Temporal Coherence Theory for the Detection and Measurement of Visual Motion

NORBERTO M. GRZYWACZ,*† SCOTT N. J. WATAMANIUK,* SUZANNE P. MCKEE*

Received 9 May 1994; in revised form 15 March 1995

A recent challenge to the completeness of some influential models of local-motion detection has come from experiments in which subjects had to detect a single dot moving along a trajectory amidst noise dots undergoing Brownian motion. We propose and test a new theory of the detection and measurement of visual motion, which can account for these signal-in-Brownian-noise experiments. The theory postulates that the signals from local-motion detectors are made coherent in space and time by a special purpose network, and that this coherence boosts signals of features moving along non-random trajectories over time. Two experiments were performed to estimate parameters and test the theory. These experiments showed that detection is impaired with increasing eccentricity, an effect that varies inversely with step size. They also showed that detection improves over durations extending to at least 600 msec. An implementation of the theory accounts for these psychophysical detection measurements.

INTRODUCTION

Two of the seminal models in psychophysics are the Motion-Energy and Elaborated-Reichardt models (Hassenstein & Reichardt, 1956; van Santen & Sperling, 1984; Adelson & Bergen, 1985; Watson & Ahumada, 1985). They propose mechanisms for the elementary visual-motion detectors in the brain. These detectors respond more strongly to motion in a particular direction than to motion in other directions, a property called directional selectivity. Moreover, these detectors can only respond to motions confined to their receptive fields. Optimally, the motions should occur during the motion detectors’ integration time, typically between 80 and 120 msec (Watson & Nachmias, 1977; Burr, 1981; Mikami, Newsome & Wurtz, 1986). These models can successfully account for a wide range of phenomena in the perception (van Santen & Sperling, 1984; Anderson & Burr, 1989, 1991; McKee & Watamaniuk, 1994) and physiology (Movshon, Adelson, Gizzi & Newsome, 1985; Reid, Soodak & Shapley, 1987, 1991; Emerson, Bergen & Adelson, 1992) of visual motion.

However, at best, the Motion-Energy and Elaborated-Reichardt models only provide a partial description of motion perception. A combination of the output of many of the motion detectors postulated by these models is necessary to estimate velocity (Heeger, 1987; Reichardt, Egelhaaf & Schüllig, 1988; Grzywacz & Yuille, 1990, 1991). Furthermore, it may take up to 400 msec for humans to discriminate motion speed or direction precisely under noisy or jittery conditions (Snowden & Braddock, 1989; Watamaniuk, Sekuler & Williams, 1989; Werkhoven, Snippe & Toet, 1992), a time longer than the integration time of the postulated motion detectors. The visual system seems to integrate motion information over distances larger than the size of the elementary motion detectors to give rise to phenomena related to the Gestaltists’ Common-Fate law (Koffka, 1935; Nakayama & Tyler, 1981; Ramachandran & Anstis, 1983a; Williams & Sekuler, 1984). This long-distance interaction helps to disambiguate local-motion measurements of two-dimensional stimuli, as for instance, when plaids (Ko0i, De Valois, Switkes & Grosof, 1992) or wiggly lines (Nakayama & Silverman, 1988a, b; Yuille & Grzywacz, 1989a) are flanked by unambiguous motions.

A serious challenge to the completeness of the Motion-Energy and Elaborated-Reichardt models comes from experiments on the detection of a single dot moving in a constant trajectory amidst dots undergoing a Brownian motion (Watamaniuk, McKee & Grzywacz, 1994). In these experiments, the signal dot was indistinguishable from the noise based on one or two frames of the motion. To ensure this, each noise dot moved with the same step size and frame rate as the signal dot. Therefore, the only way to detect the signal dot was to detect its extended motion. But the motion was such that there was a high probability that a noise dot in frame "i"...
could be closer to the signal dot in frame "i-1" than the signal dot itself in frame "i." This probability is known as the probability of mismatch (Williams & Sekuler, 1984; Watamaniuk et al., 1994). Consequently, it was almost certain that a mismatch would happen after a few frames. Nevertheless, subjects could often detect the signal dot easily. Detection was easy even when the signal dot moved in a circular trajectory, which changed its direction of motion over the integration time of a single local-motion detector by more than the detector's directional bandwidth.

These results argue against detection being limited by a single, large local-motion detector and suggested that motion signals from several local-motion detectors are combined over time (Watamaniuk et al., 1994). This temporal combination appeared to be more powerful than the spatial one. While about 6% of the dots in a display must move coherently in space to allow detection of spatially global motion (Newcombe & Paré, 1988; Vaina, Grzywacz & LeMay, 1990), temporal integration allowed the detection of 1 dot amidst at least 200 noise dots (0.5% signal-to-noise ratio).

Watamaniuk et al. (1994) performed experiments to determine what controls the detection of the signal dot and described four main findings: (1) Detectability could not be accounted for by the density of the noise dots or the hop size of the motions alone. Probability of mismatch, which is a combination of density and hop size, was the key independent variable determining detection. (2) Detection was not solely based on positional cues, such as the alignment of the signal dot over time, but required the motion itself. For instance, if the noise dots moved within a range of directions of 180 deg, then when the signal dot moved opposite to the middle of this range, detection was almost perfect. (3) Detection did not depend on the trajectory being straight. Subjects performed well if the signal dot deviated from a straight trajectory by wobbling or moving in a circular trajectory. (4) Probability of detection increased over time for at least 300 msec, which again is longer than the integration time of local-motion detectors. Poor performance resulted, if the trajectory was broken into non-continuous pieces.

Of these findings, perhaps the most surprising is the apparent dependence of performance on probability of mismatch. This quantity was first introduced by Ullman (1979) in the context of his Minimal Mapping theory. Ullman proposed this theory to account for the phenomenology of long-range apparent motions, that is, apparent motions whose step sizes and intervals are large (Biederman-Thorson, Thorson & Lange, 1971; Bradick, 1974, 1980). The motivation for this theory was Ullman's insight that the main problem the visual system must solve when measuring motion is the correspondence problem. This problem is to determine what image feature in frame "i-" corresponds to an image feature in frame "i+1." The Minimal Mapping theory was proposed to solve the problem by matching features in consecutive frames such that the total displacement of the features was minimal. While this theory seemed to account successfully for a variety of long-range phenomena, it did not seem appropriate to deal with short-range motions, for which the correspondence problem may not be serious. Adelson and Bergen (1985) argued that for these motions, the frames are close (perhaps even continuous) and can be dealt with by receptive-field like mechanisms. In this case, how could the probability-of-mismatch result of Watamaniuk et al. (1994), which seems to indicate a correspondence problem, be explained given that their experiments were performed with short-range parameters?

In this paper, we extend previous theoretical work to reconcile the probability-of-mismatch results with Adelson and Bergen's arguments, and to account for the other data of Watamaniuk et al. The new theory begins with local-motion detectors such as those postulated by the Motion-Energy and Elaborated-Reichardt models. Then the theory computes the energy of particular local speeds with a modified version of the Grzywacz and Yuille model (1990, 1991). After that, the theory integrates local-velocity signals over space and time in a manner similar, but not identical to that of Grzywacz, Smith and Yuille (1989). And finally, the theory performs an outlier-selection procedure to decide whether a special feature, such as Watamaniuk et al.'s signal dot, is present.

We implemented the new theory with a simplified neural network and compared its performance with the results of Watamaniuk et al. (1994). Additional psychophysical experiments were performed to estimate parameters and test the theory. These experiments are described in the Appendix. In the next section, we argue that models of local-motion detection cannot account for the Watamaniuk et al. (1994) data and present the new theoretical framework in detail. The section "Computer simulations" presents the results of computer simulations with an implementation of the theory and their fits to the data. Finally, we discuss the implications of the theory, and in particular, address the buildup of accurate velocity discrimination, and address ambiguous local motions, such as Non-Fourier and Second-Order motions.

THEORY

Can local-motion detectors account for the data?

In the "Introduction" we summarized the results of Watamaniuk et al. (1994) and argued that they were not easily explained by models describing local-motion detection.

We now bring another argument against the completeness of these models based on estimates of the quantity of noise in local detectors. Although the noise filling the screen in the Watamaniuk et al. experiments is impressive, the noise associated with each local-motion unit is small. For instance, if the diameter of the receptive field of the responding motion unit corresponds to the distance moved by the signal dot in 100 msec (the integration time of the unit), then for a
step size of 0.16 deg, a typical value in these experiments, the receptive field has an area of 0.32 deg\(^2\) (assuming a circular receptive field). At a noise density of 3 dots/deg\(^2\) (also typical of the experiments), about 1 noise dot falls in this unit every frame, corresponding to a signal-to-noise ratio of one. This ratio is an upper bound, since the area of the receptive field is smaller than the estimate above (motion receptive fields are probably elliptical with the long axes parallel to preferred direction) and most of the noise dots have little effect on the local-motion detector response. This is because these dots often fall in insensitive, fringe parts of the receptive field and because most of the motion vectors specified by the noise dots are apt to be orthogonal or opposite to the preferred direction of the unit. Therefore, considering that Newsome, Britten and Movshon (1989) reported that a correlated signal of 10% is detectable in noise both by human observers and single cells in primates' middle temporal cortex (MT), a single-motion-detector hypothesis would predict almost no effect of noise on performance.

To quantify this argument further, we simulated the responses of a model of motion-energy units to some of the stimuli of Watamaniuk et al. (1994). The simulations were parametric on probability of mismatch, which is

\[
p_m = 1 - e^{-nhd},
\]

where \(h\) is the hop size and \(d\) the density of the noise dots (Williams & Sekuler, 1984; Watamaniuk et al., 1994).

