

# NEURAL DYNAMICS OF MOTION INTEGRATION AND SEGMENTATION WITHIN AND ACROSS APERTURES

Stephen Grossberg, Ennio Mingolla and Lavanya Viswanathan<sup>1</sup>

Department of Cognitive and Neural Systems  
and  
Center for Adaptive Systems  
Boston University  
677 Beacon Street, Boston, MA 02215

January 2000  
Revised March 2001

**Technical Report CAS/CNS-2000-004**

Correspondence should be addressed to:

Professor Stephen Grossberg  
Department of Cognitive and Neural Systems  
Boston University  
677 Beacon Street, Boston, MA 02215  
*email:* steve@bu.edu  
*fax:* 617-353-7755

**Running Title: Motion Integration and Segmentation**

**Keywords:**

motion integration, motion segmentation, motion capture, aperture problem,  
feature tracking, MT, MST, neural network

- 
1. Authorship in alphabetical order. SG, EM and LV were supported in part by the Defense Advanced Research Projects Agency and the Office of Naval Research (*ONR N00014-95-1-0409*). SG was also supported in part by the National Science Foundation (*NSF IRI-97-20333*), and the Office of Naval Research (*ONR N00014-95-1-0657*). LV was also supported in part by the National Science Foundation (*NSF IRI-94-01659*), and the Office of Naval Research (*ONR N00014-92-J-1309* and *ONR N00014-95-1-0657*).
  2. Acknowledgments: The authors wish to thank Diana Meyers for her valuable assistance in the preparation of the manuscript and figures.

# Abstract

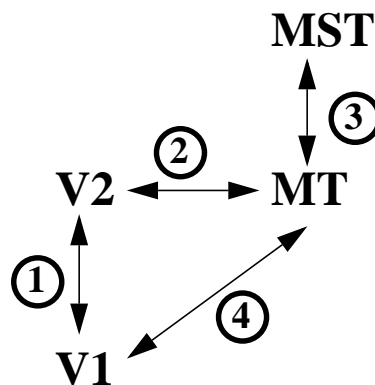
A neural model is developed of how motion integration and segmentation processes, both within and across apertures, compute global motion percepts. Figure-ground properties, such as occlusion, influence which motion signals determine the percept. For visible apertures, a line's terminators do not specify true line motion. For invisible apertures, a line's intrinsic terminators create veridical feature tracking signals. Sparse feature tracking signals can be amplified before they propagate across position and are integrated with ambiguous motion signals within line interiors. This integration process determines the global percept. It is the result of several processing stages: Directional transient cells respond to image transients and input to a directional short-range filter that selectively boosts feature tracking signals with the help of competitive signals. Then a long-range filter inputs to directional cells that pool signals over multiple orientations, opposite contrast polarities, and depths. This all happens no later than cortical area MT. The directional cells activate a directional grouping network, proposed to occur within cortical area MST, within which directions compete to determine a local winner. Enhanced feature tracking signals typically win over ambiguous motion signals. Model MST cells which encode the winning direction feed back to model MT cells, where they boost directionally consistent cell activities and suppress inconsistent activities over the spatial region to which they project. This feedback accomplishes directional and depthful *motion capture* within that region. Model simulations include the barberpole illusion, motion capture, the spotted barberpole, the triple barberpole, the occluded translating square illusion, motion transparency and the chopsticks illusion. Qualitative explanations of illusory contours from translating terminators and plaid adaptation are also given.

# 1. Introduction

Visual motion perception requires the solution of the two complementary problems of *motion integration* and of *motion segmentation*. The former joins nearby motion signals into a single object, while the latter keeps them separate as belonging to different objects. Wallach (1935; translated by Wuerger, Shapley and Rubin, 1996) first showed that the motion of a featureless line seen behind a circular aperture is perceptually ambiguous: for any real direction of motion, the perceived direction is perpendicular to the orientation of the line, called the normal component of motion. This phenomenon was later called the *aperture problem* by Marr and Ullman (1981). The aperture problem is faced by any localized neural motion sensor, such as a neuron in the early visual pathway, which responds to a moving local contour through an aperture-like receptive field. Only when the contour within an aperture contains features, such as line terminators, object corners, or high contrast blobs or dots, can a local motion detector accurately measure the direction and velocity of motion.

To solve the twin problems of motion integration and segmentation, the visual system needs to use the relatively few unambiguous motion signals arising from image features to veto and constrain the more numerous ambiguous signals from contour interiors. In addition, the visual system uses contextual interactions to compute a consistent motion direction and velocity when the scene is devoid of any unambiguous motion signals. This paper develops a neural network model that demonstrates how a hierarchically organized cortical processing stream may be used to explain important data on motion integration and segmentation (Figure 1). An earlier version of the model was briefly reported in Viswanathan, Grossberg, and Mingolla (1999). The Discussion section compares our results with those of alternative models.

FIGURE 1. Neural pathways for interactions between form and motion mechanisms. See text for details.



**1. Vector average.** The vector average solution is one in which the velocity of the plaid appears to be the vector average of the normal components of the plaids constituent gratings (Figure 2)

## 1.1 Plaids: Feature Tracking and Ambiguous Line Interiors

The motion of a grating of parallel lines seen moving behind a circular aperture is ambiguous. However, when two such gratings are superimposed to form a plaid, the perceived motion is not ambiguous. Plaids have therefore been extensively used to study motion perception. Three major

mechanisms for the perceived motion of coherent plaids have been presented in the literature..

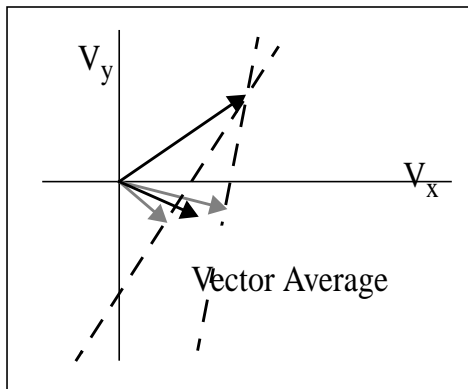


FIGURE 2. Type II plaids: Vector average vs. intersection of constraints (IOC). Dashed lines are the constraint lines for the plaid components. The gray arrows represent the perceived directions of the plaid components. For these two components, the vector average direction of motion is different from the IOC direction.

**2. Intersection of constraints.** A *constraint line* is the locus in velocity space of all possible positions of the leading edge of a bar or line after some time interval  $\Delta t$ . The constraint line for a featureless bar, or a grating of parallel bars, moving behind a circular aperture is parallel to the bar. Adelson and Movshon (1982) suggested that the perceived motion of a plaid pattern follows the velocity vector of the intersection in velocity space of the constraint lines of the plaid components. This *intersection of constraints* (IOC) is the mathematically correct, veridical solution to the motion perception problem. It does not, however, always predict human motion perception even for coherent plaids.

**3. Feature tracking.** When two one-dimensional (1D) gratings are superimposed, they form intersections which act as features whose motion can be reliably tracked. Other features are line endings and object corners. The visual system may track such features. At intersections or object corners, the IOC solution and the trajectory of the feature are the same. In some non-plaid displays described below, feature tracking differs from IOC.

No consensus exists about which mechanism best explains motion perception. Vector averaging tends to uniformize motion signals over discontinuities and efficiently suppresses noise, especially when the features are ambiguous as with features formed by occlusion. However, Adelson and Movshon (1982) showed that observers often do not see motion in the vector average direction. Ferrera and Wilson (1990, 1991) tested this by classifying plaids into Type 1 plaids, for which the IOC lies inside the arc formed by the motion vectors normal to the two components, and Type 2 plaids, for which this is not true (Figure 2). The vector average always lies inside this arc. They found that the motion of Type 2 plaids may be biased away from the IOC solution. Rubin and Hochstein (1993) showed that moving lines can sometimes be seen to move in the vector average, rather than the IOC direction. Mingolla, Todd and Norman (1992), using multiple aperture displays, showed that, in the absence of features, motion was biased toward the vector average. However, when features were visible within apertures, the correct motion direction was perceived. Clearly, the IOC solution does not always predict what the visual system sees.

These data suggest that feature tracking signals as well as the normals to component orientations

contribute to perceived motion direction. Lorenceau and Shiffrar (1992) showed that motion grouping across apertures is prevented by feature tracking signals that capture the motion of the lines to which they belong. In the absence of feature tracking signals, ambiguous signals from line interiors can propagate and combine with similar signals from nearby apertures to select a global motion direction. Consistent with these data, the present model analyzes how both signals from line interiors and feature tracking signals may determine perceived motion direction. Feature tracking signals can propagate across space and veto ambiguous signals from line interiors. Line endings may thus decide the perceived motion direction of the line to which they belong. When such signals are absent, ambiguous signals from line interiors may propagate across space and combine with signals from nearby apertures. Thus, in the absence of feature tracking signals, the model can select the vector average solution.

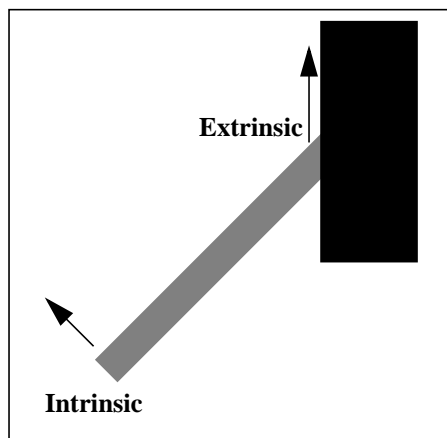


FIGURE 3. Type II plaids: Vector average vs. intersection of constraints (IOC). Dashed lines are the constraint lines for the plaid components. The gray arrows represent the perceived directions of the plaid components. For these two components, the vector average direction of motion is different from the IOC direction.

## 1.2 Intrinsic vs. Extrinsic Terminators

The present model is a synthesis of three earlier models: a model of 3D vision and figure-ground separation, of form-motion interactions, and of motion processing by visual cortex. The first model is needed because not all line terminators are capable of generating feature tracking signals. When a line is occluded by a surface, it is usually perceived as extending behind that surface. The visible boundary between the line and the surface belongs not to the line but to the occluding surface. Nakayama, Shimojo and Silverman (1989) proposed classifying of line terminators into *intrinsic* and *extrinsic* terminators (Figure 3). Bregman (1981) and Kanizsa (1979) earlier used this distinction to create compelling visual displays. The motion of an extrinsic line terminator tells us little about the line's motion. Such motion says more about occluder shape. The motion of an intrinsic line terminator often signals veridical line motion. As we shall soon see, the visual system treats intrinsic terminator motion as veridical signals if their motion is consistent. This makes it possible to fool the visual system by making the occluder invisible by coloring it the same color as the background. Then line terminators may be treated as intrinsic, but their motion is not the line's veridical motion. The preferential treatment displayed by the visual system for

motion signals from intrinsic terminators over those from extrinsic terminators is incorporated into our model through figure-ground processes that detect occlusion events in a scene and assign edge ownership at these locations to near and far depth planes. Such figure-ground processes were modeled as part of the FACADE theory of 3D vision and figure-ground separation; e.g., Grossberg (1994, 1997), Grossberg and Kelly (1999), Grossberg and McLoughlin (1997), Grossberg and Pessoa (1998), and Kelly and Grossberg (2001). FACADE theory describes how 3D boundary and surface representations are generated within the blob and interblob cortical processing streams from cortical area V1 to V2. The theory predicts that the key figure-ground separation processes that are needed for the present analysis are completed within the pale stripes of cortical area V2; see Figure 1. These figure-ground processes help to segregate occluding and occluded objects, along with their terminators, onto different depth planes. The effects of this figure-ground separation process are assumed in the present model in order to make the simulations computationally tractable. The original articles provide explanations and simulations of how the model realizes the desired properties.

How do these figure-ground constraints influence the motion processing that goes on in cortical areas MT and MST? This leads to the need for form-motion interactions, also called formotion interactions. Grossberg (1991) suggested that an interaction from cortical area V2 to MT can modulate motion-sensitive MT cells with the 3D boundary and figure-ground computations that are carried out in V2; see Figure 1. This interaction was predicted to provide MT with completed object boundaries to facilitate object tracking, and with sharper depth estimates of the objects to be tracked. Francis and Grossberg (1996) and Baloch and Grossberg (1997) developed this hypothesis to simulate challenging psychophysical data about long-range apparent motion, notably Korté's laws, as well as data about the line motion illusion, motion induction, and transformational apparent motion.

Chey, Grossberg and Mingolla (1997, 1998) developed the third component model, which is a neural model of biological motion perception by cortical areas V1-MT-MST; see Figure 1. This model is called the *Motion Boundary Contour System* (or *Motion BCS*). It simulated data on how speed perception and discrimination are affected by stimulus contrast and duration, dot density and spatial frequency, among other factors. It also provided an explanation for the barber pole illusion, the conditions under which moving plaids cohere, and how contrast affects their perceived speed and direction. Our model extends the Motion BCS model to account for a larger set of representative data on motion grouping in 3D space, both within a single aperture and across several apertures. Because the model integrates information about form as well as motion perception, it is called the *Formotion BCS* model. The next section describes in detail the design principles underlying the construction of the Formotion BCS model as well as the computations carried out at each stage and their functional significance. Simulation of a moving line illustrates how each stage of the model functions, before other more complex data are explained and simulated.

