indicate the judged quality of motion on a scale of 0 (no motion) to 10 (indistinguishable from real motion). (d–f) Two-dot presentations: (d) spatial arrangement; (e) spatio-temporal arrangement; (f) data. The contour lines indicate the proportion of trials on which non-zero (motion) judgements were made. In the area labelled 0, the fraction of non-zero motion judgements was less than 10%. (After Sperling, 1976.)

the direction of motion on any trial was random and, for multi-stimulus displays, the total path length was the same for all displays.

In multi-stimulus displays, the subjects used all ratings except ten, depending on the combination of Δx and Δt. The most significant finding was that whenever a motion path was made to more closely approximate real motion by doubling the number of points along the path (i.e. by halving Δx and Δt), the rated quality of motion always increased. The most interesting result of the two-stimulus experiments was that, once subjects had seen the multi-stimulus displays, they virtually never used a rating of more than one for the two-stimulus displays. Therefore the two-stimulus motion data are presented in terms of the proportion of non-zero motion responses. The highest proportion of non-zero motion responses occurred with large Δx and Δt. In fact, these data are essentially equivalent to those of Neuhaus' (1930) classic study of motion in two-stimulus point displays. In multi-stimulus displays, the best motion was seen with small Δx and Δt, directly the opposite result of the two-stimulus experiments.

Subsequently, Morgan (1979) and Watson *et al.* (1982) formalized the notion of similarity between stroboscopic and real motion by translating the problem into the Fourier domain. A multi-stimulus display and a continuous motion display of the same velocity have the same fundamental Fourier component, they differ only in that the stroboscopic display has many higher harmonic Fourier components. The more points there are along the multi-stimulus path, the higher is the spatial and temporal frequency of the lowest of these harmonics; when the frequency of even the lowest harmonic exceeds the resolving power of the visual system, the stroboscopic and continuous motion are visually indistinguishable. The theory that stroboscopic motion can become sufficiently similar to real motion to be indistinguishable from it is obvious

enough; however, it is only in the Fourier domain that a useful general metric of similarity is available.

Multi-stimulus motion has a reasonable psychophysics and the beginnings of a quantitative theory. Two-stimulus motion is a different matter because Fourier theory does not apply in an obvious way. A number of researchers have commented on the duality of motion processes (e.g. Pantle, and Picciano, 1976; Braddick, 1974), although whether the motion processes are distinguished by being *long range* and *short range* (as Braddick proposed) is not so clear. The moral of the story is that two-stimulus psychophysics of isolated points or bars in apparent motion is uninformative about the multiple stimuli that occur in practical applications, and the attempt to apply two-stimulus psychophysics to multi-stimulus presentations would lead to diametrically incorrect conclusions about Δx and Δt. This moral about the danger of incorrectly overgeneralizing from a paradigmatic experiment has relevance beyond the domain of motion perception.

AMBIGUOUS MULTI-STIMULUS MOTION DISPLAYS

The problem with subjective judgements of the quality of stroboscopic motion is that they confound several different possible cues, not all of which are directly related to motion. For example, a subject may observe that there is flicker in the display, or that a dot appears to remain on after another dot has flashed, or that a dot remains stationary and then moves, and so on. These cues may be, at least in part, computed by other systems than the motion system. However, they play a significant role in evaluating the quality of perceived motion. Thus, for deriving a theory of a visual motion process, it would be desirable to obtain pure, unconfounded judgements of motion even though, ultimately, it would be desirable to have a theory that dealt with all the complicated interactions of motion and other systems. A stroboscopic display that seemed not to involve any cognitive factors was developed by Burt and Sperling (1981) together with a corresponding motion theory.

The Burt–Sperling displays are illustrated in Fig. 2. A row of dots flashed on a computer screen; the spacing between dots is x. In the next flash, Δt later, a second row of dots is flashed Δy below and Δx to the right of the first row. This process continues until the dots disappear at the bottom of the screen, at which point it repeats from the beginning. When Δx is about $0.4x_0$, and Δt is large (e.g. 60 ms), then the rows of dots