(To derive this equation, one assumes that the probability that a noise dot falls in a disk centered on the signal dot and with a radius \(h\) is small, and assumes that the number of noise dots is large. In this case, this probability follows a Poisson distribution.) The model first convolved the image with a space-time oriented Gabor filter (Gabor, 1946; Daugman, 1985), which can be written as

\[
G(\vec{x}, t) = \frac{1}{(2\pi)^{1/2}(\sigma_r \gamma)} \exp \left(-\frac{|\vec{r}|^2}{2\sigma_r^2}\right) \exp \left(-i\Omega t \vec{u} \cdot \vec{r}\right) \exp \left(-\frac{t^2}{2\sigma_t^2}\right) \exp \left(-i\Omega t, t\right),
\]

where \(\vec{r}\) is position relative to the center of the filters, \(t\) is time, \(\sigma_r > 0\) and \(\sigma_t > 0\) are positive parameters related to spatial scale and temporal integration time of the filter respectively, \(\Omega\), and \(\Omega_t\) are positive parameters corresponding to the optimal spatial and temporal frequencies, respectively, and \(\vec{u}\) is the unit vector of preferred direction. The model's output was the square of the absolute value of the filtering. A Gaussian additive noise, whose standard deviation corresponded to about 10% of the maximal amplitude at probability of mismatch = 0.3 was included in the model to ensure the robustness of the simulations. This model was insensitive to the spatial phase of the stimulus, since by using a complex exponential, the model employed a quadrature pair (Pollen & Ronner, 1981; Adelson & Bergen, 1985), that is, sine and cosine Gabor functions. Because we were interested in the possibility that large motion-energy units could account for some of the temporal integration, the main independent variable of the simulations was \(\sigma_t\). The models' parameters were chosen to be consistent with human and non-human primate literature. To have the integration time of the filter (a time covering 95% of the area under the Gabor's Gaussian) be 100 msec, we used \(\sigma_t = 25.6\) msec. The spatial- and temporal-frequency bandwidth of the filter were 1.5 (for a review see Graham, 1989) and 4 octaves (estimated from the transient-channel data of Kulikowski & Tolhurst, 1973), respectively. For a Gabor filter, the relationship between these bandwidths (\(b_x\)) and the respective optimal frequencies can be computed by expressions derived by Grzywacz and Yuille (1990) as

\[
\Omega = \frac{(\ln(2))^{1/2}(2b_x + 1)}{\sigma_x (2b_x - 1)}.
\]

For each \(\sigma_t\), we stimulated this motion-energy model with five hundred repetitions of Watamaniuk et al.'s paradigm. The signal dot underwent preferred- and null-direction motions. These motions lasted 5 steps (120 msec) and thus lasted longer than the motion-energy unit's integration time. The motions were also symmetrical about the center of the unit for optimal stimulation. To measure how well the model could detect the motion, that is, how well the model produces responses to one direction of motion that are significantly larger than responses to the opposite direction, we used the following estimator:

\[
\chi^2 = \frac{R_p - R_n}{\sqrt{|R_p - R_n|}} \frac{(R_p - R_n)^2}{R_p + R_n},
\]

where \(R_p\) and \(R_n\) are the maximal amplitudes of the preferred and null responses, respectively, the second fraction is a \(\chi^2\) measure around the mean, \((R_p + R_n)/2\), and the first fraction gives the sign of \(R_p - R_n\). We chose this measure of directional selectivity rather than the ratio between preferred- and null-direction responses (for example, Snowden, Treue, Erickson & Andersen, 1991) or the directional-selectivity index (Grzywacz & Koch, 1987), since the \(\chi^2\), but not these other measures of directionality, captures its statistical significance, that is, is sensitive to absolute response amplitude. Figure 1 shows the \(\chi^2\) results of the simulations for two probabilities of mismatch.

Peak directional selectivity with correct preferred direction occurred around \(\sigma_t\) (receptive field radius) of 2.3 steps. (According to equation 3, this \(\sigma_t\) corresponds to 3.15 c/deg for the typical step size of 0.24 deg). This receptive-field size (4.6 steps diameter) was close to what the signal dot would cover in 100 msec, the integration time of the filter. For smaller receptive fields, significant directional selectivity occurred down to radiiues of approx. 0.8 steps. For radiiues smaller than about 1.7 steps, aliasing (van Santen & Sperling, 1984, 1985) caused reversals of preferred and null directions. This aliasing was due to the skipping of excitatory or inhibitory lobes of the receptive field during null-direction
apparent motion with a step size longer than these lobes. Cells with even smaller receptive fields did not respond well to the motion, since they were so small that a moving dot would skip them altogether or touch them and then leave without generating a strong motion signal. In turn, for larger receptive fields, directional selectivity was close to zero due to the effect of noise. (The small, but detectable reversal of preferred and null directions for large receptive fields is not a general property of models of local-motion detection, but can occur in strict space-time oriented filters, such as the Gabor model. The reversal will happen when the receptive fields are so large that only a tiny fraction of their extent is stimulated during their integration time.)

Hence, significant directional selectivity occurs only for a limited range of spatial scales near to the motion’s hop size. [For the simulations in the rest of the paper, we used hop size and not 2.3 hop sizes, since at the present time, one should not place too much significance on the absolute value of the receptive-field size where peak directional selectivity occurs. This is because this size depends on the model chosen to implement local-motion detection. For instance, Motion-Energy-like models with impulse responses comprising a delayed short-lived excitation followed by a strong, prolonged shunting inhibition would be unforgiving to noise dots. Therefore, these models would express optimal directional selectivity in the Watamaniuk et al. (1994) experiment at shorter (essentially noise-free) scales than those in Fig. 1.]

Arguably, the most important consequence of Fig. 1 arises from the finding that the theoretical peak directional selectivity was only slightly affected by probability of mismatch when it varied between 0.3 and 0.5. The only significant effect of increasing probability of mismatch was to reduce the amplitude of reversal at very large step sizes.

Because the experimental data showed a large drop in performance in the same range of probability of mismatch (from around 82 to 60% correct detection, Watamaniuk et al., 1994), the Motion-Energy model cannot account for the dependence of signal-dot detection on probability of mismatch, confirming the conceptual arguments above.

**Conceptual description of the temporal-coherence theory**

We now propose a new theory, which complements models of local-motion detection and can account for the Watamaniuk et al. data. To aid in the explanation of the theory, we use Fig. 2, which illustrates a stimulus at the top and the stages of the theory at the bottom. The stimulus panel of the figure shows one hop (two frames) of the type of motion used by Watamaniuk et al. (1994). While the signal dot (green arrow in the figure) moves in a straight trajectory with constant step size and speed, the noise dots (yellow arrows) move in a Brownian motion with the same step size as the signal dot. The long arrows labeled “Signal” and “Noise” indicate that in our explanation, we will use the signal dot and an exemplar noise dot to illustrate how the different stages of the theory work.

The theory comprises three hierarchical stages: The first, which we call the Local stage, estimates motion in localized regions of a multi-dimensional space whose independent variables are space, time, direction of motion, speed, and spatial scale (or spatial frequency). Except for time, we think of these measurements as being performed by cells tuned to particular values of these variables. The second stage, which we call the Coherence stage, smooths these local measurements in neighbor-
FIGURE 2. Schematic representation of the theory. To illustrate the theory, we use one of the protocols of Weltman et al. (1986). In it, the noise dots (yellow arrows) undergo Brownian motion with the same step size as the signal dot, while the signal dot (green arrow) moves in a straight trajectory (only one step of the motion is shown in the upper panel). The long arrows follow the responses to the signal dot and an exemplar noise dot. To represent the Local stage of the theory, we use eight 7 x 7 arrays of cells (white dots representing the center of the receptive fields) each tuned to a different preferred direction (indicated by the direction of the red arrows in each array). Cells responding to the local motion have an arrow attached to them. The representation of the Coherence stage of the theory uses the same eight arrays. However, this time the two preceding steps of the motion are shown for each dot by the black arrows (which do not represent activation levels of cells, but are there just to show the history of the dot in question). Only for the signal dot, the preceding steps are in the same direction as the present direction. Consequently, the responses to this dot (thick red arrows) are facilitated by temporal coherence. In this figure, the area of spatial coherence is indicated by the gray disk. Noise dots falling in this area can inhibit the responses to the signal dot.

Finally, in the Outlier stage of the theory, a statistical analysis over a broad area (dark-blue disk) is performed to detect an outlier response. In this analysis, the signal dot is often detected.
hoods in this multi-dimensional space. Consequently, this stage transforms the local multi-dimensional space into a new multi-dimensional space with the same independent variables. These variables’ values emphasize coherence of motion measurements along time and all other dimensions. The only exception for coherence occurs when two neighboring regions have very different responses in their local detectors. In this case, their responses are not mixed to allow for transparency and the detection of boundaries. As in the case of the Local stage, we think of the responses in the Coherence stage as being measured by cells tuned to particular values of the variables, except time. Finally, the third stage of the theory, which we call the Outlier stage, is of particular importance to the experiments of Watamaniuk et al. (1994) (although perhaps not necessarily important for general computations used in the analysis of motion patterns). It computes whether the distribution of responses in neighborhoods of a coherence cell contains an outlier. If this is the case, the model then reports this outlier as the perceived signal. As such, the outlier allows a comparison of human detection performance to simulated responses of the computational model. Finally, the neighborhoods in the Outlier stage are not necessarily the same as the neighborhoods in the Coherence stage. (Outlier or robust strategies have been used before in the context of computer vision, not to detect signals, but rather, to clean data for further analysis; see Geiger & Yuille, 1991, for a brief review.

**Local stage of the theory**

The Local stage of the theory begins with processes similar to those described in the context of the Motion-Energy and Elaborated-Reichardt models (for reviews see Grzywacz, Harris & Amthor, 1994; Mather, 1994). They include spatial and temporal preprocessing, and the generation of directionally selective signals through linear filters followed by a non-linear static operation such as squaring, multiplication, or rectification. Moreover, these stages may include a retinal rectification (Victor & Shapley, 1979; Grzywacz, Amthor & Mistler, 1990), which might be important to deal with some Non-Fourier motions (Chubb & Sperling, 1988; Turano & Pantle, 1989; Werkhoven, Sperling & Chubb, 1993) and may include signal normalization (Heeger, 1992, 1993). The output of these stages are combined linearly with an appropriate weighting function to build local units, whose responses are tuned to velocity (Heeger, 1987; Reichardt et al., 1988; Grzywacz & Yuille, 1990, 1991—see also Bravo & Watamaniuk, 1995, for psychophysical evidence for the computation of local speed). The only difference from previous models of these units is that, to account for motion transparency phenomena (Adelson & Movshon, 1982; Welch, 1989), the summation is in a neighborhood of spatial frequencies, rather than spanning all spatial frequencies, and also local velocities are not selected through a winner-take-all (see Smith & Grzywacz, 1993, for theoretical details). The signal in these units are like "likelihoods" of particular velocities and in this sense, resemble an early stage of a recent model by Nowlan and Sejnowski (1993, 1995).