## 2. Formotion BCS Model

Figure 4 is a macrocircuit showing the flow of information through the model processing stages. We now describe the functional significance of each stage of the model in greater detail.

### 2.1 Level 1: Figure-Ground Preprocessing by the FACADE Model

One sign of occlusion in a 2D picture is a T-junction. The black bar in Figure 5A forms a T-junction with the gray bar. The top of the T belongs to the occluding black bar while the stem belongs

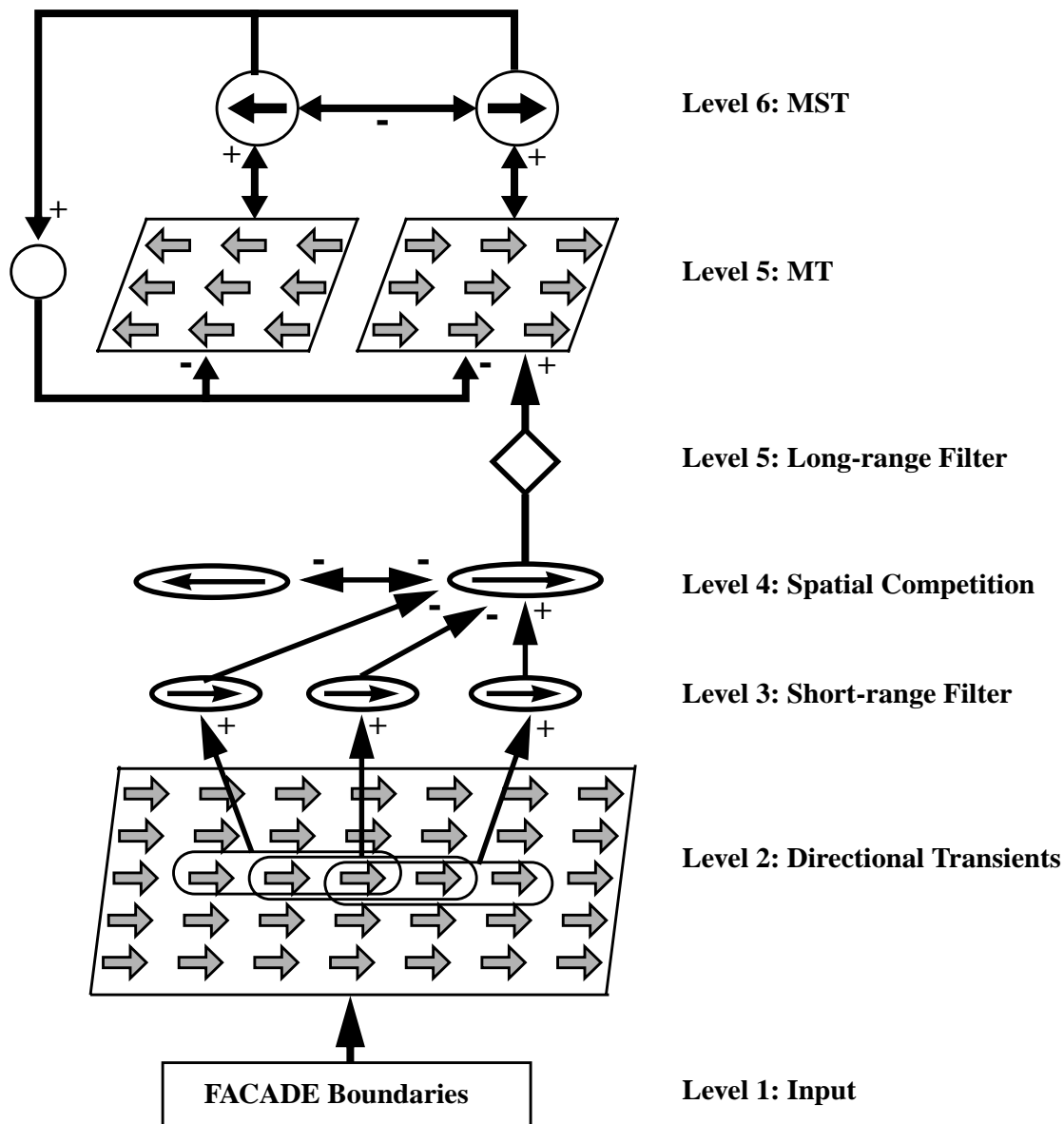


FIGURE 4. Extrinsic vs. intrinsic terminators: the boundary that is caused due to the occlusion of the gray line by the black bar is an extrinsic terminator of the line. This boundary belongs to the occluder rather than the occluded object. The unoccluded terminator of the gray line is called an intrinsic terminator because it belongs to the line itself.

to the occluded gray bar. This boundary ownership operation supports the percept of a black horizontal bar partially occluding a gray vertical bar which lies behind it. When no T-junctions are present in the image, such as in Figure 5B, the two gray regions no longer look occluded. Figures 5A and 5B are two extremes in a continuous series of images wherein the black bar is gradually made gray and then white. When the black horizontal bar is replaced by a horizontal gray bar that is much lighter than the two gray regions, the two gray regions may appear to be separate regions that are each closer than the horizontal gray bar, and not a single region that is partially occluded by it. Because only the relative contrasts, and not the shapes, in this series of images are changed, it illustrates that geometrical and contrastive factors may interact to determine which image regions will be viewed as occluding or occluded objects. In the present data explanations, unambiguous figure-ground separations, like the one in Figure 5A, are assumed to occur. Since extrinsic terminators are generated due to occlusions, T-junctions help distinguish between extrinsic and intrinsic object contours. The present model achieves this by using the FACADE boundary representations that are formed in model cortical area V2. These figure-ground-separated boundaries input to model cortical area MT via a formation interaction from V2 to MT.

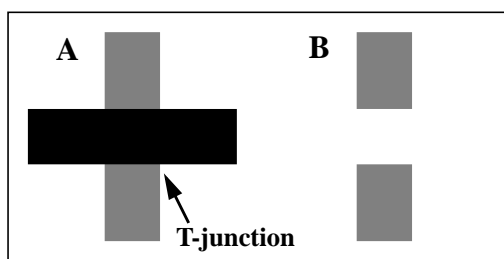


FIGURE 5. T-junctions signalling occlusion. In the 2D image (A), the black bar appears to occlude the gray bar. When the black bar is colored white, and thus made invisible, as in (B), it is harder to perceive the gray regions as belonging to the same object.

The FACADE model detects T-junctions without using T-junction detectors. It uses circuits that includes oriented *bipole cells* (Grossberg and Mingolla, 1985) which model V2 cells reported by von der Heydt, Peterhans and Baumgartner (1984). Consider a horizontally oriented bipole cell, for definiteness. Such a cell can fire if the inputs to each of the two oriented branches of its receptive field are simultaneously sufficiently large, have an (almost) horizontal orientation, and are (almost) collinear. The bipole constraint ensures that the cell fires beyond an oriented contrast such as a line-end only if there is evidence to a link with another similarly oriented contrast, such as a another collinear line-end. Various investigators have reported psychophysical data in support of bipole-like dynamics, including Field *et al.* (1993) and Kellman and Shipley (1992).

At a T-junction, horizontal bipole cells get cooperative support from both sides of their receptive field from the top of the T, while vertical bipole cells only get activation on one side of their receptive field from the stem of the T. As a result, horizontal bipole cells are more strongly activated than vertical bipole cells and win a spatial competition for activation. This cooperative-competitive interaction leads to detachment of the vertical stem of the T at the location where it joins the horizontal top of the T, creating an end-gap in the vertical boundary (Figure 6). This end-gap begins the process whereby the top of the T is assigned to the occluding surface (Grossberg, 1994, 1997). Grossberg, Mingolla and Ross (1997) and Grossberg and Raizada (2000) have predicted how the bipole cell property can be implemented between collinear coaxial pyramidal cells in layer 2/3 of visual cortex via a combination of known long-range excitatory horizontal connections and short-range inhibitory connections that are mediated by interneurons. This implementation of bipole cells has been embedded into a detailed neural model of how the cortical layers are



organized in areas V1 and V2, and how these interactions can be used to quantitatively simulate data about cortical development, learning, grouping, and attention; see Grossberg and Raizada (2000), Grossberg and Williamson (2001), Raizada and Grossberg (2001), and Ross, Grossberg, and Mingolla (2000) for details. Thus accumulating experimental and theoretical evidence support the theory's predictions about how bipole cells initiate the figure-ground separation.

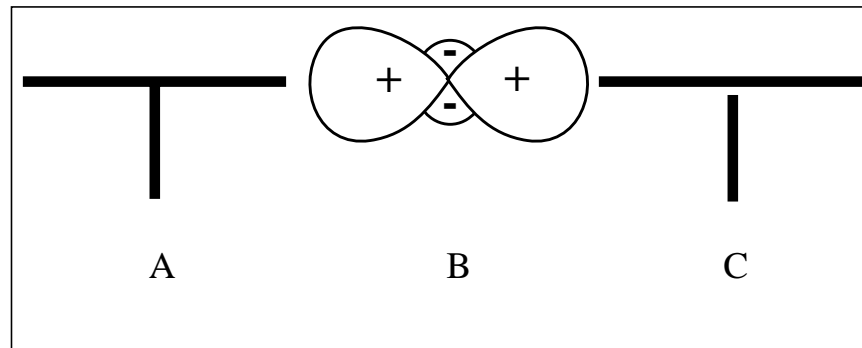


FIGURE 6. (A) T-junctions can signal occlusion. (B) A horizontally-oriented bipole cell (+ signs) can be more fully activated at a T-junction than can a vertically-oriented bipole cell. As a result, the inhibitory interneurons of the horizontal bipole cell (- signs) can inhibit the vertically-oriented bipole cell more than conversely. (C) A break in the vertical boundary that is formed by vertically-oriented bipole cells can then occur. This break is called an *end gap*. End gaps induce the separation of occluding and occluded surface, with the unbroken boundary typically "belonging" exclusively to the occluding surface. [Reprinted with permission from Grossberg, 1997.]





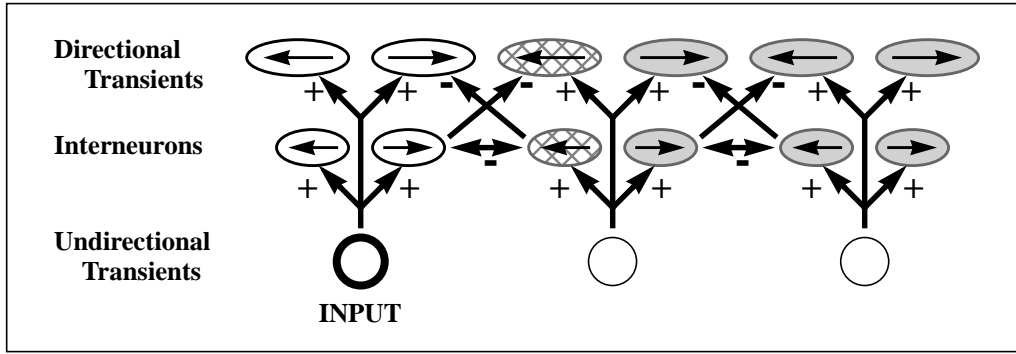
	<b>VISIBLE OCCLUDERS</b>	<b>INVISIBLE OCCLUDERS</b>
<b>Image</b>		
<b>FACADE boundary</b>		

FIGURE 7. FACADE output at the far depth with visible and invisible occluders.

FACADE mechanisms generate the type of boundary representations shown in Figure 7 at the farther depth for a partially occluded line and an unoccluded line. When the occluders are invisible, the occluded line does not appear to be occluded. These boundaries, computed at each frame of a motion sequence, are the model inputs. Any other boundary-processing system that is capable of detecting T-junctions in an image and assigning a depth ordering to the components of the T could

also provide the model inputs.