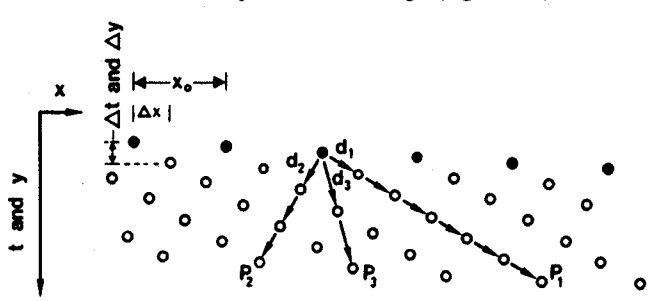

Figure 2. An ambigous motion stimulus. Rows of dots are flashed successively at intervals of Δt and are separated vertically by Δy. (Time, t, and vertical position, y, are perfectly correlated in this display.) The distance between adjacent points on the same row is x_0; successive rows are displaced laterally by Δx. Three competing paths P_1, P_2, P_3 are indicated. The distance between successive points on path i is d_i and the time is Δt_i.

appear to march down and to the right. However, when Δt is small (e.g. about 15 ms), then the rows of dots appear to move down and to the left. The same geometric display leads to opposite directions of apparent movement depending upon Δt. The perceived direction of movement can be influenced by tracking a movement path with the eyes; it cannot be voluntarily influenced when the eyes maintain stable fixation. By the judicious removal of dots along one path or the other, it can be shown that, even while motion in only one direction is being perceived, a perfectly adequate stimulus exists for motion in the other direction. The subject simply is unaware of the alternative perceptual possibility.

By varying the between-flash time Δt while keeping the geometry fixed, a 'balance point Δt_{ij}' can be determined at which the alternative perceptual modes i and j have equal probability of being seen. At Δt_{ij}, the *strengths* of perceptual modes i and j are equal. By varying the geometry of the display $(x, \Delta x, \Delta y)$, and determining the balance point Δt_{ij} between alternative modes, a great many pairs of paths of precisely equal strength can be obtained. From such data, the following principles emerged:

1. *Scale invariance.* If two paths are in balance, then changing the viewing distance (i.e. changing the scale of the display) leaves the balance unperturbed. Scale invariance breaks down when successive points along a path are separated by less than about 6 min of arc (for parafoveal viewing).

2. *Shape indifference.* If two paths are in balance, they remain in balance when the shape of the elements is changed. By arranging successive points along a path so that they alternate in shape (e.g. dots and rings, or right slanting line segments and left slanting segments) such a path is not weaker than a path with all elements having the same shape (cf., Kolers, 1972).

3. *Log–linear Δt versus d.* Balance points Δt_{ij} were determined for 36 displays with different geometries. Let the distance of successive points along a path i be d_i. The Δt_{ij} were discovered to be related to the d_i and d_j by $\Delta t_{ij} = A - B \log (d_i/d_j)$.

From these three observations, and the assumption that only time and distance of successive points along a path enter into the computation of its perceptual strength, Burt and Spreling (1981) were able to derive and to prove the uniqueness of a simple additive theory to account for their data. The motion strength S_i of Path i is an increasing monotonic function, M, of the sum of two component strengths: *time* strength $(a \log \Delta t - \Delta t)$ and *distance* strength $(-b \log d_i - c/d_i)$, where a, b, c are positive constants. [The equivalent multiplicative form of the theory asserts that $S_t = M(t^{\alpha\beta} e^{\beta t - \gamma/d}/d)$ where $\beta = 1/b$ and $\gamma = -\beta c$.] Here, c/d_i represents the correction factor for small distances that becomes negligible for $d_i \gg 6$ min. Except for very small d_i, distance strength is simply inversely proportional to $\log d_i$, which follows directly from scale invariance. The parameter b represents the relative importance of time to distance. The temporal factor is a unimodal function with a maximum strength at Δt of 18 and 24 ms for the two subjects (determined by parameter a). This theory, with only three parameters for each subject, predicts the outcomes of the 36 experiments extremely well.

Two noteworthy observations are:

1. This strength theory exemplifies a principle of *additive evidence* proposed by Sperling *et al.* (1983) as a general principle for the perceptual resolution of ambiguous, multi-stable displays. Surprisingly often it has been possible to formulate a theory in which the various factors (the sources of evidence) that contribute to the strengths of

the various competing perceptual modes simply add (without interactions) to determine the outcome.

2. The Fourier analog to the Burt–Sperling display is a combination of two sine-wave gratings oriented at an angle to each other and drifting with different velocities. These displays were investigated by E. H. Adelson while he was a postdoctoral fellow in the laboratory of G. Sperling, but no quantitative account of the results was discovered (Adelson and Movshon, 1982).