In normal conditions, the range of spatial scales of cells which measure local motion is wide (Dow, 1974; Levinson & Sekuler, 1975; Anderson & Burr, 1987). However, in the discussion of Fig. 1, we argued that the design of the Watamaniuk et al. (1994) experiments is such that they confine the measurements of local motion to detectors with a limited range of spatial scales (receptive-field sizes) close to the motion's hop size. The behavior of directional selectivity within this range is complex and may include aliasing. In the "Discussion," we will argue that the new theory is particularly apt to disambiguate aliasing and other kinds of ambiguous motions. Hence, for the sake of simplicity, we will from now until the "Discussion" think of the Watamaniuk et al. experiments as tapping a single spatial scale close to the motion's hop size. In addition, since in these experiments, the time interval between consecutive image frames was constant, we will consider only units tuned to one speed.

In Fig. 2, the Local stage is represented by eight arrays of directionally selective cells (white dots), whose receptive fields are centered at 7 x 7 positions. In each array, the cells have one of eight preferred directions, indicated by the direction of the arrows in the arrays (45 deg separation between neighbor arrays). The main role of this stage is to decompose the stimulus into smaller chunks of data. Because in this paper, we wish to emphasize the Coherence stage of the theory, which is necessary to account for Watamaniuk et al.'s data, we will describe our choice of the Local-stage model in the Appendix.

**Coherence stage of the theory**

There are several motivations for the Coherence stage of the theory. Spatial integration of local-motion signals is important to account for motion cooperativity (Williams & Sekuler, 1984; Watamaniuk et al., 1989; Watamaniuk & Sekuler, 1992), motion capture (Ramachandran & Anstis, 1983a; Yo & Wilson, 1992), and changes of perceived speed with density modulations of moving textures (Watamaniuk, Grzywacz & Yuille, 1993). Furthermore, spatial smoothing of motion is important to reduce noise in the motion system (Yuille & Grzywacz, 1988, 1989a; Bülthoff, Little & Poggio, 1989) and deal with the aperture problem (Horn & Schunck, 1981; Marr & Ullman, 1981; Hildreth, 1984). Evidence for integration within neighborhoods is not only available for the spatial dimension but also for direction (Watamaniuk et al., 1989; Watamaniuk & Sekuler, 1992), speed (Watamaniuk & Duchon, 1992; Bravo & Watamaniuk, 1995), and spatial scale (Cleary & Bradtlick, 1990a, b). In the "Introduction," we listed several reasons why motion signals from several local-motion detectors are combined over time. To add to that list, we mention the motion-inertia phenomenon (Ramachandran & Anstis, 1983b; Anstis & Ramachandran, 1987), which indicates a role for immediately preceding motion in the
disambiguation of current motion signals (Grzywacz et al., 1989).

The cells in the Coherence stage of the theory try to be as consistent as possible with the input from the Local stage, with neighbor "coherence" cells, and with past information about the motion. In the next section, we will model this consistency requirement through a minimization of a energy function, a procedure that is something like a least-squares fit. As explained above, cells in a neighborhood in the Coherence stage code space, direction of motion, speed, and spatial scale. Because for paradigms such as that in Fig. 2 we are focusing on a single spatial scale and speed, we only illustrate the directional (45 deg) and spatial (open disk) neighborhoods.

There are some important differences in the manner the theory treats temporal coherence and coherence in the dimensions of space, direction of motion, speed, and spatial scale. To begin with, time information must be taken asymmetrically, that is, only the past is considered. Furthermore, it is unlikely that there are cells tuned to absolute time as there are cells tuned to absolute position or direction of motion. Therefore, temporal coherence is probably imposed by transmitting responses from place to place through delay lines.

In the Coherence stage of the theory, the responses of a cell "A" with particular preferred direction are made consistent with the past responses of other cells with roughly the same preferred direction and whose receptive fields are located in positions for which the preferred direction point roughly to the receptive field of "A." Because the past information is only "rough," temporal coherence does not require the motion to be in a straight trajectory, that is, small deviations of direction are tolerated. The optimal speed and spatial scales of these other cells are also roughly similar to those of "A." The distance from these cells to "A" and the delay in the transmission of information need not match the speed of the motion. Grzywacz et al. (1989) argued that only directional information should be taken into account, since for general three-dimensional translations, the direction, but not the speed, of the image motion is time invariant. Psychophysical evidence on motion inertia supports this argument (Anstis & Ramachandran, 1987), as does evidence that the visual system does not appear to take acceleration into account in the temporal integration of motion signals (Gottsdanker, 1956; Snowden & Braddick, 1989; Werkhoven et al., 1992). Hence, the theory also discards speed information and does not bother to implement temporal coherence optimally (in the sense of Kalman filters, Ayache & Faugeras, 1987; Matthis, Kanade & Szeliski, 1989; Clark & Yuille, 1990), but rather uses a strategy that is sufficiently good and that can be implemented easily with neural machinery (Ratliff, 1965; Grzywacz & Poggio, 1990). In this sense, temporal coherence is similar to temporal integration of motion signals implemented by some electronic (Delbruck, 1993) or developmental (Martin & Marshall, 1993) models of motion detection.

Figure 2 illustrates how the neural networks implementing the Coherence stage work. The Coherence-stage panels illustrate not only the current directional signals (red arrows), but also the two preceding steps of the motion (black arrows). The only dot for which the preceding steps are always in the same direction as the current step is the signal dot. For this reason, the responses to it can be large (represented by the thicker arrows), while the responses to the noise dots are typically small. These large responses occur because of facilitatory (or disinhibitory) connections between cells in the Coherence stage. Consequently, different than the Local stage where cells are relatively independent of each other, the cells in the Coherence stage are interconnected for purposes of motion processing. These interconnections are not all facilitatory, since cells with weak responses can exert an inhibitory influence on neighbors (gray disks in the figure) with strong responses. This inhibition occurs to close the gap between the weakly and strongly responding cells, that is, to smooth the responses in the coherence neighborhood. (A coherence cell with "zero" response does not inhibit a coherence cell responding to the signal dot, because the local responses of these cells are very different and thus, as mentioned in the introduction of "Conceptual description of the temporal-coherence theory," these responses are not mixed to allow for transparency and detection of boundaries.) In the "Discussion," we will address the biological plausibility of these interconnections and propose an intracortical network to implement them.

**Outlier stage of the theory**

The last part of the theory is the Outlier stage, which is necessary since in the Watamaniuk et al. (1994) experiments, the goal was to detect the signal dot, which in each time, had one different property from the noise dots. Very little data are available to constrain this stage. And in some respect, it involves a higher, cognitive process of the visual system, since its action could depend on the instructions given to the subject before the experiment. For instance, if the subject were told that the signal dot is going to pass through the center of the screen, then the search for the dot might span a much narrower area than if the subject were not given this information. Consequently, given the uncertainty about the Outlier stage, we will seek a solution that is adapted to the signal-vs-noise-dots paradigm of Watamaniuk et al.

If the brain has to decide that the responses elicited by the signal dot are "unusually" high, then it must compare this response to those elicited by other dots. How many dots or how large an area (dark-blue disk in Fig. 2) should be used in the comparison? The Watamaniuk et al. experiment in which the noise dots moved with a range of directions of 180 deg showed that detection improved if the signal dot moved in a direction opposite to the mean direction of the noise and deteriorated when these directions were equal. Imagine an extreme version of this experiment in which all noise dots always move in the same direction, that is, they
move in parallel straight trajectories. If the signal dot moves in the same direction as the noise dots, then it will never be detected, since the signal and noise dots have exactly the same motion parameters. In contrast, if the signal dot moves in the opposite direction, it will almost always be detected. Chance or almost perfect detection would occur in these conditions regardless of the density of the noise dots, and in particular, would happen at very low densities. Hence, the signal dot can be compared with noise dots at arbitrarily large distances. It is possible that the entire image is used in the comparison.

Another possibility is that the comparison area is somehow adjusted to contain a fixed number of noise dots. This possibility would have the advantage that the comparison samples would be similar from a statistical perspective and that the comparison areas would be smaller, allowing for more than one signal dot (not moving in the same direction) to be detected. For this reason, we adopted this strategy in our implementation of the theory.

A final consideration of the theory suggests that direct inputs from the Local stage should gate those cells from the Coherence stage that participate in the Outlier stage computations. Both temporal coherence and spatial coherence smear the signal dot's responses in the Coherence stage. These responses tend to have a "tail" and spread to neighbor cells. This tends to make the responses of coherence cells in the neighborhood of the signal dot more alike, and thus render outlier selection difficult. To alleviate this problem, a representative of each Coherence stage neighborhood could be selected for the outlier computation, such that this representative has the largest signals among its neighbors and its counterpart cell in the Local stage is firing. This is a local, gated winner-take-all computation, which can be implemented biologically (Hadedel, 1974; Elias & Grossberg, 1975; Yuille & Grzywacz, 1989b).

Size of spatial neighborhood

Now that all the stages of the theory have been described in conceptual terms, we can address an important constraint on the size of the spatial neighborhood of the Coherence stage. From the discussion of Fig. 1, we concluded that the dependence of performance on probability of mismatch is probably not due to the local-motion detectors.

We now argue that the probability-of-mismatch result of Watamaniuk et al. (1994) constrains the size of the spatial neighborhood of a cell in the Coherence stage to be proportional to the cell's spatial scale, that is, to the cell's receptive-field size. To understand this claim, recall that this result showed that performance does not fall independently with hop size or density of noise dots, but rather with a combination of these variables as expressed in equation (1). What this really means is that performance depends on a variable proportional to $h^2d$, where again, $h$ is the hop size and $d$ is density. But we argued above that the Watamaniuk et al. experiments essentially confined the operative local-motion detectors to those with a spatial scale near to the hop size. Consequently, performance depends on a variable proportional to the area of the receptive field multiplied by density, or the mean number of noise dots falling in the receptive field. Because performance is limited by the Coherence stage, the area of the receptive field of the local-motion detector cannot be important per se. It must thus be the case that the area of spatial neighborhood is proportional to the area of the local detector so that performance still depends on a variable proportional to $h^2d$. The important point is that the constant of proportionality should be the same for all directionally selective cells.