### A: Frame 1



### B: Frame 2

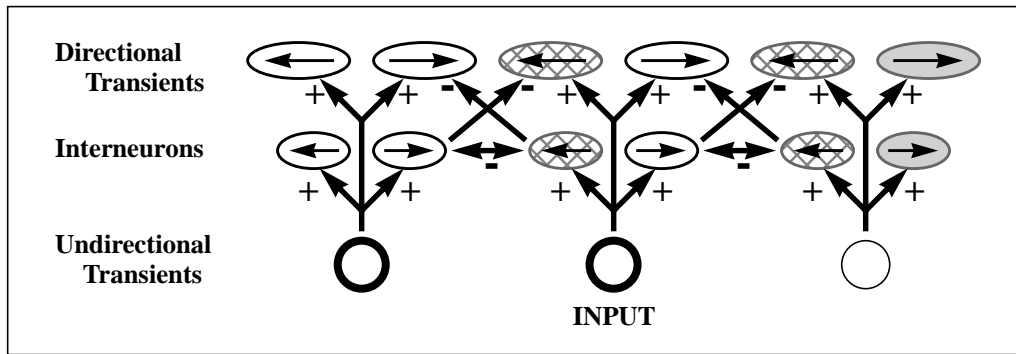


FIGURE 8. Schematic diagram of a 1D implementation of the transient cell network showing the first two frames of the motion sequence. Thick circles represent active undirectional transient cells while thin circles are inactive undirectional transient cells. Ovals containing arrows represent directionally-selective neurons. Unfilled ovals represent active cells, cross-filled ovals are inhibited cells and gray-filled ovals depict inactive cells. Excitatory and inhibitory connections are labelled by ‘+’ and ‘-’ signs respectively.

## 2.2 Level 2: Transient Cells

The second stage of the model comprises undirectional transient cells, directional interneurons and directional transient cells. Undirectional transient cells respond to image transients such as luminance increments and decrements, irrespective of whether they are moving in a particular direction. They are analogous to the Y cells of the retina (Enroth-Cugell and Robson, 1966; Hochstein and Shapley, 1976a, 1976b). A directionally selective neuron fires vigorously when a stimulus is moved through its receptive field in one direction (called the *preferred direction*), while motion in the reverse direction (termed the *null direction*) evokes little response. The connectivity between the three different cell types in Level 2 of the model incorporates three main design principles that are consistent with the available data on directional selectivity in the retina and visual cortex: (a) directional selectivity is the result of asymmetric inhibition along the preferred direction of the cell, (b) inhibition in the null direction is spatially offset from excitation, and (c) inhibition arrives before, and hence vetoes, excitation in the null direction.

Figure 8 shows how asymmetrical directional inhibition works in a 1D simulation of a two-frame motion sequence. When the input arrives at the leftmost transient cell in Frame 1, all interneurons at that location, both leftward-tuned and rightward-tuned, are activated. The rightward-tuned interneuron at this location inhibits the leftward-tuned interneuron and directional cell one unit to the right of the current location. When the input reaches the new location in Frame 2, the leftward-tuned cells, having already been inhibited, can no longer be activated. Only the rightward-tuned cells are activated, consistent with motion from left to right. Further, mutual inhibition between the interneurons ensures that a directional transient cell response is relatively uniform across a wide speed range. Directional transient cells can thus respond to slow and fast speeds. Their outputs for a 2D simulation of a single moving line are shown in Figure 9A. The signals are ambiguous and the effects of the aperture problem are clearly visible.

## 2.3 Level 3: Short-range Filter

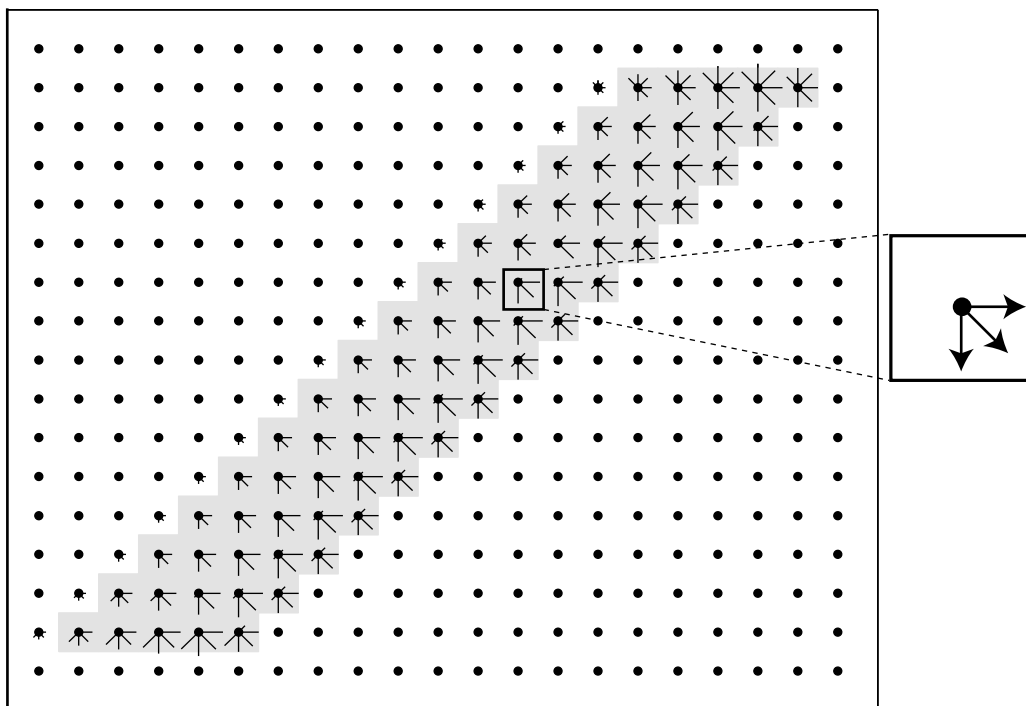
Although known to occur *in vivo*, the veto mechanism described in the previous section exhibits two computational uncertainties in a 2D simulation. First, the short spatial range over which it operates results in the creation of spurious signals near line endings, as can be seen in Figure 9A. Second, vetoing eliminates the wrong (or null) direction, but does not selectively activate the correct direction. It is important to suppress spurious directional signals while amplifying the correct motion direction at line endings because these unambiguous feature tracking signals must be made strong enough to track the correct motion direction and to overcome the much more numerous ambiguous signals from line interiors. In Level 3 of the model (see Figure 4), the directional transient cell signals are space- and time-averaged by a short-range filter cell that accumulates evidence from directional transient cells of similar directional preference within a spatially anisotropic region that is oriented along the preferred direction of the cell. This computation strengthens feature tracking signals at unoccluded line endings, object corners and other scenic features. It is not necessary to first identify form discontinuities that may constitute features and then to match their positions from frame to frame. We thus avoid the *feature correspondence problem* which correlational models (Reichardt, 1961; van Santen and Sperling, 1985) need to solve.

The short-range filter uses multiple spatial scales. Each scale responds preferentially to a specific speed range. Larger scales respond better to faster speeds by thresholding short-range filter outputs with a self-similar threshold; that is, a threshold that increases with filter size. Larger scales thus require "more evidence" to fire (Chey, Grossberg, and Mingolla, 1998). Outputs for a single moving line are shown in Figure 9B. Feature tracking signals occur at line endings, while the line interior exhibits the aperture problem.

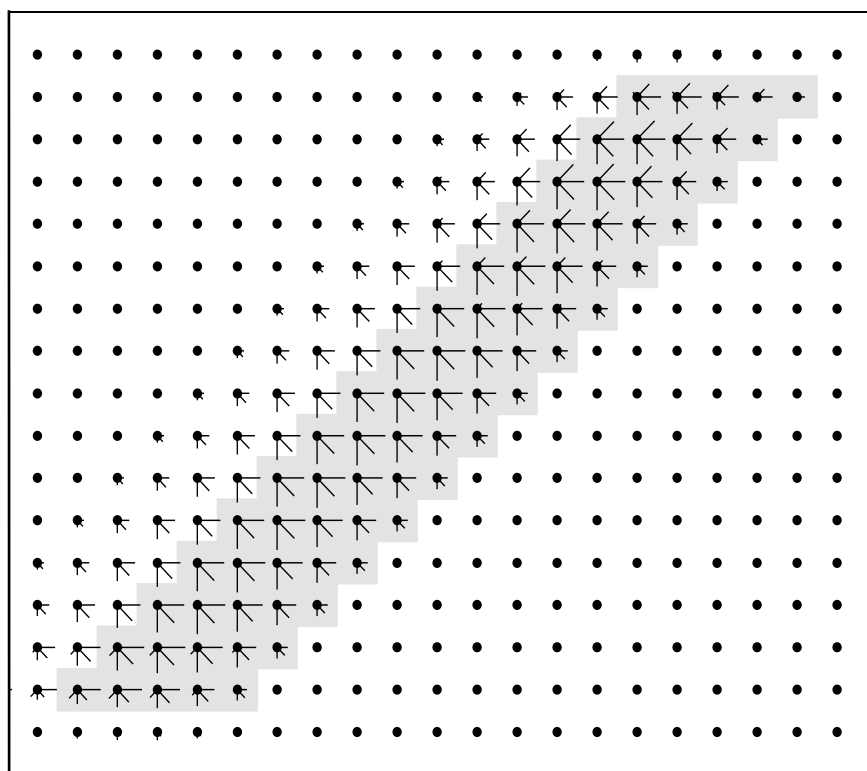
## 2.4 Level 4: Spatial Competition and Opponent Direction Inhibition

Spatial competition among cells of the same spatial scale and that prefer the same motion direction further boosts the amplitude of feature tracking signals relative to that of ambiguous signals. This contrast-enhancing operation within each direction works because feature tracking signals, being at motion discontinuities, tend to get less inhibition than ambiguous motion signals that lie within an object interior. This enhancement occurs without making the signals from line interiors so small that they will be unable to group across apertures in the absence of feature tracking signals. Spatial competition also works with the self-similar thresholds to generate speed tuning