A GENERAL THEORY OF STROBOSCOPIC MOTION VIA ELABORATED REICHARDT DETECTORS (ERDs)

While the Burt–Sperling theory gives an excellent description of the outcome of certain kinds of ambiguous dot motion, it cannot readily be generalized to any other kind of display. For general prediction, a *process* rather than a *descriptive* theory is needed. Remarkably, a general theory of insect motion detection was developed by Reichardt (1957). With some elaborations, this same theory was quite successfully applied to the human detection of continuous motion (van Santen and Sperling, 1983a, b, 1984b). The subsequent application of the elaborated Reichardt model to stroboscopic motion will be described below.

The basic principle that most theorists who have grappled with the problem of motion detection have been driven to is a *delay-and-compare* principle. A Reichardt motion detector receives light input from two nearby points of the visual field, designated for convenience as the left and right inputs. To detect rightward motion, input arriving at the left point is delayed (by a general time-invariant linear filter) and then compared with the *un*delayed right input. When the delay of the object travelling from left to right in the external world matches the internal delay, the left and right inputs to the comparator contain matching signals.

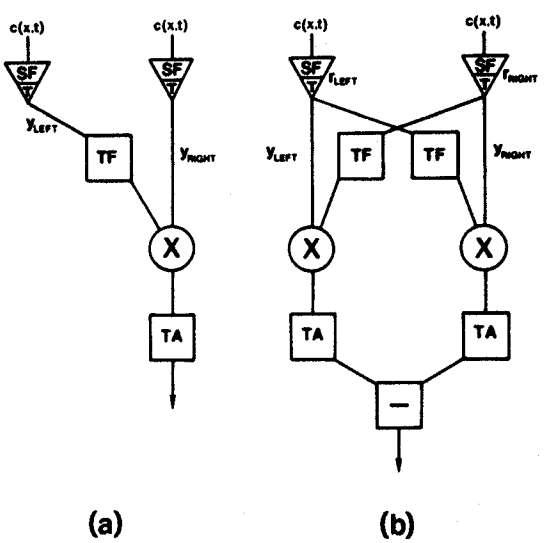

Figure 3. (a) The right subunit of an elaborated Reichardt detector (ERD). The luminance input is $c(x, t)$: *SF* denotes the left and right spatial receptive fields; *TF* is a temporal delay filter; × denotes the comparison multiplier; *TA* represents time averaging. (b) An ERD consisting of mirror symmetric left and right subunits whose outputs subtract to produce ERD output. (After van Santen and Sperling, 1985.)

The comparator in the Reichardt detector is a multiplier. It is followed by a time-averaging component (Fig. 3a). [In the elaborated Reichardt model of van Santen and Sperling (1985), for periodic displays, infinite time-averaging, which is unrealistic, is replaced by averaging over an integral number of periods of the display, which is exactly equivalent. For temporally confined displays, such as two-flash or *n*-flash displays, time-averaging is confined to the finite period of the display and its immediate aftermath.] The output of the time-averager is, in fact, the covariance between the two inputs to the comparator. Covariance is a comparison operation of great statistical efficiency that is optimal in many circumstances. Its evolution in the visual motion system is quite plausible.

What has been described so far, is only the right subunit of a Reichardt detector. To account for human motion detection, the subunit must receive input not merely from a point but from an area, a *receptive field*, an elaboration developed in detail by van Santen and Sperling (1984b). With most choices of receptive fields and temporal filters, an ERD subunit will be fooled into giving positive responses to non-moving stimuli, such as flickering fields and the like. To solve this problem (and simultaneously many others) a Reichardt detector contains a mirror image left subunit and the outputs of the two subunits are subtracted to determine the detector's motion response. Positive outputs indicate rightward motion; negative outputs, leftward motion (Fig. 3b). Some very general constraints on the receptive fields and temporal filters of which an ERD is composed suffice to enable it to correctly classify the direction of motion of any sinusoidal input (van Santen and Sperling, 1984b).

To develop a Reichardt detector into a model of human motion perception, requires combining the outputs of many local detectors according to a *voting* rule to determine the ultimate response of the entire system. Van Santen and Sperling considered various rules (addition of outputs, selection according to maximum, etc.) and developed theorems for their elaborated Reichardt *model* that were valid independently of the particular voting rule. (The term detector applies to a single unit; the term model is reserved for an aggregation of detectors.)

The most astounding properties of an elaborated Reichardt detector (ERD) concern its response to sinusoidal inputs, such as those that would be produced by drifting sinewave gratings.

Ninety-degree rule. The ERD response to a drifting sinusoid is always at a maximum when the left and right inputs differ in temporal phase by 90°. This is true for every ERD no matter how the temporal delay filter or other filters are chosen.