Mathematical formulation of the coherence stage

Because several authors have dealt with mathematical aspects of the Local stage of the theory (Hassenstein & Reichardt, 1956; van Santen & Sperling, 1984; Adelson & Bergen, 1985; Watson & Ahumada, 1985), we refrain from doing so here (see Appendix). Furthermore, because it is hard at the present time to formulate a general theory of the Outlier stage, we leave our choice for the implementation of this stage to the Appendix.

The mathematical formulation we adopt for the Coherence stage is a modification of that used by Grzywacz et al. (1989). Let us characterize directionally selective cells in both the Local and Coherence stages by the position of the middle of the receptive fields (f), preferred direction of motion (unit vector $\hat{u}$), preferred speed (s), and spatial scale or receptive field size ($\lambda$). The activity of one such cell in the Local stage can be written as $R(i, f, \hat{u}, s, \lambda)$, where the quantities following the colon are parameters of the cell. Similarly, we write the activity of the corresponding cell in the Coherence stage as $R(i, f, \hat{u}, s)$. This function is discrete, since a finite number of cells represent it, but for mathematical convenience, we think of it as continuous and smooth.

The mathematical formulation presented in this section postulates that the Coherence stage finds the $R$ that minimizes the following energy function:

$$E(i) = \sum_{f, \lambda, \hat{u}, s} (R(i, f, \hat{u}, s, \lambda) - R(i, f, \hat{u}, s, \lambda))^2$$

$$+ \int_{f, \lambda, \hat{u}, s} \psi_s(D, R_s)^2 + \psi_s(D, R_s)^2$$

$$+ \psi_s(D, R_s)^2 + \psi_s(D, R_s)^2$$

$$+ \int_{f, \lambda, \hat{u}, s} W(s) \left( \frac{\partial R}{\partial t} + s \nabla R \cdot \hat{u} \right)^2,$$ (5)

where the $\psi_s$ are positive constants, which set the relative importance of each term, $D_s$ are differential operators to be explained below, and $W(s')$ is a density probability function to be determined experimentally. Although this mathematical formulation is in terms of energy minimization, it could also be recast as finding a maximal-likelihood Bayesian estimator under Gaussian noise (Clark & Yuille, 1990).

By minimizing this energy function, the cells in the Coherence stage of the theory try to be as consistent as possible with (1) the input from the Local stage, (2)
with neighbor “coherence” cells, and (3) with past information about the motion. Minimization of the summation imposes consistency with local inputs, since it reduces the differences between $R_i$ and $R_j$. Minimizing the first integral smoothes the responses across neighborhoods by reducing the magnitude of the differentiation of $R_i$. Although this integral sums over the entire space, the $D_i$ operators are local, and thus impose neighborhood coherence. These operators could normally be partial derivatives, gradients, or more general operators using a combination of derivatives of various orders (Yuille & Grzywacz, 1989a). However, in some instances, they may not be symmetrical, since if at some point in the multi-dimensional space, the local data $(R_i)$ changes too fast along some direction, then data around that point are not used in the derivative estimate. For instance, let us say that normally, $D_i R_i = \partial R_i / \partial \lambda$ and that this is implemented around $\lambda_0$ as the three-point formula $D_i R_i(\lambda_0) = (R_i(\lambda_0 + \Delta \lambda) - R_i(\lambda_0 - \Delta \lambda)) / (2\Delta \lambda)$. Let us also say that $|R_i(\lambda_0 + \Delta \lambda) - R_i(\lambda_0)|$ is larger than a preset criterion. Then, we might estimate $D_i R_i$ as the two-point formula $(R_i(\lambda_0) - R_i(\lambda_0 - \Delta \lambda)) / \Delta \lambda$. Therefore, the $D_i$ operators are not just differential operators, but are functions of the local data, that is, $D_i = D_i(R_i)$. (In practice, such local-data-dependent smoothing could be implemented by a line-process mechanism, Geman & Geman, 1984; Blake & Zisserman, 1987; Poggio, Gamble & Little, 1988, or as in equations 12 and 13 in the Appendix.) Moreover, the spatial operator must also be a function of $\lambda$, that is, $D_i = D_i(\lambda, R_i)$. This is because, as we argued above, the size of the spatial neighborhood of a cell must be proportional to its spatial scale. One way to implement this requirement is to have the interneighbor spatial distances be proportional to $\lambda$, and to estimate derivatives from neighbor responses. Another way is to define $D_i$ as an operator using a combination of derivatives of various orders. Such operators can implement smoothing with Gaussian approximation (Yuille & Grzywacz, 1989a), whose standard deviation can be made proportional to $\lambda$.

The bottom integrals try to impose temporal coherence by using an integrand that is reminiscent of the “image constraint equation” of Fennema and Thompson (1979), and Horn and Schunck (1981). The main difference between their formulation and that in equation (5) is that the new equation uses spatio-temporal derivatives of the responses instead of brightness. Hence, in a sense, this equation measures the movement of the local responses, rather than the movement of the image. The other difference is that equation (5) eliminates speed ($s'$) dependence through integration with the weight function $W(s')$. To obtain the bottom integrand (except for $W$), one postulates that there is a minimal difference between the responses of a cell “A” with particular preferred direction and the past responses of other cells with the same preferred direction and whose receptive fields are located in positions for which the preferred direction point to the receptive field of “A.” Mathematically, this difference can be approximated as

$$\frac{1}{\Delta t} (R_i(t; \vec{u}, s, \lambda) - R_i(t - \Delta t; \vec{u} - s' \Delta \vec{u}, s, \lambda)) \approx \frac{\partial R_i}{\partial t} + s' \nabla R_i \cdot \vec{u},$$

(6)

where the approximation assumes that $\Delta t$ and $s'$ are sufficiently small.

Besides having a new temporal-coherence term, equation (5) has two biologically-relevant, fundamental differences with past theoretical formulations of integration of motion signals (Horn & Schunck, 1981; Hildreth, 1984; Yuille & Grzywacz, 1988, 1989a; Bülow, 1989). The first is that what is being smoothed is not velocity, but actually a measure of velocity likelihood, $R_i$. Although this measure might result in less smooth velocity fields than produced by past formulations, it is more biological in that it is directly encoded in the brain. The second difference with past formulations is that neighborhood integration is not just spatial. We actually include other dimensions (direction of motion, speed, and spatial scale), which might be relevant to biological motion systems. Such a mixing of dimensions has been used in computational models of texture perception (Lee, Mumford & Yuille, 1992).

**COMPUTER SIMULATIONS**

The goal of our simulations was not to produce an exact fit of the data, since many of the experimental details necessary to build a complete implementation of the theory are as yet unavailable. Rather, we wanted to demonstrate that it captures the main qualitative features of the data and that even by using simplistic models of the various stages, the fits were quantitatively reasonable. Accordingly, we used a simplistic implementation of the Local stage, approximated the Coherence stage equation (equation 5) through local means (first temporal and then spatial), and used an approximation for the outlier-detection statistic. For instance, one simplistic approximation was that the directional bandwidth of the Local stage cells was sharp (equation 9) and 45 deg wide. This means that responses only occurred when a dot moved in a direction less than 22.5 deg away from the preferred direction. These responses were all-or-none, that is, had a fixed amplitude rather than being computed from a full Motion-Energy or Elaborated-Reichardt model. Because of this assumption, the local detectors were not affected by the noise, an extreme version of the result in Fig. 1.

This section will proceed directly to the simulations’ results leaving the details of the implementation to the Appendix.

**Computational results**

In the first set of simulations, the signal dot moved in straight trajectories, while the noise underwent a Brownian motion with the same step size as the signal dot. The raw data for one such motion appears in Fig. 3.
In this case, the signal dot moved to the right and the probability of mismatch was 0.3. The figure shows the output signals of the theory ($R_e$ in equation 14 with added internal noise) in three $22 \times 22$ arrays of cells and in three time snapshots. These arrays are representative of the cell population and their preferred directions are indicated by the arrows. The signal is proportional to the diameter of the black disks, whose centers indicate the middle of the cells’ receptive field.

Figure 3 shows that while the response to the noise dots increase little over time, the response to the signal dot (large black disk in the right panel of the middle row) can rise dramatically. After 1 step, no dot produced a signal that dominated the others, since temporal coherence did not have time to work. However, in this particular example, the signal-dot response was already dominant after only 5 steps. And after 9 steps, the ratio between the responses to the signal dot and the strongest noise response was quite large (2.9). This demonstrates that temporal coherence can bring the responses to the signal dot out of the noise even under large probability of mismatch.

What is the prediction of the theory for the dependence of percent-correct detection of the signal dot on probability of mismatch? To answer this question, we simulated the responses to 12-step straight trajectories, with the noise undergoing Brownian motion. Detection of the signal dot occurred when the Dixon-like statistic crossed a predetermined threshold (equation 15). These simulations were performed at various probabilities of mismatch and the result is shown as the solid line in Fig. 4. In addition, the figure shows the corresponding experimental data.

Even with a naive implementation, the theory generates a reasonable fit to the data. That the theory predicts a single curve despite the stimulus being at various step sizes is mainly due to our assumption that the protocols of Watamaniuk et al. (1994) select a local spatial scale close to the step size and that the diameter of the coherence neighborhood is proportional to that scale (see “Size of spatial neighborhood”). In this sense, the theory captures the important qualitative feature that the main independent variable dominating performance is probability of mismatch. Moreover, the theoretical fit has a steepness that is similar to the data. The largest discrepancy occurs at low probabilities of mismatch where the model underestimates performance (see “Discussion”).
We also investigated how the theory's performance improves as a function of the duration of the trajectory. To do so, we simulated the responses of the theory's implementation to the signal dot moving in circular trajectories. These trajectories were always at 1.14 deg eccentricity with the signal dot's direction of motion changing by 12 deg from step to step. In turn, the noise dots underwent Brownian motions with probabilities of mismatch of either 0.224 or 0.304. The percent of correct detection of the signal dot (again based on the Dixon-like detection) as a function of time (each step corresponds to 20 msec) appears in Fig. 5. For comparison, this figure also displays corresponding psychophysical data, whose methods are in the Appendix.