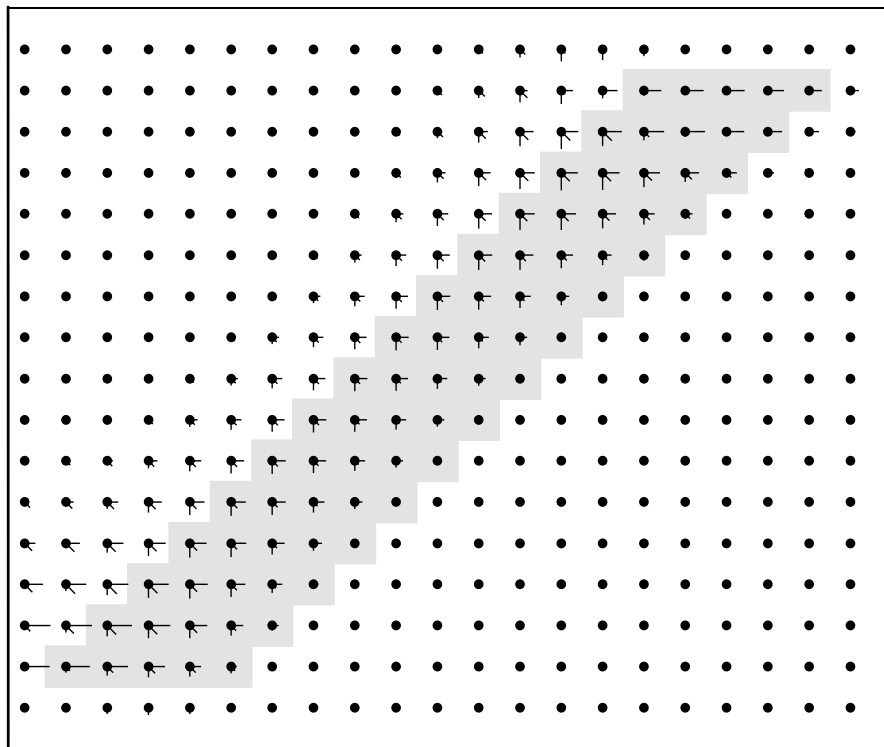
A



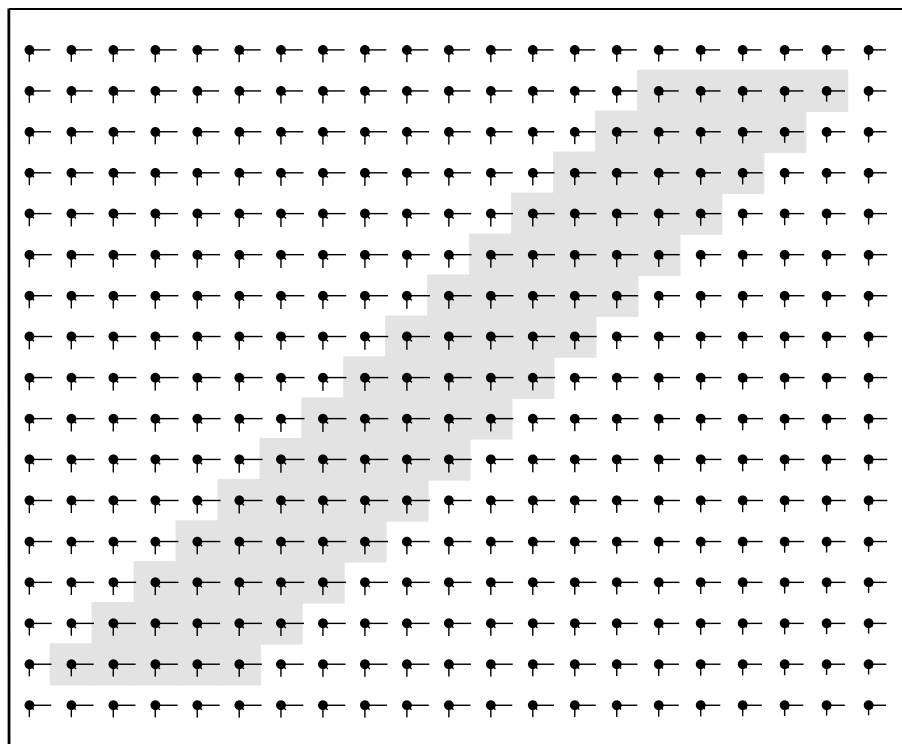
B



C



D



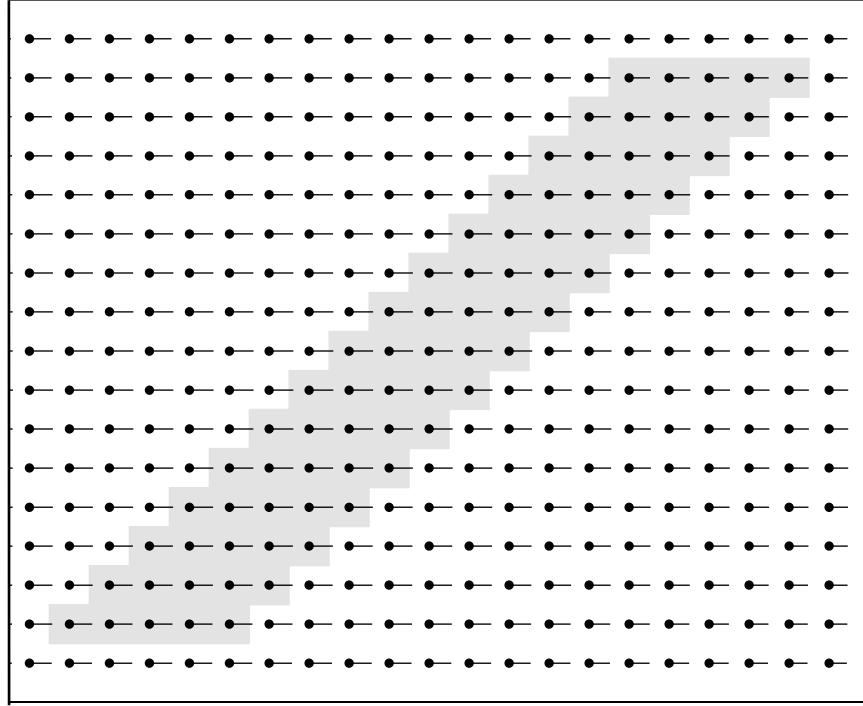
**E**

FIGURE 9. Model activities for a 2D simulation of a moving tilted line. (A) Directional transient cells. (B) Thresholded short-range filter cells. (C) Competition network cells. (D) MT cells. (E) MST cells: model output. The gray region in each diagram represents the position of the input at the current frame. The inset diagram in (A) enlarges the activities of cells at one x-y location. The dot represents the center of the x-y pixel. Since all simulations in this paper use eight directions, there are eight cells, each with a different directional tuning at every spatial location. At the location shown, three of the eight cells, those tuned to east, south-east and south directions, are active. This is depicted through velocity vectors oriented along the preferred directions of each cell. The length of each vector is proportional to the activity of the corresponding cell. This convention is used for all the model outputs in the paper. The simulations for panels (a) - (e) were done on a 30 X 17 grid of locations; the leftmost 9 columns of the grid were cropped for figure display.

curves for each scale; see Chey, Grossberg, and Mingolla (1998).

This model stage also uses opponent inhibition between cells tuned to opposite directions; cf., Albright (1984) and Albright, Desimone, and Gross (1984). This ensures that cells tuned to opposite motion directions are not simultaneously active. Outputs for a moving line are shown in Figure 9C. Feature tracking signals are highly selective and larger than ambiguous signals.

## 2.5 Levels 5 and 6: Long-range Filter, Directional Grouping, and Attentional Priming

Levels 5 and 6 of the model consists of two cell processing stages, which are described together because they are linked by a feedback network. Level 5 models a spatially long-range filter and its effect on model MT cells. Level 6 models MST cells. The long-range filter pools signals, over larger spatial areas than the short-range filter of similar directional preference, opposite contrast

polarity, and multiple orientations. It turns MT cells into true "directional" cells. A model MT cell can, for example, pool evidence about diagonal motion of a rectangular object that is lighter than its background from both the vertical dark-to-light leading edge of the rectangle and the horizontal light-to-dark trailing edge. This pooling operation is also depth-selective, so it is restricted to cells of the same scale that are tuned to the same direction. Despite this directional selectivity, the network can respond to a band of motion directions at ambiguous locations due to the aperture problem, as in Figure 9C. Thus, although the model MT cells are competent directional motion detectors, they cannot, by themselves, solve the aperture problem. A suitably defined feedback interaction between the model MT and MST cells solves the aperture problem by triggering a wave of motion capture that can travel from feature tracking signals to the locations of ambiguous motion signals. This feedback interaction comprises the grouping, matching, and attentional priming network of the Formotion BCS model. It works as follows.

Bottom-up directional signals from model MT cells activate like-directional MST cells, which interact via a winner-take-all competition across directions. We propose that this occurs in ventral MST, which has large directionally tuned receptive fields that are specialized for detecting moving objects (Tanaka, Sugita, Moriya, and Saito, 1993). The winning direction is then fed back down to MT through a top-down matching and attentional priming pathway that influences a region that surrounds the location of the MST cell (Figure 4). Cells tuned to the winning direction in MST have an excitatory influence on MT cells tuned to the same direction. However, they also nonspecifically inhibit all directionally tuned cells in MT. For the winning direction, the excitation cancels the inhibition, so the winning direction survives the top-down matching process, and may even be a little amplified by it. But for all other directions, having lost the competition in MST and not receiving excitation from MST to MT, there is net inhibition in MT. This matching process within MT by MST leads to net suppression of all directions other than the winning direction within a region surrounding a winning cell. If the winning cell happens to correspond to a feature tracking signal, then the direction of the feature tracking signal is selected within the spatial region that its top-down matching signals influence, due to the relatively large size of feature tracking signals compared with ambiguous motion signals. This selection, or motion capture, process creates a region dominated by the direction of the feature tracking signal. The bottom-up signals from MT to MST from this region then force the direction of the feature tracking signal to win in MST. Feedback from MST to MT then allows the feature tracking direction to suppress more ambiguous motion signals in the contiguous region of MT via top-down matching signals. A feature tracking signal can hereby propagate its direction into the interior of the object, much like a travelling wave, using unidirectional bottom-up and top-down feedback exchanges between model MT and MST. Motion capture is hereby achieved, as shown in Figures 9D and 9E, which display the activities of MT and MST cells after feedback has a chance to respond to a single tilted line moving to the right.

Motion capture is a *preattentive* process, since it is driven by bottom-up signals, even though it makes essential use of top-down feedback. This particular kind of top-down matching process can select winning directions, without unduly biasing their speed signals (Chey, Grossberg, and Mingolla, 1997), while suppressing losing directions. Such a matching process has also been used for top-down *attentional priming*. This kind of attentional priming was proposed by Carpenter and Grossberg (1987) as part of Adaptive Resonance Theory (ART). In the present instance, it realizes a type of directional priming, which is known to exist (Groner, Hofer, and Groner, 1986; Sekuler and Ball, 1977; Stelmach, Herdman, and McNeil, 1994). Cavanagh (1992) has described an attention-based motion process, in addition to low-level or automatic motion processes, and has shown that it provides accurate velocity judgments. The facts that ART-style MST-to-MT matching preserves the velocity estimates of attended cells, and suppresses aperture-ambiguous direc-

tion and velocity estimates, are consistent with his data. Neural data are also consistent with this attentional effect. Treue and Maunsell (1996) have shown that attention can modulate motion processing in cortical areas MT and MST in behaving macaque monkeys. O’Craven *et al.* (1997) have shown by using fMRI that attention can modulate the MT/MST complex in humans.

These data are consistent with the following model predictions. One prediction is that the same MT/MST feedback circuit that accomplishes *preattentive* motion capture also carries out *attentive* directional priming. Cooling ventral MST should prevent MT cells from exhibiting motion capture in the aperture-ambiguous interiors of moving objects. Another prediction is that a directional attentional prime can reorganize preattentive motion capture. A third prediction derives from the fact that MST-to-MT feedback is predicted to carry out ART matching, which has been predicted to help stabilize cortical learning (Carpenter and Grossberg, 1987; Grossberg, 1980, 1999b). This property suggests how directional receptive fields develop and maintain themselves. In addition, it is predicted that inhibition of the MT-to-MST bottom-up adaptive weights can prevent directional MST cells from forming, and inhibition of the MST-to-MT adaptive weights can destabilize learning in the bottom-up adaptive weights. Grossberg (1999a) has also proposed how top-down ART attention is realized within the laminar circuits from V2-to-V1, and by extension from MST-to-MT; also see Grossberg and Raizada (2000) and Raizada and Grossberg (2001). By extension, a predicted attentional pathway is from layer 6 of ventral MST to layer 6 of MT (possibly by a multi-synaptic pathway from layer 6 of MST to layer 1 apical dendrites of layer 5 MT cells that project to layer 6 MT cells) followed by activation of a modulatory on-center off-surround network from layer 6-to-4 of MT. Preattentive motion capture signals, as well as directional attentional priming signals, from MST are hereby predicted to strongly activate layer 6 of MT, to modulate MT layer 4 cells via the on-center, and to inhibit layer 4 cells in the off-surround.

### 3. Model Computer Simulations

This section describes some motion percepts and how the model explains them.

#### 3.1 Classic Barber Pole

Due to the aperture problem, the motion of a line seen behind a circular aperture is ambiguous. The same is true for a grating of parallel lines moving coherently. Wallach (1935) showed that if such a grating is viewed behind an invisible rectangular aperture, then the grating appears to move in the direction of the longer aperture edge of the aperture. For the horizontal aperture, in Figure 10A, the grating appears to move horizontally from left to right, as in Figure 10B.

Line terminators help to explain this illusion by acting as features with unambiguous motion signals (Hildreth, 1984; Nakayama and Silverman, 1988a, 1988b). As in the tilted line simulation, our model uses line terminators to generate feature tracking signals. In the short-range filter stage (Level 3), line terminators generate feature tracking signals that are strengthened by spatial competition (Level 4). In a horizontal rectangular aperture, there are more line terminators along the horizontal direction than along the vertical direction (Figure 10). Hence there are more feature tracking signals signalling rightward than downward motion. Rightward motion therefore wins in the interdirectional competition of the long-range directional grouping MT-MST network. Top-down priming of the winning motion direction from MST to MT suppresses all losing directions across MT. Thus, in the presence of multiple feature tracking signals (here, grating terminators) that signal motion in different directions, interdirectional and spatial competition ensure that the



direction favored by the majority of features determines the global motion percept as shown in the simulation in Figure 11A.