Segregation of sinusoids. When the input is composed of two or more drifting sinusoids of *different* temporal frequency, the output consists simply of the sum of the outputs that each of the sinusoids would have produced individually. Considering that the comparator performs a multiplication operation (making the ERD an extremely nonlinear system) the apparent non-interaction of sinusoidal inputs is quite unexpected. However, when two component sinusoids have the *same* temporal frequency, interactions can be extremely nonlinear.

The ninety-degree rule and frequency segregation are not the only interesting properties of ERDs, but these are sufficient to derive some very provocative predictions. For example, and ERD gives zero response to any stationary pattern. Therefore, any stationary pattern can be added to a moving grating, and the detectability of the direction of movement of the grating will not be impaired. This prediction was tested by van Santen and Sperling (1984b) for a variety of stationary patterns, even sinusoids

with the same *spatial* frequency as the drifting sinusoid and was verified in all instances. Similarly, an ERD gives zero response to any spatially-uniform flickering field. Therefore, adding a uniform flickering field to a moving grating should not impair detectability of the grating's direction of movement, provided that the *temporal* frequencies of the two stimuli are different. When the flickering field and drifting sinusoid have the same temporal frequency, the property of frequency segregation no longer applies and anything is possible, even reversal of apparent direction. These predictions also were tested and verified.

The two experiments just described are spatial–temporal duals, that is, experiments in which the space and time relationships are interchanged. In most of these instance of duals (adding a stationary grid to a drifting grating and adding a homogeneous flickering field to a drifting grating) the results were the same: no interference with discrimination of the direction of movement. However, in one pair of instances, profoundly different results were observed:

(a) added stationary grids (zero temporal frequency) but with the same *spatial* frequency as the drifting grating did not impair discrimination of its direction of movement,

(b) added uniform fields (zero spatial frequency) but flickering with the same *temporal* infrequency as the drifting grating caused reversal of apparent direction (total interference).

These results underscore the asymmetry of space and time in the ERD.

An ERD compares the temporal patterns that occur in neighboring spatial locations (by computing their covariance). It does not compare the spatial patterns that occur in successive frames, the principle of most computer models of motion perception, a principle that is thoroughly disproved for human motion perception by the experiments described above. Nor is an ERD symmetric in space and time like spatio-temporal Fourier analysis, a computation that is disproved for human motion perception by experiments to be described below.

The ERD and stroboscopic motion

The general solution for the response of ERDs to multiframe displays (including stroboscopic motion) was derived by van Santen and Sperling (1985). The astonishing aspect of the ERD's response is that spatial and temporal aspects of the response are completely separable. For two-flash displays, the spatial component of ERD output y_{spatial} is especially simple: $y_{\text{spatial}} = R_{\text{right},1}(x)R_{\text{left},2}(x) - R_{\text{right},2}(x)R_{\text{left},1}(x)$, where R indicates the response of the receptive field of an ERD at location x; (*right, left*) indicates the receptive fields; and (1, 2) indicates the flash. Together, the spatial and temporal components of the ERD's response determine its output.

Van Santen and Sperling (1984a, 1985) derived and tested ERD predictions for one-dimensional two-stimulus presentations consisting of simple spatial sinusoids, pairs of sinusoids, sinusoidal triples, and random bar gratings. ERDs respond differently to these stimuli depending on their relative size (their *scale*) and on their location relative to landmarks of the stimulus. For example, random-bar two-stimulus presentations are correctly analysed according to the direction of movement by detectors at most locations for a midrange of detector sizes, but detectors with very small receptive fields

respond almost randomly with respect to the direction of movement, and detectors with very large receptive fields respond hardly at all. This agrees with subjects' impressions in viewing these presentations: there is local jitter but overall movement.

With double sinusoids composed of spatial frequencies f and $3f$, the optimal response occurs when each components is displaced 90° of its spatial frequency. Rigid displacements (in which each sinusoid is displaced by the same physical distance and thus by a phase that is proportional to its spatial frequency) produce ambiguous responses, small detectors reporting one direction, larger ones the other. In viewing f and $3f$ presentations, subjects report better, 'even more rigid' movement in the optimal nonrigid displacements than in rigid displacements.

Finally, two-flash motion was studied with stimuli composed of triple sinusoids. The sinusoids in both flashes were combined either in amplitude-modulation phase or in quasi-frequency-modulation phase, but the two kinds of two-flash stimuli were equivalent in all other respects. These internal phase relations between stimulus components were critical both for the ERD and for human perception. The direction of motion was most reliably discriminated with amplitude-modulation phase. Fourier analysis of the stimuli would yield no suggestion of why some phase relations between components aid motion-direction discrimination and other phases impair it. For such predictions, a phase theory is needed; the ERD is the only detector to make the correct predictions.