Once again, despite a naive implementation, the theory led to reasonable fits of the data. Qualitatively, both the theoretical and experimental curves rise rapidly for durations lower than about 9 steps (200 msec) and then slowed down, but kept increasing for durations up to 29 steps (600 msec). (The only exception was for subject SM at probability of mismatch of 0.304, for which performance appeared to level off at around 19 hops, 400 msec.) Therefore, at constant eccentricity, the temporal buildup in performance continued for a longer time than when the signal dot moves across eccentricities.
as in the straight-trajectory case. A major reason for that was that straight trajectories crossed large eccentricities where detection is poor (see Appendix). When eccentricity was constant, as in Fig. 5, the theory's prediction for the separation of the two probability-of-mismatch curves was reasonable. The main discrepancies of the theoretical fits occurred at large durations. These fits overestimated performance for subject SW at the high probability of mismatch and underestimated performance for subject SM at the low probability of mismatch. The theory also could not capture the rise-and-fall behavior of subject's SM high-probability-of-mismatch data. We currently have no good explanation for these theoretical discrepancies.

Finally, we tested whether the theory could predict the dependence of performance on the directional bandwidth of the noise dots. In these simulations, the signal dot moved in 12-step straight trajectories. The noise dots moved with the same step size as the signal dot and in such a manner that the directions of each dot was selected randomly from step to step from a homogeneous distribution 180 deg wide. The main independent variable was the difference between the signal dot's direction and the mean direction of the noise dots. This difference varied from 0 to 180 deg in steps of 10 deg. The results appear as a polar plot in Fig. 6, where percent-correct detection is the radius, difference between the direction of motions is the angle, and the curve was reflected about the horizontal axis to estimate performance at angles higher than 180 deg. Again, the corresponding experimental data from Watamaniuk et al. (1994) are plotted for the sake of comparison.

The theoretical curves captured the qualitative features of the data. When the mean direction of the noise was the same as the signal dot's direction, performance was poor. As the difference between these directions increased, the detection performance of both the model and data rose. This indicated that the detection of the signal dot depends on motion mechanisms and is not just due to positional cues. In addition, these results indicated that a destructive interaction between the noise and signal dots occurs in limited directional neighborhoods (see equation 5). The quantitative fits to the detection of the signal dot were reasonable, especially at directional differences smaller than 90 deg and at differences around 180 deg. However, the fits overestimated the performance at directional differences around 135 deg (see "Discussion").

**DISCUSSION**

We have proposed a new theory to account for the temporal integration of local-motion signals. This theory postulates that local-motion signals are made coherent over time, by taking into account past information from the trajectory of the motion. Coherence is also imposed across cells of similar spatial positions, preferred directions, optimal speeds, and spatial scales (or spatial frequencies). Due to temporal coherence, the responses to image features moving in a trajectory with no sharp turns are facilitated. Hence, these signal features can be detected with an outlier-selection procedure. Simulations with a simple neural-network implementation of the theory show that it can account qualitatively for all reported experimental features related to the detection of a signal dot amidst a background of Brownian motion. In addition, the fits were for the most part quantitatively reasonable.

That the fits produced by the theory were not always quantitatively perfect can be largely explained by our oversimplified implementation of the theory. For instance, one simplistic approximation was that the directional bandwidth of the Local stage cells was sharp (equation 9) and 45 deg wide. Consequently, in the experiment with limited directional-bandwidth noise (Fig. 6), the simulations predicted, contrary to experimental evidence, that when the difference between the signal's direction and the mean direction of the noise dots was larger than 135 deg, the performance would be perfect. This is because there would be no noise dots stimulating the detectors responding to the signal dot. If we were to relax the assumption of sharp directional bandwidth, the predicted performance at these directional differences would be affected by noise dots, making the fits conform better to the data. Another example for which the fits did not conform perfectly to the data was the probability-of-mismatch curves (Fig. 4). The steepness of the data curve was higher than that of the simulations. A reason for this discrepancy might have been the use of a fixed response amplitude for the local detectors (equation 9), rather than using a full Motion-Energy or Elaborated-Reichardt model.

![Figure 6](image-url) Directional bandwidth of the signal–noise interaction. In this experiment, the signal dot moved in 12-step straight trajectories. The motion of the noise dots had the same hop size as the signal dot, but with the directions of each dot selected randomly from frame to frame from a homogeneous distribution 180 deg wide. The independent variable was the difference between the mean direction of the noise and the direction of the signal. In the figure, this difference is the angle of the polar plot, while its radius is percent-correct detection. The simulations (solid and dashed lines) were performed for angles between 0 and 180 deg and the results were reflected about the horizontal axis. The experimental data (closed symbols) are the same as in Watamaniuk et al. (1994). Both the experimental and theoretical performances rise as the difference between the direction of signal and noise motions increase. The fits are reasonable for differences smaller than 90 deg and around 180 deg. The only significant discrepancies in the fit happen for intermediate differences, that is, around 135 deg (see "Discussion").
Because of this assumption, the local detectors were not affected by the noise. If they were, its effect would have been larger for the larger probabilities of mismatch, thus increasing the steepness of the simulation's curve.

Is motion coherence biologically plausible?

Our simple implementation of neighborhood and temporal smoothing at the Coherence stage of the theory (equation 5) involves the computation of local weighted means of present and past responses. Are computations like this consistent with present knowledge about cortical anatomy and physiology? We argue that they are and use Fig. 7 to help explain why. This figure schematically illustrates an array of $5 \times 5$ cells in the Coherence stage (with receptive-field positions concentric with the circles) and the network connectivity of one of them. For the purpose of simplicity, the inputs from the Local stage do not appear, and these cells have the same preferred direction (shown by the arrows), spatial scale, and preferred speed.

As pointed out in “Coherence stage of the theory,” temporal-coherence computations are probably performed by transmitting responses from place to place through delay lines. In the figure, this is illustrated by the “shower” of delayed (indicated by the $\Delta t$ box) excitatory inputs (closed dots) of cells whose preferred direction point to the target cell. Such delay lines could be mediated by the long-range cortical connections for which evidence has been recently mounting (Gilbert & Wiesel, 1979, 1988; Martin & Whitteridge, 1984; Kitano, Niiyama, Kasamatsu, Sutter & Norcia, 1994). These delays do not have to be proportional to the retinal distance between the connected cells, since speed information is probably discarded (see “Coherence stage of the theory”). In turn, neighborhood coherence involves both excitatory (closed dots) and inhibitory (open dots) connections (again, see “Coherence stage of the theory”). In the figure these connections are only with the immediate near neighbors, but it remains to be established how far they can go in the context of motion processing. The excitatory and inhibitory synapses implementing the connections should operate differently so that their effects do not cancel each other. This is consistent with the intracortical excitatory pathways being more massive than the inhibitory ones (Douglas, Martin & Whitteridge, 1988, 1991; Douglas & Martin, 1992) and with inhibition, but not excitation, being division-like to implement response normalization (Heeger, 1992, 1993). If one thinks of the excitatory synapse as being linear, then the response of a cell before inhibition is taken into account is the weighted sum of its own response and those of its neighbors. If the effect of inhibition is to normalize the responses, then instead of the response being a weighted sum, it becomes a weighted mean. In this regard, the response can approximate well the simple smoothing implementation we chose for our theory. The exact form of smoothing would, of course, depend on the details of inhibition and excitation. But in any event, it could be consistent with inhibition of signal responses by noise dots and temporal facilitation of trajectory motions.

Role of temporal coherence

Our naive implementation of the theory assumed a single spatial scale for the local-motion detectors despite potential complications due to aliasing at other scales (Fig. 1). We now argue that this is not a serious problem, since the theory deals with aliasing ambiguities well. To illustrate this point, Fig. 8 illustrates how the theory would solve a version of this problem. In this version, the motion signals are completely ambiguous locally, but this ambiguity propagates in time to the right. In this case, the responses of “coherence” cells with rightward, but not other, preferred directions get progressively facilitated (increasingly thicker arrows), allowing the detection of the motion of the ambiguity. Preliminary simulations confirmed that this is exactly what happens in the theory. This not only demonstrates that the theory would “disregard” the aliasing reversals, but also that temporal coherence could deal with some Non-Fourier motions, which essentially are movements of ambiguities of the type illustrated in the figure (Lekens & Koenderink, 1984; Chubb & Sperling, 1988; Turano & Pantle, 1989; Victor & Conte, 1990). In addition, we speculate that the theory could explain some types of
Second-Order motion, which are motions of subjectively defined boundaries (Cavanagh, Arguin & von Grunau, 1989; Cavanagh & Mather, 1989; Zanker, 1990, 1993). Temporal coherence would help in some of these cases, since to define boundaries, image features appear suddenly at the boundaries, giving rise to transient, ambiguous motion signals there. In other words, such a subjective boundary would stimulate local-motion detectors of all preferred directions, including those pointing in the direction of the movement of the boundary, which thus could facilitate them through temporal coherence. In this sense, the theory works differently from the correlator-of-correlators model of Zanker (1993), which can only account for Second-Order motions of boundaries defined by First-Order motion stimuli. It is of significance that Fig. 8 suggests that in principle, early rectification (Chubb & Sperling, 1988; Turano & Pantle, 1989; Werkhoven et al., 1993) or more than one motion pathway (Wilson, Ferrera & Yo, 1992; Wilson & Mast, 1993; Zhou & Baker, 1993) are not necessary to deal with many types of Non-Fourier and Second-Order motions. Moreover, since according to the Temporal-Coherence theory, these types of motions and luminance-based motions could share the Coherence stage, they would under many circumstances result in similar psychophysical performances (Turano & Pantle, 1989; Smith, Hess & Baker, 1994). The buildup of signals in “Coherence” cells would favor motion “correspondences” along trajectories over “correspondences” with near neighbors, thus being consistent with motion-inertia, and the temporal buildup of the upper displacement limit (Nakayama & Silverman, 1984; Snowden & Braddick, 1989) and deblurring of images (Burr, Ross & Morrone, 1986; Watamaniuk, 1992) in visual translations.