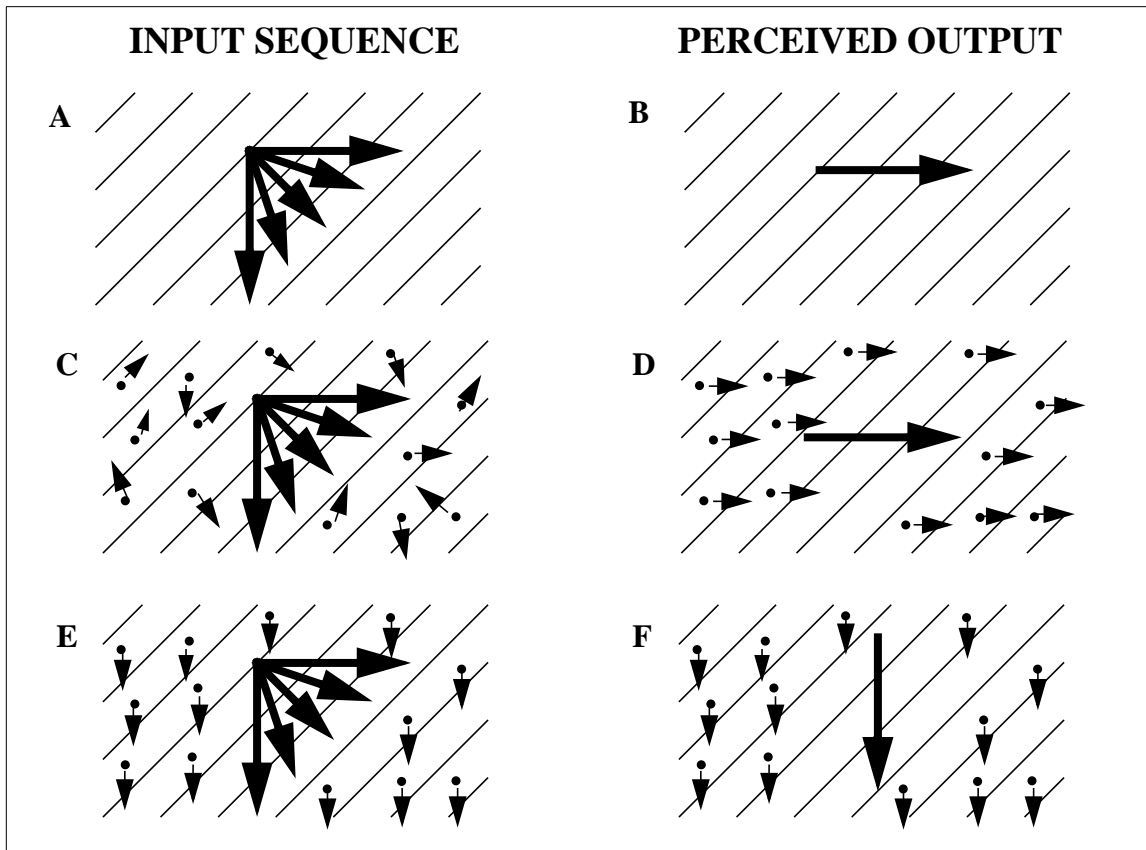


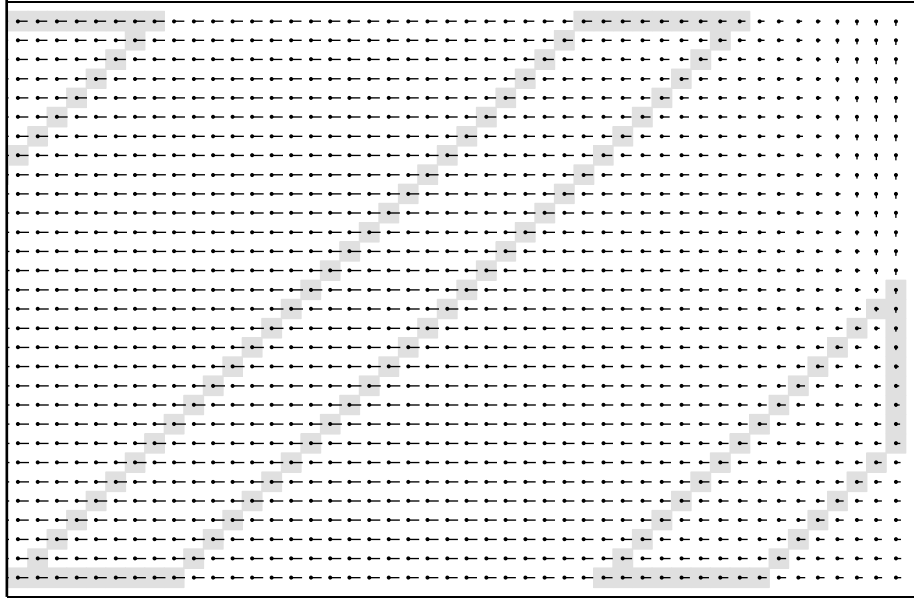
FIGURE 10. Moving grating illusions. The left column shows the physical stimulus presented to observers and the right column depicts their percept. (A,B) Classic barber pole illusion. (C,D) Motion capture. (E,F) Spotted barber pole illusion.

### 3.2 Motion Capture

The barber pole illusion demonstrates how the motion of a line is determined by unambiguous signals formed at its terminators. Are motion signals restricted to propagate only from unambiguous motion regions to ambiguous motion regions within the same object or can they also propagate from unambiguous motion regions of an object to nearby ambiguous motion regions of other objects? Ramachandran and Inada (1985) addressed this question with a motion sequence in which random dots were superimposed on a classic barber pole pattern such that the dots on any one frame of the sequence were completely uncorrelated with the dots on the subsequent frame. Despite the noisiness of the dot motion signals from frame to frame, subjects saw the dots move in the same direction as the barber pole grating (Figures 10C and 10D). The dot motion was *captured* by the grating motion. Solving the aperture problem is also a form of motion capture.

The Formotion BCS model explains motion capture as follows: Since the dots are not stationary but flickering, they activate transient cells in Level 2. However, due to the noisy and inconsistent dot motion in consecutive frames, no feature tracking signals are generated for the dots in the short-range filter. The dot signals lose the competition in the MT-MST loop. The winning barber

A



B

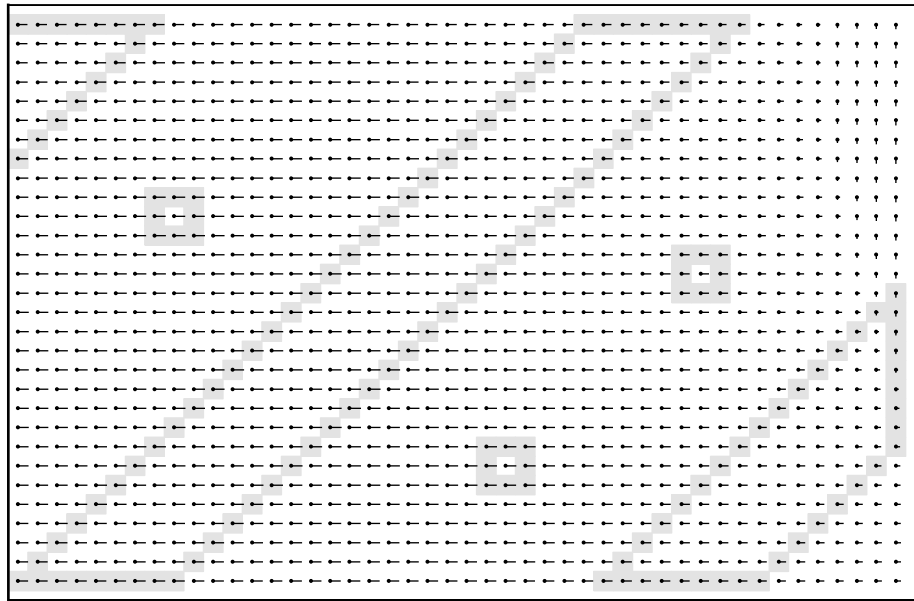


Figure 11. Caption on next page.

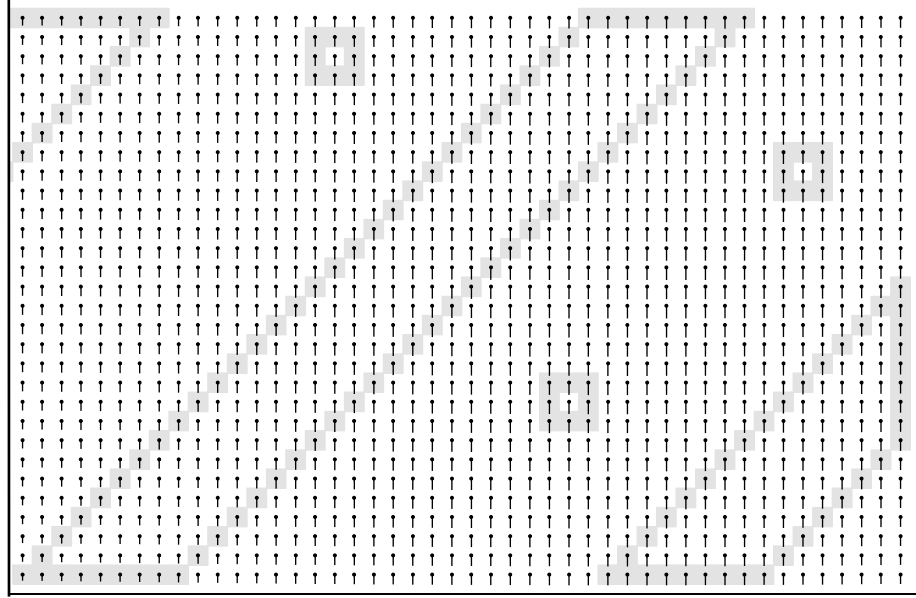
**C**

FIGURE 11. Model MST outputs for the grating illusions. (A) Classic barber pole illusion. (B) Motion capture. (C) Spotted barber pole illusion. The simulations for panels (a) - (c) were done on a 60 X 30 grid of locations; the leftmost 14 columns of the grid were cropped for figure display.

pole motion direction inhibits the inconsistent motion directions of the dots, which now appear to move with the grating, as shown in the computer simulation of Figure 11B.

### 3.3 Spotted Barber Pole

The spotted barber pole (Shiffrar, Li, and Lorenceau, 1995) also involves superposition of random dots on a barber pole, as in motion capture. Unlike motion capture, the dots move coherently downwards (Figure 10E). Observers here see the grating move downwards with the dots (Figure 10F). Thus, the motion of the dots now captures the perceived motion of the grating.

This phenomenon may seem to be difficult to explain. One may expect that, as in the classic barber pole, for each line of the grating, the unambiguous motion of its terminators would determine its perceived motion. Since the stimulus contains more lines with rightward moving terminators than downward moving terminators, it would seem that the grating should appear to move rightward rather than downward. However, unambiguous motion signals need not propagate only within a single object. They can also influence the perceived motion of spatially adjacent regions using long-range filter kernels that are large enough to overlap feature tracking signals from spatially contiguous regions. The superimposed dots thus generate strong feature tracking signals signalling downward motion. When these downward signals combine with those produced by the few downward moving grating terminators, they outnumber the rightward signals formed by the remaining grating terminators. Downward energy predominates over rightward energy in the MT-MST loop and wins the interdirectional competition. Both grating and dots appear to move downward, as shown in the computer simulation of Figure 11C.

### 3.4 Line Capture

The previous simulations have demonstrated the importance of line terminators in determining the perceived motion direction. However, all terminators are not created equal. While intrinsic terminators appear to belong to the line, extrinsic terminators, which are artifacts of occlusion, do not. The following simulations, which are related to the motion capture stimuli of Ramachandran and Inada (1985), predict how the visual system assigns differing degrees of importance to intrinsic and extrinsic terminators to determine the global direction of motion in a scene.

#### 3.4.1 Partially Occluded Line

When a line's terminators are occluded and thus extrinsic, their motion signals are ambiguous. In the absence of other disambiguating motion signals, the visual system accepts the motion of these terminators as the most likely candidate for the line's motion (Figure 12A). Extrinsic terminators can produce feature tracking signals, but these are weaker than those produced by intrinsic terminators. They play a role in determining the global percept (Figure 12B) only when intrinsic features are lacking. This effect is simulated in Figure 13A.

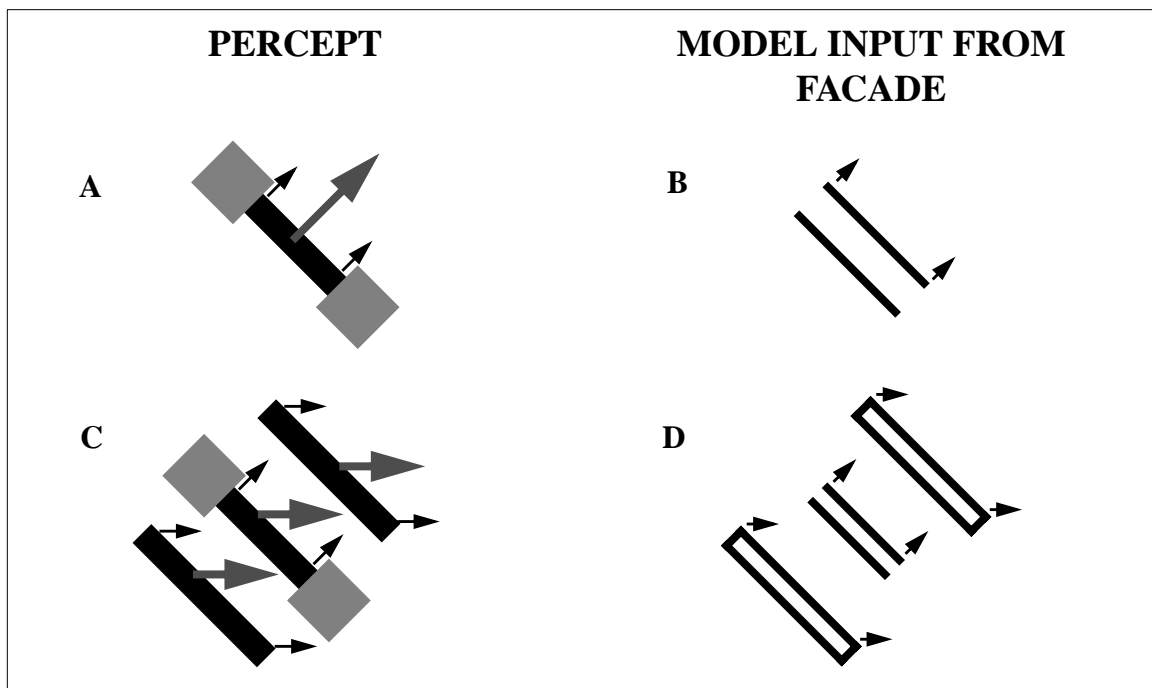
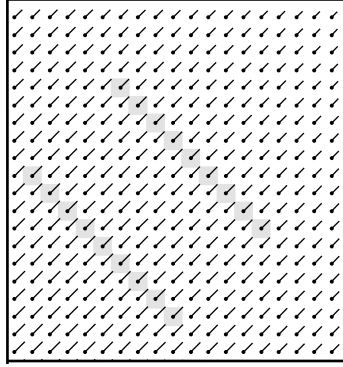


FIGURE 12. Line capture stimuli: Percept and model input from FACADE. Small arrows near line terminators depict the actual motion of the terminators. Larger gray arrows represent the perceived motion of the lines. (A,B) Single line translating behind visible rectangular occluders. (C,D) Line behind visible occluders with flanking unoccluded rightward moving lines.