Taken all together, these results establish the ERD as essentially duplicating the direction discrimination responses of the human short-range motion system to the one-dimensional stimuli for which the ERD was developed and tested. The theory has not yet been fully developed for two-dimensional patterns, such as the dot patterns studied by Burt and Sperling. Another unresolved issue is why direction discrimination of motion in sinusoidal and random-bar two stimulus presentations conforms to multi-stimulus predictions, whereas the subjective judgement of motion quality in two-stimulus dot presentations of Section 1 does not.

Different versions of the ERD

Watson and Ahumada (1983) propose a motion mechanism based on a comparison operation consisting of simple addition. Their mechanism has many properties that are similar to the ERD but is based on a totally different principle. However, it is not yet a detector. When van Santen and Sperling (1985) elaborated Watson and Ahumada's mechanism to make it into a detector, they discovered that this detector was actually equivalent to an ERD. It performed all the same operations as an ERD, merely the order of computation was different. The equivalence to an ERD of the directional motion energy detector proposed by Adelson and Bergen (1984) was similarly proved. Finally, van Santen and Sperling showed that, with carefully chosen filters, even a subunit of an ERD could be fully equivalent to an entire ERD.

With respect to the physiological basis of the multiplication operation of the ERD, Thorsen (1966) showed that it could be closely approximated by shunting inhibition (Furman, 1965; Sperling and Sondhi, 1968), and Torre and Poggio (1978) demonstrated that many forms of nonlinear physiological interaction could mimic the multiplication operation of the ERD because the term xy occurs in the Taylor series that describes the nonlinearity. Thus it appears that the ERD's computation is quite robust: it can be carried out in a number of different ways and by a variety of different components.

SUMMARY AND CONCLUSIONS

The first paradigm involved judgements of the perceived quality of motion in stroboscopic dot displays. The conclusion was that responses to two-flash and to multi-flash displays were described by quite different laws, and probably were governed by different processes. The preliminary theory asserted that the perceived similarity of multi-flash displays to continuous motion was governed by the similarity of the low frequencies in their spatio-temporal Fourier transforms.

The second paradigm involved constructing ambiguous motion displays an determining the balance point at which the competing perceptual modes had equal strength. The theory provided a unique description of strengths in terms of separable factors due to the time Δt and to the distance Δx of successive points along the paths. While this theory had great predictive power, it did not readily generalize beyond the domain of dot displays.

The third paradigm involved motion displays that were composed of component stimuli, some of which did not move or moved with different velocities and directions than other components. Both two-flash and multi-flash presentations were investigated. The theory was formulated in terms of an elaborated Reichardt detector (ERD), a system composed of two mirror symmetric subunits, each of which receives information from undelayed inputs, and then time-averages to compute the *covariance* between the inputs. The computation performed by ERDs can be computed by detectors composed of quite different components and also by apparently quite differently composed detectors that, upon analysis, turn out to be equivalent to ERDs. In circumstances where motion direction discrimination can reasonably be assumed to involve the monocular short-range motion system, ERDs account for the main facts of motion direction discrimination in both simple and complex visual displays.

Acknowledgement
Much of this research and the preparation of this article was supported by AFOSR Grant 80-0279.

REFERENCES

Adelson, E. H. and Bergen, J. (1984). Motion channels based on spatio-temporal energy. *Investigat. Ophthalmol. Visual Sci.* **25**, 14 (abstract).

Adelson, E. H. and Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature* **300**, 523–525.

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

Burt, P. and Sperling, G. (1981). Time, distance and feature trade-offs in visual apparent motion. *Psychol. Rev.* **88**, 171–195.

Exner, S. (1875). Experimentelle Untersuchung der einfachsten psychischen Processe. *Archiv Gesamte Physiol. Mensch. Tiere* **11**, 403–432.

Furman, G. G. (1965). Comparison of models for subtractive and shunting lateral-inhibition in receptor-neuron fields. *Kybernetik* **2**, 257.

Kolers, P. (1972). *Aspects of Motion Perception.* Pergamon Press, New York.

Morgan, M. J. (1979). Perception of continuity in stroboscopic motion: a temporal frequency analysis. *Vision Res.* **19**, 491–500.

Neuhaus, W. (1930). Experimentelle Untersuchung der Scheinbewegung. *Archiv Gesamte Pyschol.* **75**, 315–458.

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

Reichardt, W. (1957). Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. *Zh. Naturforchung* **12b**, 447–457.