Is there a correspondence problem?

Despite Adelson and Bergen’s (1985) arguments against a motion correspondence problem, the visual system must disambiguate local-motion signals such as those produced by aliasing, and that implies that the system suffers from a form of this problem. However, in agreement with them, the problem is probably not due to large separation of image features from frame to frame and is probably not solved by matching near neighbors (Ullman, 1979). The “nearest-neighbor-like” performance in the Watamaniuk et al. (1994) experiments, that is, their dependence on probability of mismatch can be accounted for by strong inhibitory effects from spurious noise dots falling in the coherence neighborhood of the signal motion (Fig. 7). In this case, all one needs to explain the experimental results is that size of the coherence neighborhood and the dependence on eccentricity of the “visibility” of local motions at fixed spatial scales (see “Parameters”) are proportional to the size of the relevant local-motion detectors. Our experimental and theoretical results provide evidence for these interesting scale-similarity processes in the human visual system.

REFERENCES

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two dimensional


Acknowledgements—We would like to thank Drs Mary Bravo, Julie Harris, Steve Heinen, Misha Mahowald, and Alan Yuille for critically reading the manuscript. Support for this work came from a grant from the Air Force Office of Sponsored Research (F49620-92-J0156) and from a core grant from the National Eye Institute to Smith-Kettlewell (EY06883).
APPENDIX
Experimental Methods

The stimuli were dynamic random dot cinematosgrams in which each dot took an independent two-dimensional random-walk with a constant step size. Each noise dot’s displacement was chosen randomly from a uniform distribution spanning 360 deg each frame and was independent of both its previous displacements and the displacements of other dots. A signal dot, moving on a circular path centered on a stationary fixation point, was presented amidst this random-direction motion noise. The signal dot moved clockwise or counter-clockwise, chosen randomly each trial. Each trial, the signal dot started at a randomly determined position on its circular path. Within a single trial, noise and signal dots took the same constant spatial displacement (hop size) each frame. Thus the signal was indistinguishable from the noise on any two-frames. The only characteristic that could be used to separate the signal from the noise was the signal’s consistent pattern of movement over time.

Stimuli were displayed, under computer control via A/D converters, on an x-y cathode ray tube display (CRT) equipped with a P4 phosphor. Observers viewed the CRT through a 10 deg diameter circular aperture, from a distance of 57 cm and fixated a spot located at the center of the screen. The height of the CRT was set so that the center of the aperture was approximately at eye level. Stimuli were presented at a frame rate of 50 Hz and each stimulus dot subtended 0.07 deg. All experiments took place with the overhead room lights on creating a background luminance of 43 cd/m². Space-averaged dot luminance was 61 cd/m². This value was obtained by plotting a matrix of non-overlapping dots (center-to-center spacing was 0.08 deg) at the same frame rate as used in the experiments and then measuring the luminance of this matrix with a Minolta luminance meter. Push buttons connected to the computer initiated each trial and signaled observer responses.

Observers judged the presence of the signal (circular-trajectory) in the stimuli in a two-alternative forced-choice paradigm. Observers were shown two stimulus intervals of which only one, randomly selected each trial, contained the signal. Observers judged in which interval the signal was presented. Feedback was provided. Each experimental run consisted of 80 trials, preceded by 15 practice trials. Percent correct detection was evaluated for each experimental run. Reported performance values were evaluated for each condition by averaging values from many separate runs. Error bars on all graphs are ±1 standard errors reflecting the between-run variability for each condition and subject. Two of the authors (SW and SM), both experienced psychophysical observers, provided data for all experiments. Both observers had corrected to normal vision.

In the eccentricity-dependence experiment, we measured detectability of several circular paths covering a range of eccentricities (0.5-3.0 deg) at a constant probability of mismatch of 0.18. Performance was measured for three hop sizes (0.08, 0.16, and 0.32 deg) while duration was held constant at 400 msec. Within a single experimental run, hop size was fixed but the signal could appear at any of the specified eccentricities. At least 100 trials were run for each stimulus condition.

In the temporal-dependence experiment, detectability for a circular path with an eccentricity of 1.14 deg was measured at two probability of mismatch values, 0.224 and 0.304. The probability of mismatch value was changed by varying the density of noise dots while keeping the hop size constant at 0.24 deg. Performance was measured for a range of durations from 100-500 msec. Data for a single duration and probability of mismatch were collected within a single block of 80 trials. Three 80-trial blocks were run for each stimulus condition giving a total of 240 trials per data point.

Eccentricity Results

We studied the effects of eccentricity on the detection of the signal dot by using circular trajectories of various radii and fixating at the center of the circles [Fig. A1(A)]. All measurements used the same probability of mismatch (0.18), since this is the most important independent variable controlling performance. However, the step sizes were varied. The results of these experiments appear in Fig. A1(B).

The data show that the detection of the signal dot falls with eccentricity even at eccentricities as near as 1 deg. Furthermore, the fall is faster at shorter hop sizes. To quantify this effect, we fit Gaussian profiles to the curves. (We chose Gaussians, since visually, some of the curves appeared to have small slopes at short eccentricities.) Before producing these fits, percent-correct values ($P_e$) were transformed into percent-detection values ($P_d$) by assuming that half of those trials in which the signal dot was not seen yielded correct responses by guessing. This transformation is

$$P_d = 2(P_e - 0.5).$$

(7)

The Gaussians were fit to the $P_d$ and then transformed back to $P_e$ values by the inverse of equation (7). The free parameters in the fit (performed with the “fmin” algorithm of Mathlab-MathWorks, Inc., South Natick, MA, U.S.A.) were the amplitude (constrained to be less than or equal to 100%) and standard deviation ($\sigma$) of the Gaussians. Figure 9(b) shows the fits as the solid, dashed, and dashed-dotted lines. For subject SW, the fits gave $\sigma = 0.78$, 1.61, and 2.42 deg for the curves corresponding to the step sizes 0.08, 0.16, and 0.32 deg respectively. For subject SM, the same curves yielded $\sigma = 0.82$, 1.64, and 2.03 deg, respectively. Measured with a z² test, the fits were statistically indistinguishable from the data, with the poorest fit yielding $P > 0.30$ (for subject SW at step size of 0.08 deg) and the best fit yielding $P > 0.99$ (for subject SM at step size of 0.32 deg). That $\sigma$ increases with step size is a quantitative statement that decline in detection with eccentricity is faster at shorter step sizes.

Computational Methods

The goal of our simulations was not to produce an exact fit of the data, since many of the experimental details necessary to build a complete implementation of the theory are as yet unavailable. Rather, we wanted to demonstrate that it captures the main qualitative features of the data and that even by using simplistic models of the various stages, the fits were quantitatively reasonable.

Implementing the Local stage

Accordingly, our implementation of the Local stage was as simple as possible. This stage comprised eight $M \times M$ square-lattice arrays of directionally selective cells. What differentiated the arrays was the cells’ preferred directions, $\theta_i$, where $1 \leq i \leq 8$. The $\theta_i$ differed by 45 deg between neighbor arrays and one of the arrays had a preferred direction to the right. The choice of 45 deg throughout the paper was based on the “wobble” experiment of Watamaniuk et al. (1994). They showed that abrupt changes of direction of motion of more than 41 and 57 deg for subjects SM and SW, respectively caused the detection of the signal dot to be significantly impaired. A compromise between these two values that divided nicely 360 deg was the directional bandwidth of 45 deg. Reducing this angle to 30 deg and accordingly, increasing the number of preferred-direction arrays to twelve (Williams, Tweten & Sekuler, 1991) would cause a small increase in the sensitivity to changes in direction of motion, and thus, a slightly worse performance in circular trajectories. The distance between the middles of the receptive fields of two neighbor cells in each array was equal to the dots’ step size. The value of $M$ was determined from each simulation’s probability of mismatch such that 50 noise dots always participated in each simulation. (This value of 50 was an arbitrary choice for the number of dots involved in the Outer stage of the model.) From equation (1), before rounding to nearest integer, $M$ was

$$M = 1 + \left(\frac{50\pi}{\tan(1 - \theta_i)}\right)^{1/2}.$$

(8)

After determining $M$, we moved the dots, which we will label by $j$ (1 $\leq j \leq 51$, including the signal dot) from position $(\theta_i - \Delta t \theta_i)$ to position $(\theta_i + \Delta t \theta_i)$, where $\Delta t$ is the time interval between frames. We defined the local velocity as $\theta_i(t) - (\theta_i(t) - \theta_i(t - \Delta t)) / \Delta t$ and the middle position as $\theta_i(t) = (\theta_i(t) + \theta_i(t - \Delta t)) / 2$. (This definition of velocity neglects the false-matching problem, an assumption justified by the relative independence of directional selectivity in motion-energy units on the Watamaniuk et al. noise.) Fig. 1) Then, our simple implementation of the Local stage modeled the sum of the responses of a set of
cells with parameters $\langle \hat{t}, \hat{\delta} \rangle$ (recall that we are considering $s$ and $\lambda$ constant for the Watamaniuk et al. experiments) as

$$R_d(t; \hat{t}, \hat{\delta}) = \begin{cases} R & \text{if } \exists j, \cos^{-1}(\hat{t} \cdot \hat{\delta}_j / |\hat{\delta}_j|) < 22.5 \text{ and } |\hat{t} - \hat{t}_j| < h \\ 0 & \text{otherwise} \end{cases}, \quad (9)$$

where $R$ is a positive constant, which depends on eccentricity. What "sum of the responses of a set of cells" means is that $R$ should not be interpreted as the response of a single cell, but as the "confidence" that a set of local-motion detectors have in a particular motion. This confidence falls with eccentricity, since the number of local-motion detectors with a particular scale decreases. Equation (9) assumed that the receptive field and directional bandwidth of local-motion detectors were sharp, in the sense that the response was all or none. The response only occurred if there was a dot that fell within a distance $h$ from the middle of the receptive field and moved in a direction less than 22.5 degrees away from the preferred direction. Furthermore, the response was always the same, that is, saturated, regardless of the number of stimulating dots. We also justify this constant-response assumption because of the relative independence of directional selectivity in motion-energy units on the Watamaniuk et al. noise.