#### 3.4.2 Horizontal Line Capture

When the same partially occluded line is presented with flanking unoccluded lines (Figure 12C), the perceived motion of the ambiguous line is captured by the unambiguous motion of the flank-

A



B

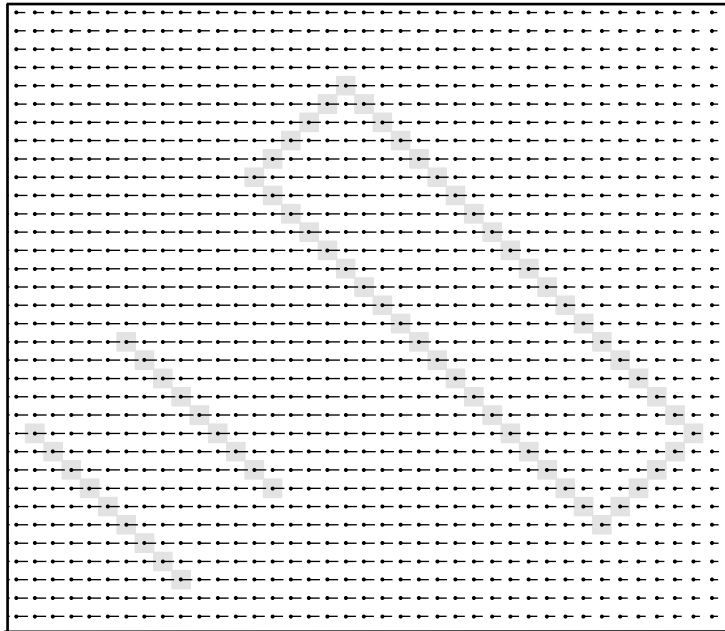


FIGURE 13. Model MST output for line capture. (A) Partially occluded line. (B) Horizontal line capture. The simulation for panel (a) was done on a 31 X 31 grid of locations; the leftmost 12 columns and bottommost 11 rows of the grid were cropped for figure display. The simulation for panel (b) was done on a 71 X 71 grid of locations; the leftmost 32 columns and bottommost 31 rows of the grid were cropped for figure display. The cropped region included another line input, identical in shape, orientation, motion to the one displayed in the upper right of the grid in panel (b).

ing lines. The terminators of the unoccluded lines, being intrinsic, generate strong feature tracking signals in the short-range filter (Figure 12D). These can capture not only the motion of the line that they belong to but also that of nearby ambiguous regions, such as the partially occluded line which only has extrinsic terminators, as shown in the computer simulation in Figure 13B).

### 3.5 Triple Barber Pole

Shimojo, Silverman and Nakayama (1989) studied the relative strength of feature tracking signals

at intrinsic and extrinsic line terminators. They combined three barber pole patterns (Figure 14). When the occluding bars are visible (when the horizontal barber pole terminators are extrinsic), observers saw a single downward-moving vertical barber pole behind the occluding bars. When the occluding bars are invisible (when the barber pole terminators are intrinsic), the percept was of three rightward-moving horizontal barber pole patterns. The similar Tommasi and Vallortigara (1999) experiment emphasized figure-ground segregation in the percept.

The three barber pole gratings appear to move rightward when the occluders are invisible because, in each grating, rightward moving terminators outnumber downward moving terminators. Although this is still true with visible occluders, the rightward moving line endings, being extrinsic, produce very weak feature tracking signals while the downward moving endings, being intrinsic, produce strong feature tracking signals. Downward activities, although fewer, are larger than the more numerous, but weaker, rightward activities, so downward motion wins the MT-MST competition. Figures 15A and 15B show simulations of cases 14A and 14B, respectively.

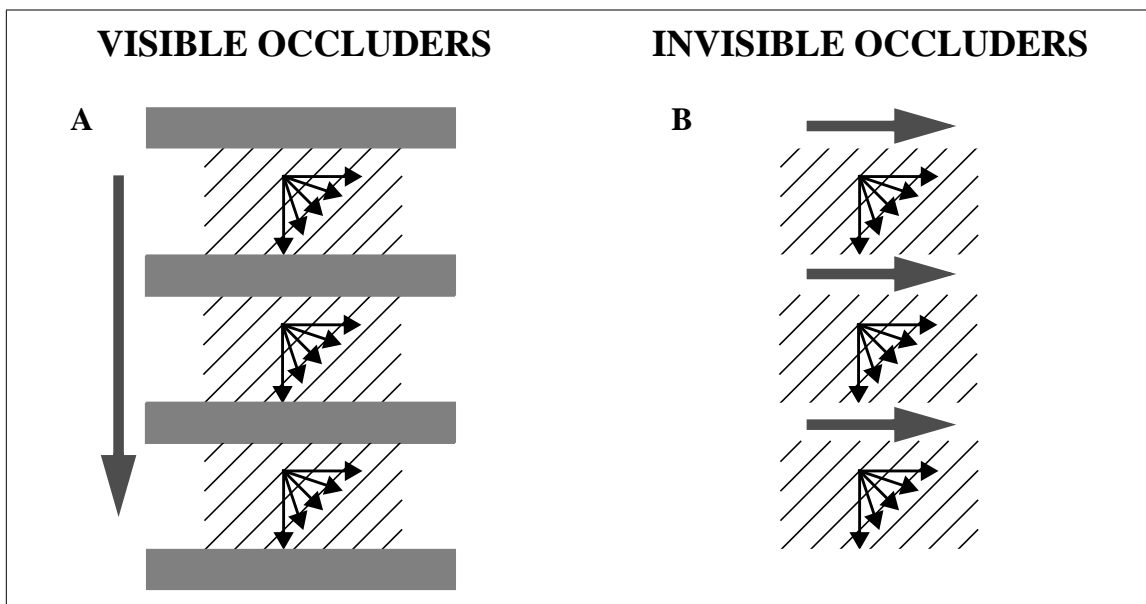


FIGURE 14. Triple Barber Pole. Thin black arrows represent the possible physical motions of the barber pole patterns. Thick gray arrows represent the perceived motion of the gratings.

### 3.6 Translating Square seen behind Multiple Apertures

All the phenomena described so far involved integration of motion signals into a global percept. We now describe data in which the nature of terminators is solely responsible for whether motion integration or segmentation takes place. Lorenceau and Shiffrar (1992) studied the effect of aperture shape and color on how humans group local motion signals into a global percept. Since the physical motion in each of the three cases described below is identical and the only parameters

varied are the occluder luminance and shape, a solution computed on the basis of the intersection of constraints (IOC) model (Adelson and Movshon, 1982) would predict the same percept for each case. The percept, however, varies widely and depends entirely on the strength of the feature tracking signals generated in each case.

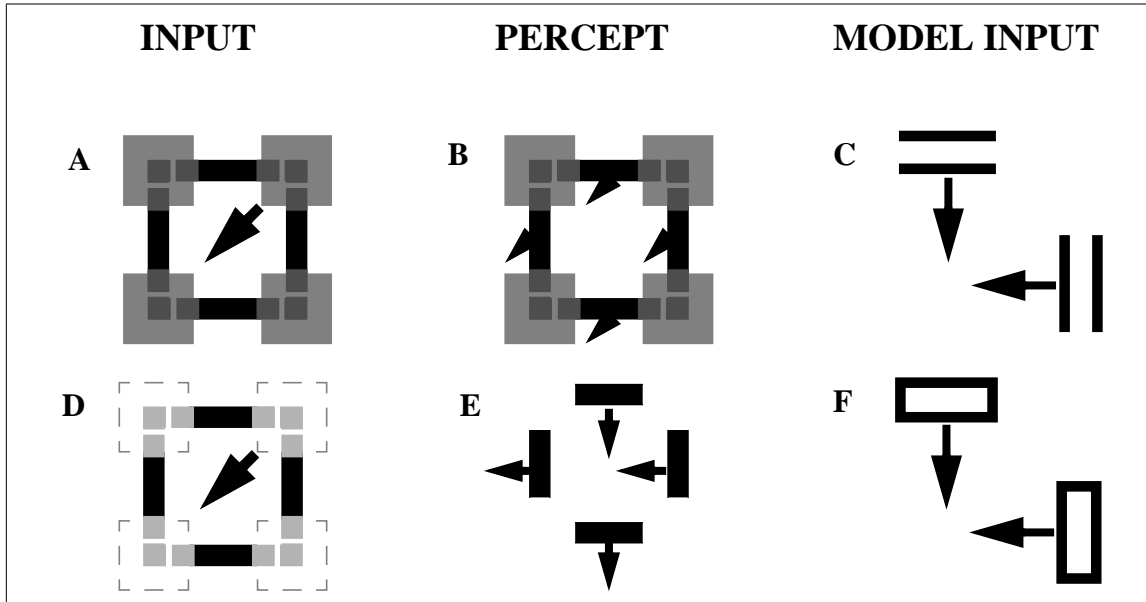


FIGURE 16. Square translating behind rectangular occluders. (A,B,C) Visible occluders. Dark gray dashed lines represent the corners of the square that are never visible during the translatory motion of the square. (D,E,F) Invisible occluders. Light gray dashed lines depict the invisible

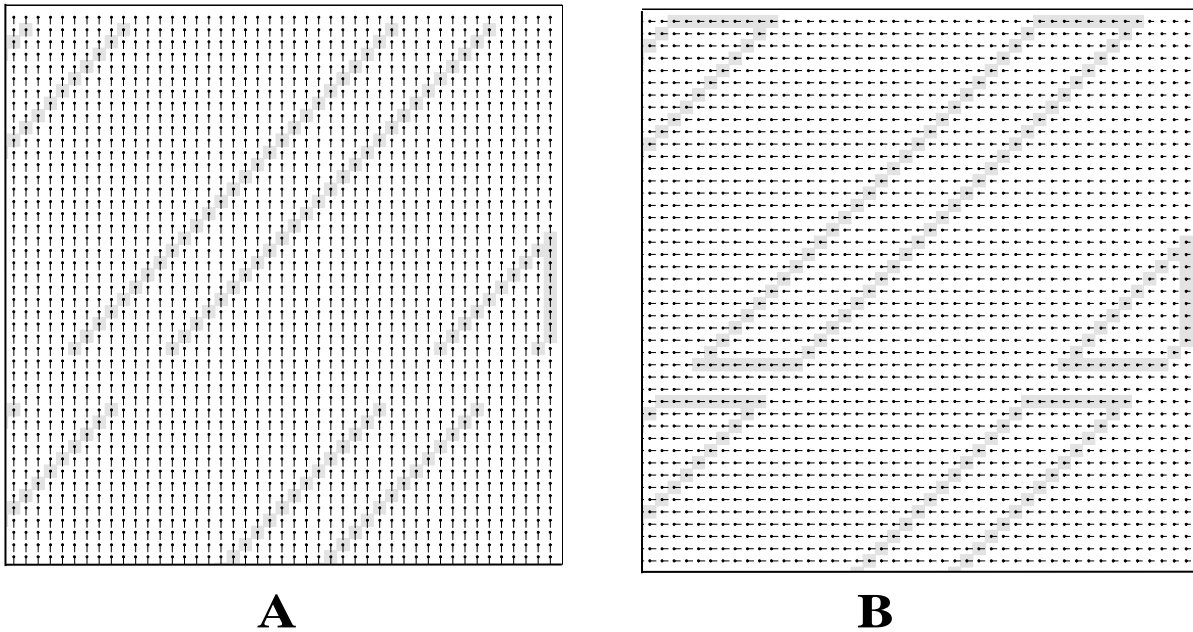


FIGURE 15. Model MST output for the triple barber pole illusion. (A) Visible occluders, i.e., extrinsic horizontal line terminators. (B) Invisible occluders, i.e., intrinsic horizontal line terminators. The simulations for panels (a) and (b) were done on a 60 X 90 grid of locations; the leftmost 15 columns and bottommost 35 rows of the grid were cropped for figure display. The cropped area contained inputs that continued the pattern shown, with a second horizontal line cutting across diagonal lines.

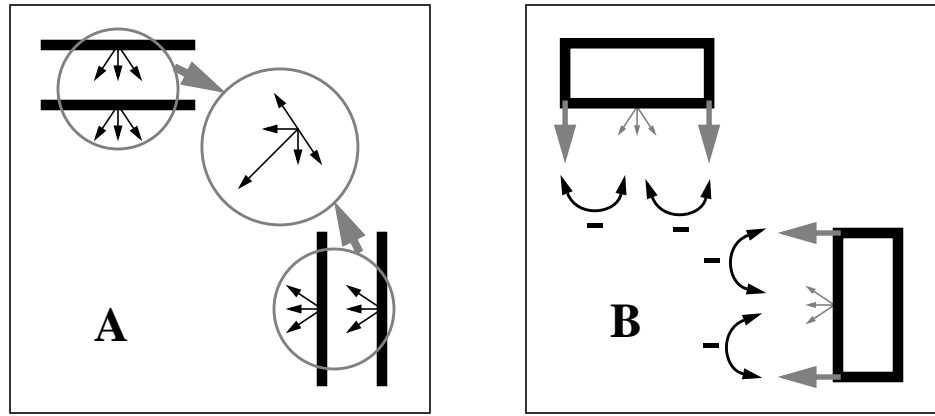


FIGURE 17. Schematic of how model mechanisms explain the translating square illusion. (A) when occluders are visible, motion integration across apertures takes place. (B) when occluders are invisible, motion segmentation occurs.

corners of the square; dashed rectangular outlines represent the invisible occluders that define the edges of the apertures.

### 3.6.1 Visible Rectangular Occluders

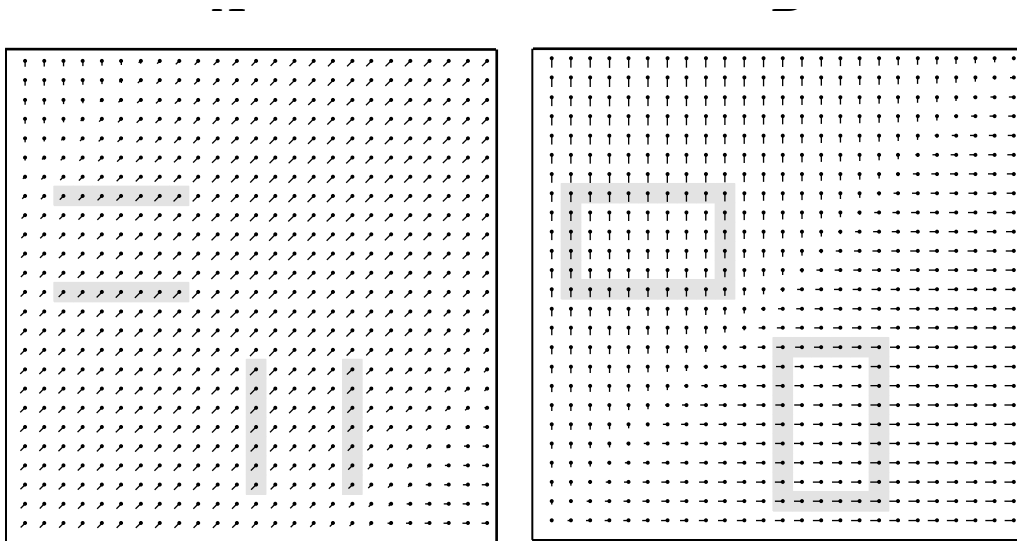
Suppose that a square translates behind four visible rectangular occluders (Figure 16A) such that the corners of the square (potential features) are never visible during the motion sequence. Observers are then able to amodally complete the corners of the square and see it consistently translating southwest (Figure 16B). For computational simplicity, we can, without loss of generality, consider just the top and right sides of the square (Figure 16C). When the occluders are visible, the extrinsic line terminators generate weak feature tracking signals that are unable to block the spread of ambiguous signals from line interiors across apertures. The southwest direction gets activated from both apertures, while the other directions only get support from one of the two apertures (Figure 17A). This is because the ambiguous motion positions activate a range of motion directions, including oblique directions, in addition to the direction perpendicular to the moving edge. The southwest direction hereby wins the interdirectional competition in MST. Top-down priming from MST to MT boosts the southwest motion signals while suppressing all others (Figure 17A). Thus, in the model computer simulation, both lines appear to move in the same diagonal direction (Figure 18A). Motion *integration* of local motion signals is said to occur.

### 3.6.2 Invisible Rectangular Occluders

This display is identical to the previous one except that the occluders are made invisible by making them the same color as the background (Figure 16D). This small change drastically affects the percept. Now, observers can no longer tell that the lines belong to a single object, a square, that is translating southwest. The lines appear to move independently in horizontal and vertical directions (Figure 16E). Consider only the square's top and right sides (Figure 16F). The intrinsic line terminators of each line produce strong feature tracking signals that veto the ambiguous interior signals. Each line appears to move in the direction of its terminators. The intrinsic terminators thus effectively block the grouping of signals from line interiors across apertures (Figure 17B). Motion *segmentation* occurs, as shown in the computer simulation in Figure 18B.

The role of inhibition between motion signals from line endings and line interiors was empha-





**C**

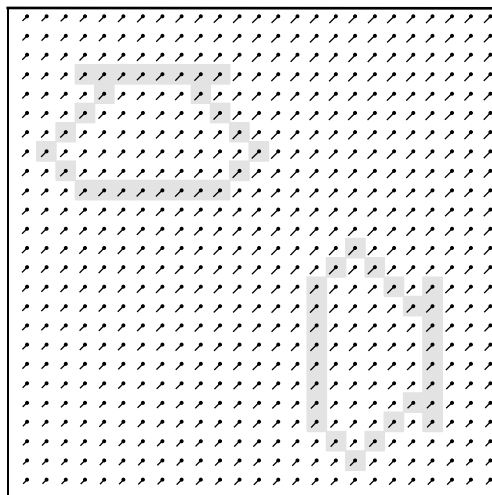


FIGURE 18. Model MST output for the translating square behind multiple apertures. (A) Visible rectangular occluders. (B) Invisible rectangular occluders. (C) Invisible jagged occluders. The simulations for panels (a) and (b) were done on a 33 X 33 grid of locations; the rightmost 8 columns and topmost 8 rows of the grid were cropped for figure display. The simulation for panel (c) was done on a 37 X 37 grid of locations; the rightmost 12 columns

sized by Giersch and Lorenceau (1999). They boosted inhibition through the use of lorazepam, a substance that facilitates the fixation of inhibitory neurotransmitter GABA on GABAA receptors. This selectively affected performance in the invisible rectangular occluders case, but not in the visible rectangular occluders case. Enhanced inhibition did not affect motion integration when the occluders were visible, but it boosted motion segmentation when the occluders were invisible.

### 3.6.3 Invisible Jagged Occluders

Lorenceau and Shiffrar (1992) showed that if the occluders are invisible as but jagged instead of rectangular, then observers can group individual line motions into a percept of a translating square (Figure 19). Clearly, intrinsic terminators do not always generate feature tracking signals that are

strong enough to block motion grouping across apertures. The jagged edges cause the motion of the line terminators to change direction constantly. The short-range filter is then unable to accumulate enough evidence for motion along any particular direction at line endings, so strong feature tracking signals are not produced. Signals from line interiors can again group across apertures, as shown in the computer simulation in Figure 18C. In summary, for features such as line endings and dots to produce reliable feature tracking signals, they must be intrinsic and generate sufficient evidence for consistent motion in a particular direction.

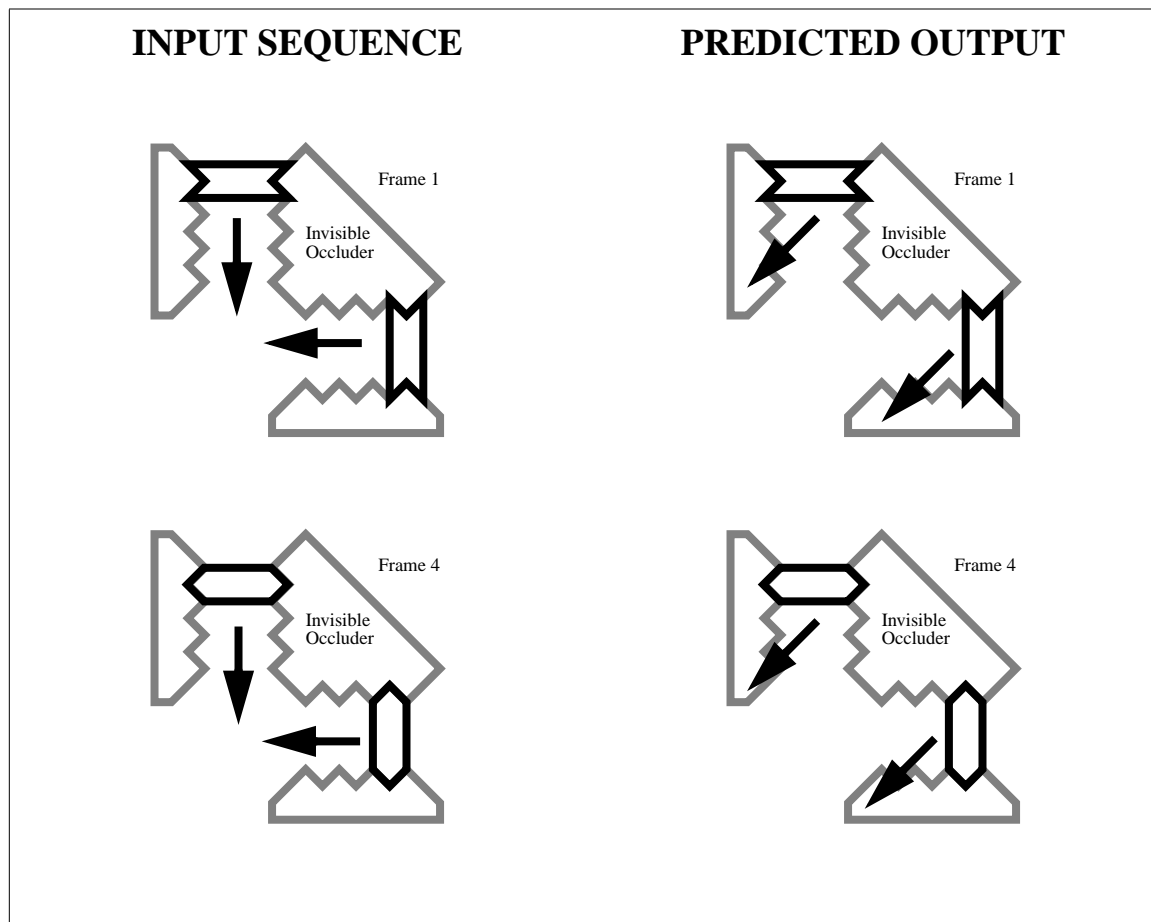


FIGURE 19. (A) Square translating behind invisible jagged apertures: Model input and predicted output. (B) 20B: (B) Opposite motion directions within multiple scales compete. In addition, directions within scales that represent nearer motions inhibit the same directions within scales that represent farther motions. This type of "asymmetry between near and far" is also found in FACADE theory.

### 3.7 Motion Transparency

Motion transparency is said to occur when transparency is perceived purely as a result of motion

cues. A typical display consists of two fields of superimposed random dots moving in different directions. Then one field of dots appears closer than the other. The motion dissimilarity between the two fields is alone responsible for their depth segregation (Figure 20A).

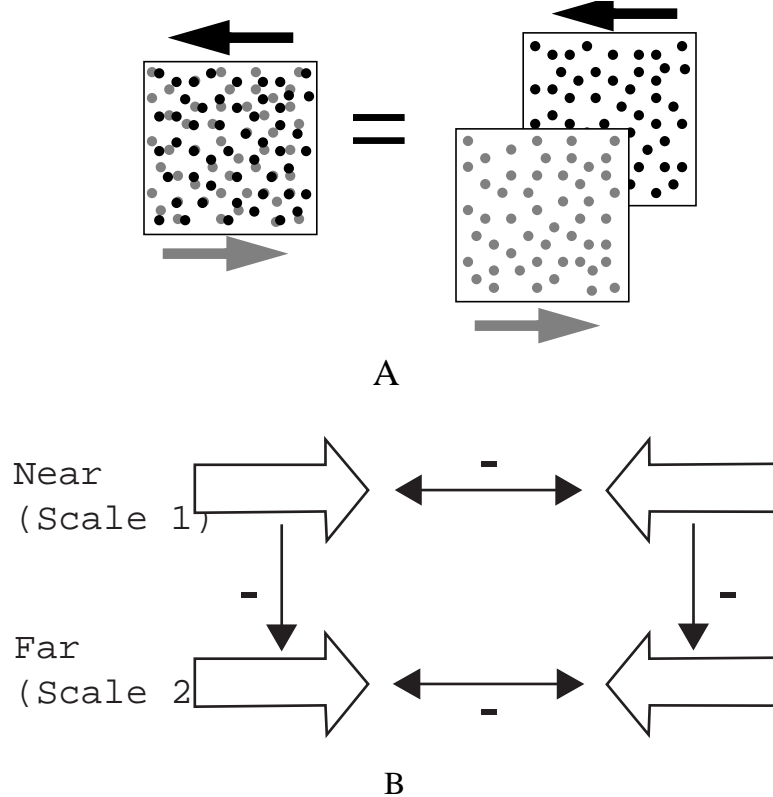


FIGURE 20. (A) Motion transparency. Note that, in this figure, shading has been used solely to identify the two fields. In the actual display, the two fields are identical in all respects except their motion. (B) Opposite motion directions within multiple scales compete. In addition, directions within scales that represent nearer motions inhibit the same directions within scales that represent farther motions. This type of “asymmetry between near and far” is also found in FACADE theory.