**Implementing the coherence stage**

The implementation of the Coherence stage was also a simplification of equation (5). The first simplification was that we performed temporal coherence before neighborhood coherence. We labeled the intermediate Coherence stage responses, that is, the responses after temporal coherence but before neighborhood coherence, by $R^\ast(t; \hat{t}, \hat{\delta})$. At $t = 0$, we assumed that $R^\ast(t; \hat{t}, \hat{\delta}) = R(0; \hat{t}, \hat{\delta}) = 0$. At later times, we used the immediate past information about the motion ($R_d(t - \Delta t)$) and the present information about local motion ($R(t)$) to compute $R^\ast(t)$ as

$$R^\ast(t) = \begin{cases} (1 - \psi^\ast) R_d(t) + \psi^\ast R_d(t - \Delta t; \hat{\delta} - h\hat{\delta}) & \text{if } (0 \cdot \hat{\delta}_0)(\hat{t}; \hat{\delta}_0) = 0 \\ (1 - \psi^\ast) R_d(t) + \psi^\ast R^\ast(t - \Delta t; \hat{t}, \hat{\delta}) & \text{otherwise} \end{cases}, \quad (10)$$

where

$$R^\ast(t - \Delta t; \hat{t}, \hat{\delta}) = \frac{1}{3} \left( R_d(t - \Delta t; \hat{t} - h\hat{\delta}) + R_d(t - \Delta t; \hat{t} - h\hat{\delta} + h\hat{\delta}) + R_d(t - \Delta t; \hat{t} - h\hat{\delta} - h\hat{\delta}) \right), \quad (11)$$
and where \( \psi^* \) is a positive parameter and \( \bar{\mu} = \text{sign}(\bar{\mu})(1,0), \quad \bar{\nu} = \text{sign}(\bar{\nu})(0,1) \), with \( \text{sign}(x) = 1 \) if \( x > 0 \), \( \text{sign}(x) = -1 \) if \( x < 0 \), and \( \text{sign}(x) = 0 \) if \( x = 0 \), and with \((1,0)\) and \((0,1)\) being the horizontal and vertical unit vectors, respectively. The parameter \( \psi^* \) is tightly related to the parameter \( \psi \), in equation (5), since they weigh the relative importance of the local data and temporal coherence. The only difference between these parameters is that \( 0 < \psi^* < 1 \), which ensures that the discrepancy between present and past signals decrease as required by the last term of equation (5). Past information was taken from the neighborhood of the current position minus \( \bar{\nu} \), that is minus one hop in the direction opposite to the preferred direction (see equation 6). The diameter of the neighborhood was just smaller than the hop size \( h \). This diameter selected one cell for horizontal and vertical preferred directions (top term of right hand side of equation 10) and three cells for diagonal directions (equation 11). For cells at the border of the network, the selected “past” cells would sometimes have to be taken from outside of it. In this case, they were picked from the other side of the network in a wrap-around scheme.

After the temporal portion of the implementation of the Coherence stage was completed, we proceeded with neighborhood coherence.

In neighborhood coherence, only neighbors that receive similar inputs from the Local stage are made coherent (to allow for transparency and the detection of boundaries). Therefore, it is useful to define the neighborhood similarity function as \( \text{sim}(\bar{f}, \bar{k}) = 1 \) if \( \underline{R}_i(\bar{f}, \bar{k}) = R_i(\bar{f}, \bar{k}) = R_i(\bar{f} + h \cos(\theta x), \sin(\theta y)) \) and \( \text{sim}(\bar{f}, \bar{k}) = 0 \) otherwise, for \( k = 1, 2, 3, \text{or } 4 \). This function tells when two immediate neighbors in the square lattice have a similar local response. With this definition, our simple implementation of neighborhood coherence yields the output of the Coherence stage as

\[
\underline{R}_i(\bar{f}, \bar{k}) = \psi^* \left( R^{**} + (1 - \psi^*) R^* \right),
\]

where

\[
R^{**} = \sum_{k \in B} \text{sim}(\bar{f}, \bar{k}) R^*_i(\bar{f} - h \cos(\theta x), \sin(\theta y))
\]

\[
= \sum_{k \in B} \text{sim}(\bar{f}, \bar{k}) R^*_i(\bar{f} - h \cos(\theta x), \sin(\theta y))
\]

and

\[
\text{if } \sum_{k \in B} \text{sim}(\bar{f}, \bar{k}) > 0,
\]

\[
0 \text{ otherwise}
\]

(13)

with \( \psi^* \) being a positive parameter. These equations impose that only neighbors that receive similar inputs from the local stage are made coherent by only setting \( \text{sim} = 1 \) for them, and thus by pooling their responses together and not pooling these responses with those of other cells. Similar to \( \psi^* \), \( \psi^* \) is tightly related to the parameter \( \psi \), in equation (5), since they weigh the relative importance of the local data and neighborhood coherence. Again, the only difference between these parameters is that \( 0 < \psi^* < 1 \), which ensures that the discrepancies between a cell and its neighbors decrease as required by the first integral of equation (5). In the simulations, the directional neighborhood was the same as the directional bandwidth of a local-motion detector, that is, 45 deg. The spatial neighborhood used information from the immediately adjacent cells (jumps of \( h \) in equation 13). Therefore, by taking into account the receptive-field radius of local-motion detectors, the diameter of the coherence neighborhood was twice larger (2\( h \)) than that of the local detectors. However, the neighborhood of a cell was not always symmetric, since only neighbors similar to it would be included in the coherence computation (as imposed by function \( \text{sim} \) in equation 13). If no neighbors were similar, then responses were reduced by a factor of \( 1 - \psi^* \). This small reduction could happen in the absence of noise, or when the noise moved opposite or perpendicularly to the signal dot. For cells at the border of the network, the neighbors would sometimes have to be taken from outside of it. In this case, as we did for temporal coherence, they were picked from the other side of the network in a wrap-around scheme.

Implementing the Outlier stage

Neighborhood coherence was followed in the implementation by the Outlier stage. The first step in this stage was to eliminate the smear of responses in the Coherence stage by gating them with direct inputs from the Local stage (see end of “Outlier stage of the theory”). To do so, we computed the responses \( \underline{R}_i(\bar{f}, \bar{k}) \) of outlier candidates as

\[
\begin{align*}
\underline{R}_i(\bar{f}, \bar{k}) & > 0 \quad \text{and} \\
\underline{R}_i(\bar{f}, \bar{k}) & > R_i(\bar{f} + h \times (1,0), \bar{k}) \quad \text{and} \\
\underline{R}_i(\bar{f}, \bar{k}) & > R_i(\bar{f} + h \times (-1,1), \bar{k}) \quad \text{and} \\
\underline{R}_i(\bar{f}, \bar{k}) & > R_i(\bar{f} + h \times (0,1), \bar{k}) \quad \text{and} \\
\underline{R}_i(\bar{f}, \bar{k}) & > R_i(\bar{f} + h \times (-1,-1), \bar{k}) \quad \text{and} \\
\underline{R}_i(\bar{f}, \bar{k}) & > R_i(\bar{f} + h \times (0,-1), \bar{k}) \quad \text{and} \\
\underline{R}_i(\bar{f}, \bar{k}) & > R_i(\bar{f} + h \times (-1,1), \bar{k}) \quad \text{and} \\
\underline{R}_i(\bar{f}, \bar{k}) & > R_i(\bar{f} + h \times (1,-1), \bar{k}) \quad \text{and} \\
0 & \text{otherwise}
\end{align*}
\]

(14)

This equation says that a cell is a candidate to be an outlier if this cell has the largest signals among its neighbors and its counterpart cell in the Local stage is firing. Consider spatial neighborhoods including 50 noise dots around each encoded preferred direction. (In the way we implemented the model, the spatial dimension of the neighborhoods are the entire network, equation 8, and the directional neighborhoods are 45 deg wide, equation 9.) In these neighborhoods, let the largest and second largest outlier-candidate responses be \( \underline{R}_i(\bar{f}, \bar{k}) \) and \( \underline{R}_i(\bar{f}, \bar{k}) \), respectively. We implemented detection if at any time during the motion

\[
D(\bar{f}, \bar{k}) = \frac{\underline{R}_i(\bar{f}, \bar{k})}{\underline{R}_i(\bar{f}, \bar{k})} > 0,
\]

(15)

where \( \theta \) is a positive parameter corresponding to detection threshold.

The value of \( D(\bar{f}, \bar{k}) \) in this equation approximates the Dixon statistic, which is often used for the removal of outliers (Dunn & Clark, 1987). Strictly speaking, the Dixon statistic is the ratio between the closest and the farthest distances of an outlier candidate to the extremities of the sample with which the candidate is being compared. Because in the present implementation, the lowest responses are close to zero, we approximated the largest distance by \( \underline{R}_i(\bar{f}, \bar{k}) \), obtaining equation (15). This approximation is not the only one made in the context of Dixon's analysis. Strictly speaking, this statistic is only valid if the sample comes from a normal distribution, which is not the case here. In any event, we felt that this statistic could provide a good direct measure of the difference between the signal-dot and noise-dot responses. This statistic approaches zero if this difference is small and approaches unity if this difference is large. In addition, in all our simulations, the statistic never detected a noise dot as signal.

\textbf{Internal noise}

Now that we have described the implementation of all the stages of the theory, it is important to address the issue of internal noise of the system. Even under optimal conditions, that is, noise-free stimuli, humans cannot improve their measurement of visual motion past a given level. For instance, for well trained subjects, speed discrimination seems always to level off at Weber fractions of around 5% regardless of the stimulus (McKeen & Welch, 1985; McKeen, Silverman & Nakayama, 1986; De Bruyn & Orban, 1988; Turano & Pantle, 1998). This limitation cannot be attributed to stimulus noise. Moreover, the Local stage probably does not limit performance, since its noise is essentially smoothed out by the Coherence stage. One must thus postulate that an internal noise after or at the output of the Coherence stage must limit perform under noise-free stimuli. Hence, we summed an additive Gaussian noise to the input to the Outlier stage, that is, to \( R_i \) in equation (14). This noise had zero mean and a standard deviation, which we labeled \( \sigma_e \).

\textbf{Parameters}

To implement the theory, we had to specify the values of \( R \) (equation 9), \( \psi^* \) (equation 10), \( \psi^* \) (equation 12), \( \theta \) (equation 15), and \( \sigma_e \).

The only parameter of the simulation that was set differently for the two subjects in the Watamaniuk et al. experiments was \( \theta \). To set it,
we imposed that percent-correct detection for the 12-step straight trajectory under 0.3 probability of mismatch would be in the middle of the range of the experimentally observed values. (This motion was one of the most representative motions in the experimental set, since its duration and probability of mismatch were roughly in the middle of the range of the stimulus parameters.) From this motion, the values of \( \theta \) were 0.44 and 0.445 for subjects SW and SM, respectively. The choice for the value of \( \sigma \) also used the same motion as a standard. To set the noise such that 90% of its distribution fell within 10% of the signal generated by this motion, we used \( \sigma = 0.011 \).

The choice for the parameter \( R \) involved eccentricity considerations. Because we are assuming that the Watamaniuk et al. experiments essentially selects a spatial scale close to their step size (see "Local scale of the theory"), the simplest interpretation for the fall in performance with eccentricity (Fig. A1) is that this decline mirrors a fall in the density of local-motion detectors with a given spatial scale. Moreover, this detector-density falloff is faster for the small-scale detectors. Therefore, ideally, to model the effects of eccentricity one would have to set the parameter (namely, \( R \)) that is monotonically related to the density of local-motion detectors to different values at different positions of the network (see the interpretation of \( R \) after equation 9). But in the spirit of our search for a simple implementation of the theory, we chose a different route to model \( R \).

We reasoned that in straight trajectories, the signal dot would only spend a fraction of its motion in zones of low eccentricity and thus would be "visible" only in this fraction. We estimated upper and lower bounds for this fraction. The estimate for the upper bound used the \( \sigma \) of the Gaussian fits for the eccentricity data (see "Experimental results"). If "visibility" falls in a Gaussian manner with eccentricity, then assuming a visibility of 1 at zero eccentricity, the average length of visible trajectory in the Watamaniuk et al. straight-trajectory experiment is

\[
L_v = \frac{1}{8\sigma^2} \int_{-\sigma^2}^{\sigma^2} \int_{-\sigma^2}^{\sigma^2} \left( x^2 + y^2 + z^2 + 2z(x \cos (\pi/4) + y \sin (\pi/4)) \right) \exp \left( -\frac{x^2 + y^2 + z^2}{2\sigma^2} \right) dx \, dy \, dz
\]

where \( w \) is the side (2 deg) of the square window (centered on the fixation point) where the middle of the straight trajectories randomly fell in the Watamaniuk et al. experiments, \( L \) is the length of the trajectory, and "8" corresponds to the number of directions in which the signal dot could move. Application of this equation with the \( \sigma \) obtained in "Experimental results" yielded upper bounds for \( L_v \) of 10.1 steps for subject SW and 9.7 steps for subject SM for the 12-step trajectory data of Watamaniuk et al. These estimates of \( L_v \) should be considered upper bounds, since the long duration of the eccentricity experiments (19 steps or 400 msec) tended to raise the eccentricity curves and thus yield overly large estimates for \( \sigma \). To estimate a lower bound, we used the Watamaniuk et al.'s curves describing percent detection as a function of time for straight trajectories. As described in "Experiments," these curves rose during the first 300 msec or so and then leveled off at less than 100% correct. The simplest explanation for this leveling off was that eccentricity caused this impairment. This can be seen directly from equation (16), since it predicts that \( L_v \) first increases and then levels off with \( L \). Consequently, one can use this equation to estimate new values for \( \sigma \), that is, to find what \( \sigma \) best predict the leveling off of \( L_v \) in the data. These estimates would be lower bounds, since as the curves get closer to 100%, they would tend to level off more quickly thus giving the impression that \( L_v \) is leveling off sooner than predicted. For subject SW, the estimates obtained from the data were \( \sigma = 0.72, 0.91 \), and 1.62 deg for the curves corresponding to the step sizes 0.2, 0.24, and 0.30 deg, respectively. For subject SM, the estimates were \( \sigma = 0.72 \) and 1.25 deg for step sizes 0.2 and 0.24 deg, respectively. These values together with equation (16) yielded lower bounds for \( L_v \) of 9.0 steps for subject SW and 8.4 steps for subject SM for the 12-step trajectory data of Watamaniuk et al.

The similarity of the lower- and upper-bound estimates of the fraction of "visible" motion based on equation (16) had important implications both for modeling and the interpretation of the probability-of-mismatch results of Watamaniuk et al. (1994). Because of this similarity, instead of modeling the effect of eccentricity on straight trajectories as local modulations of visibility, we could model this effect as a reduction of the number of visible steps with perfect visibility. Hence, on 12-step trajectories, we set \( R = 1 \) and made the number of visible steps equal to 9. But perhaps more importantly, that we could set this number to the same value regardless of step size implies that the eccentricity effect on visibility scales with this size so that straight trajectories with short and long step sizes cover approximately the same window of visibility in number of steps, or equivalently for the Watamaniuk et al. experiment, in units of the spatial scale of the relevant local-motion detector. This approximate eccentricity scaling is crucial to account for the probability-of-mismatch results, since it implies a dependence of performance on step size.

However, for circular trajectories, setting the parameter \( R \) to unity and reducing the number of "visible" steps is not a good solution. This is because all portions of these motions are equally visible. Instead, this equal visibility allows us to assume that it is density of local-motion units, and thus, \( R \) which falls with eccentricity, rather than the number of steps that is reduced. To estimate this fall, we assumed as always that \( R \) depends on eccentricity in a Gaussian manner. Because the circular-trajectory data (Fig. 5) were taken with a step size of 0.24 deg, we used \( \sigma = 1.08 \) (the mean value for both subjects) for this Gaussian. Given that the experiments were performed with a radius of 1.14 deg, we set \( R = 0.57 \).

Finally, we reasoned that the parameters \( \psi^* \) and \( \psi \) should strongly favor coherence over locality. When observing the signal dot in the Watamaniuk et al. experiments, it would sometimes disappear during the course of the trajectory. Because the stage limiting performance is the Coherence stage (see "Size of spatial neighborhood"), it must be responsible for the disappearance of the signal dot. If a noise dot enters the neighborhood area of the signal dot, then the noise dot's weak signal could pull down the responses to the signal dot by the neighborhood weighted average in equation (12). For this to occur, the value of \( \psi^* \) would have to be high. Moreover, for spurious noise dots to be able to pull down significantly the response to the signal dot, its temporally integrated response must be much higher than the responses to the noise dots. Again, for this to happen, \( \psi^* \) should be high so that the past history of the motion counts significantly (equation 10).

If both \( \psi^* \) and \( \psi \) are high, then one can think of detection in the following terms: if the signal dot can move for a number of steps without the noise dots interfering with it, then it will be detected. However, if a noise dot interferes, then the signal-dot responses are essentially reset and the signal dot has to begin building its responses essentially from scratch. In support of these ideas, in a preliminary modeling effort before the work described here, we found that one could model the Watamaniuk et al. data well with the probability that the signal dot can move without any interference for about four steps. This preliminary study made it even more apparent that \( \psi^* \) and \( \psi \) had to be high, and they were set to 0.95 in all but a pair of simulations. In this pair, \( \psi^* \) and \( \psi \) were set to 0.5, and the dependence on probability of mismatch became too shallow to be consistent with the data.

The choices described in this section for the non-free parameters \( R, \psi^*, \psi \), and \( \sigma \) were not circular in the sense of being estimated from the same psychophysical experiments that were used to test the theory. On one hand, these tests involved the data in Figs. 4, 5, and 6. On the other hand, to choose \( R \), we did not use any of these data, but rather the eccentricity data in Fig. A1. Furthermore, to choose \( \psi^* \) and \( \psi \), we only used the general consideration of the occasional disappearance of the signal dot during the course of the trajectory. This is a feature of all our experiments and not particular to any data set. Finally, the choice of \( \sigma \) was based on experiments performed by other groups, which measured a ceiling in the performance of the human motion system (see "Internal noise").

When we modulated \( R \) and \( \sigma \) up and down individually by factors of two (while holding \( \theta \) constant), the simulation results were qualitatively, but not quantitatively, similar to those reported in this paper. The same conclusion was reached when \( \psi^* \) and \( \psi \) were individually reduced by a factor of two. Either raising \( R \) or lowering \( \psi^* \) tended to diminish the relative effect of temporal coherence.
resulting in poorer performances. In contrast, lowering $\psi^*$, tended to reduce the negative effect of the noise dots on the signal dot, improving performance. Addition of noise by increasing $\sigma_n$ made outlier detection, and thus detection of the signal dot, harder.

**Stimuli**

To complete the implementation's description, we now address the motions of the signal and noise dots. The signal dot moved either in a straight or circular trajectory. In both cases, the hop size ($h$) was constant. The straight trajectories reported here were always 12-steps long. However, these motions were in random directions (100 trials each), with the middle of the trajectory being randomly placed within a $h \times h$ square centered on the middle of the network. In the circular trajectory, the direction of motion changed by 12 deg from step to step (as was the case for the experiment in Fig. 5). This trajectory was always counterclockwise, but the direction of the initial tangent to the circular trajectory was randomized (100–300 trials) and the position of the center of the circle was placed randomly within a $h \times h$ square centered on the middle of the network. (Different than in the psychophysics where the circle was always centered on the fixation point, in the theoretical implementation, we randomized the circle's position to avoid systematic effects due to the square-lattice arrangement of the cells.) The noise dots either underwent a Brownian motion with hop size $h$, or a motion that also had hop size $h$, but with the directions of each dot selected randomly from frame to frame from a homogeneous distribution 180 deg wide. The Brownian motion was used with both the straight and circular trajectories, with the main independent variables being probability of mismatch and stimulus duration. The motion with 180 deg directional bandwidth was only used with straight trajectories. Here the independent variable was the difference between the mean direction of the noise and the direction of the signal. Simulations were made with this difference set at every 10 deg from 0 to 180 deg. For each difference, there were 100 trials, each with a random signal direction.