Opponent-direction inhibition in MT can have the undesirable effect of suppressing neuron responses under transparent conditions and rendering the visual system blind to transparent motion. Snowden *et al.* (1991) showed that the response of an MT cell to the motion of random dots in the cell’s preferred direction is strongly reduced when a second, transparent dot pattern moves in the opposite direction. Recanzone, Wurtz, and Schwartz (1997) demonstrated that this result extended to cells in MST and can also be observed when discrete objects are substituted for whole-field motions. However, Bradley, Qian, and Andersen (1995) and Qian and Andersen (1994) showed that, since opponent direction inhibition occurs mainly between motion signals with similar disparities, the disparity-selectivity of MT neurons can be used effectively to extract information about transparency due to motion cues. Our model explains how the use of multiple spatial scales, with each scale sensitive to a particular range of depths according to the size-disparity correlation, achieves this functionality.

Just as the FACADE model uses multiple scales for depth sensitivity and the Motion BCS uses multiple scales for speed sensitivity, the Formotion BCS model uses multiple scales for motion segmentation in depth. The transparent motion percept is bistable and attention can determine

which of the two fields is seen in front of the other. Fluctuations within the system, whether due to small activation asymmetries or attentional biases, can break the symmetry and render one direction of motion momentarily more salient. The model implements this by attentional enhancement via MST of a randomly selected motion direction, say rightward motion, within a given scale, say scale 1, and inside a foveal region. Even a small advantage across direction can yield selection of the preferred direction through the cooperative-competitive interactions within and between model MST and MT that carry out motion capture. Attentional enhancement acts as a gain control mechanism that adds a DC value to all cells tuned to rightward motion within the attentional locus. Consistent with recent data about attentional enhancement in MT/MST (O’Craven *et al.*, 1997; Treue and Martinez Trujillo, 1999; Treue and Maunsell, 1996, 1999), the enhancement does not change the cell tuning curves and only increases their activity.

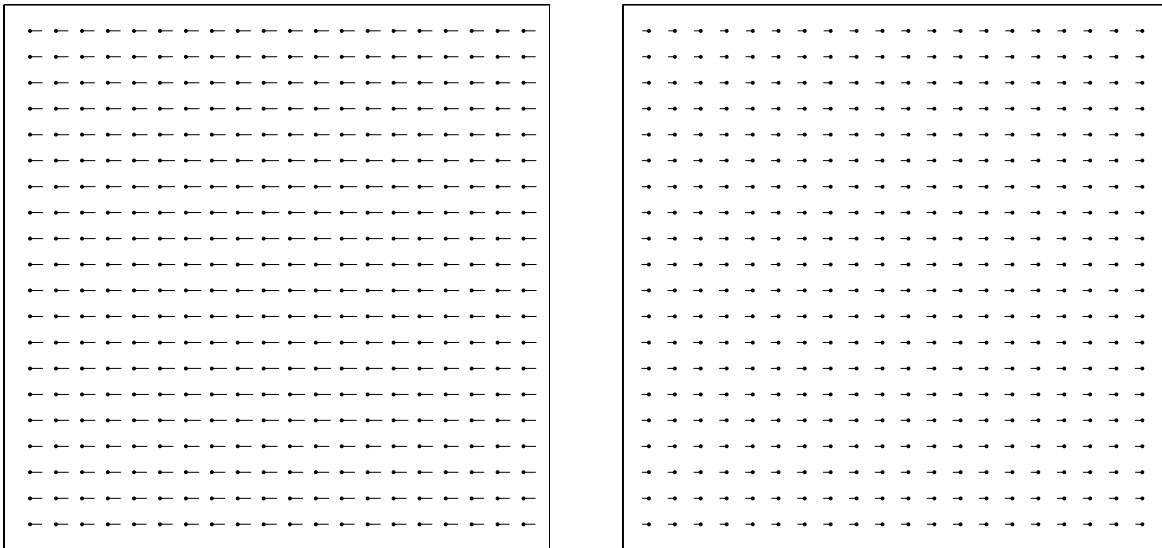


FIGURE 21. Model MST output for motion transparency. (A) Scale 1. (B) Scale 2.

The attentional gain is applied only within the selected direction and scale and inside the attentional locus. In our simulation, the locus of attention is at the center of the display and covers 6.25% of the total display area. The boost to rightward motion signals in scale 1 allows this direction to win the interdirectional competition across all of scale 1 via motion capture. Interscale inhibition from the near scale, scale 1, to the far scale, scale 2, within direction and at each spatial location suppresses rightward motion in scale 2 (Figure 20A). This is an example of the *asymmetry between near and far* (Grossberg, 1997; Grossberg and McLoughlin, 1997). Leftward motion signals in scale 2 are disinhibited and win the interdirectional competition in this scale. Two different motion directions become active at two different depths, as shown in the computer simulation in Figure 21. Thus, by using two scales representing different depths, the model explains how a 2D input sequence can lead to the perceptual segregation in depth of two surfaces based solely on motion cues. These competing directions can alternate for which appears nearer in time due to the action of habituating, or depressing, transmitters in their active pathways (*cf.*, Francis and Grossberg, 1996a; Grossberg, 1987b).

### 3.8 Chopsticks Illusion: Coherent and Incoherent Plaids

In the chopsticks illusion (Anstis, 1990), two overlapping lines of the same luminance move in opposite directions. When the lines are viewed behind visible occluders, they appear to move

together as a welded unit in the downward direction (Figures 22A and 22B). When the occluders are made invisible, the lines no longer cohere but appear to slide one on top of the other (Figures 22C and 22D). The first case is similar to coherently moving plaids while the second resembles the percept of incoherently moving plaids. Chey, Grossberg, and Mingolla (1997) simulated a variety of data concerning the conditions under which type 1 and type 2 plaids may cohere or not, including the effect of varying their component angles (Kim and Wilson, 1993), durations (Yo and Wilson, 1992), and contrasts (Stone, Watson, and Mulligan, 1990). This analysis did not consider intrinsic and extrinsic terminators, or how one component moving in front of another component could be explained. The chopsticks display provides an excellent example of how these additional factors influence perception. It contains two kinds of feature: the line terminators of each line and the intersection of the two lines. Of the line terminators, two move leftward while the other two move rightward. The line intersection moves downward. All these features have unambiguous motion signals. The model of Yo and Wilson (1992) and Wilson, Ferrera, and Yo (1992) analysed data about plaid percepts by invoking distinct channels for processing Fourier and non-Fourier signals, along with a delay in the non-Fourier motion pathway. These hypotheses are not needed in the present model. The data of Bowns (1996) do not support Fourier and non-Fourier pathways, but do support the feature tracking explanation that we further develop herein.

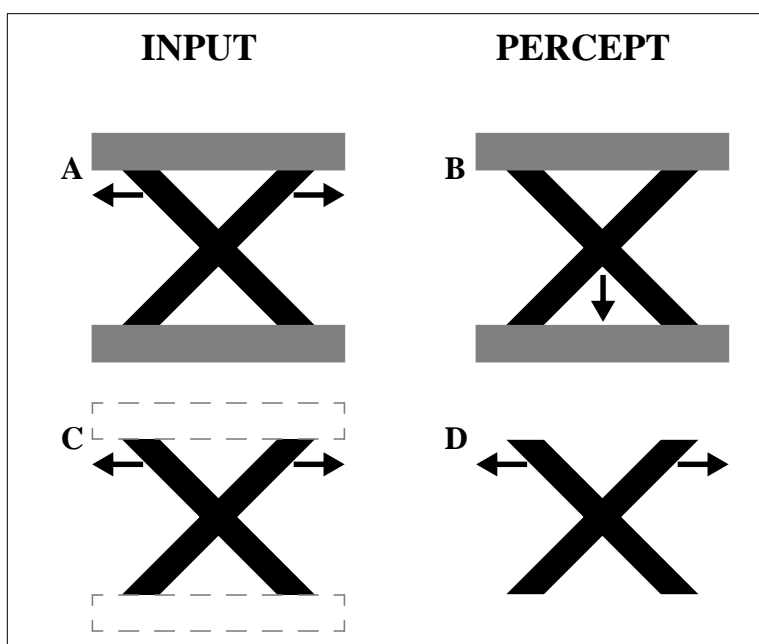


FIGURE 22. Chopsticks illusion. (A,B) Visible occluders. Two overlapping lines move in opposite directions behind visible occluders. Observers see a rigid cross translating downward. (C,D) Invisible occluders. Gray dashed lines depict the edges of the invisible occluders that define the edges of the apertures. Observers see two lines slide past each other.

### 3.8.1 Visible Occluders

When the line terminators are made extrinsic by making the occluding bars visible, their motion signals are given less importance by the visual system. The feature tracking signals due to the intersection of the two lines are stronger than those due to the extrinsic line terminators. The downward moving signals at the intersection win the competition in the MT-MST loop and propagate outward to capture the motion of the lines. Both lines appear to move downward as a single

coherent unit, as shown in the simulation in Figure 23A.

### 3.8.2 Invisible Occluders

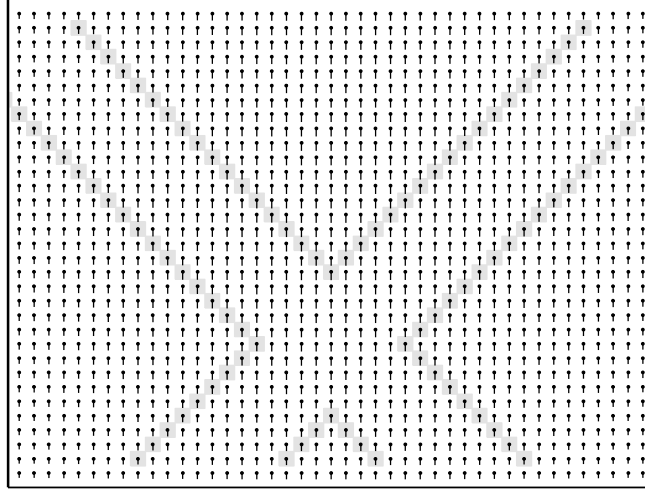
The percept of incoherency involves the interplay of more complicated mechanisms. We argue that this percept cannot be explained by considering the motion system alone, but requires a form-motion interaction of the form and motion systems; see Figure 1. In this view, incoherency is the combination of two percepts that occur simultaneously: (a) the perceived inconsistency of the motion velocities of the two lines, and (b) perceptual form transparency with one line perceived as being superimposed in front of the other. The two percepts are interlinked and can each cause the other. For instance, Stoner, Albright, and Ramachandran (1990) showed that form transparency cues at the intersections of two plaids can lead to perceptual incoherency of the plaids. This is an example of a form-to-motion interaction. However, Lindsey and Todd (1996) argued that form transparency cues are not sufficient to perceive motion incoherency. They showed that incoherency may arise from prolonged viewing, and suggested that motion adaptation may also play a role. How such adaptation could explain the Lindsey and Todd (1996) data was described and simulated in Chey, Grossberg, and Mingolla (1997), but without a simulation of incoherent motions at different depths. In the chopsticks illusion, there are no form cues that robustly lead to perceptual transparency at each moment. Motion cues lead to the percept of depth segregation of the two lines. This is a motion-to-form interaction. Models that have simulated incoherent plaids without a form-to-motion interaction (Chey, Grossberg, and Mingolla, 1997; Liden and Pack, 1999) have not produced the perceived motion at plaid intersections.

In the chopsticks illusion, when the line terminators are intrinsic (Figure 22C), their motion signals are at least as strong as those due to the line intersection. The different motion signals arising from line terminators leads to the depth segregation of the two lines (Figure 22D). When this happens, the feature arising from the intersection of the two lines no longer perceptually exists, since the lines are processed at different depth planes. This is consistent with the data of Bressan, Ganis, and Vallortigara (1993) and Vallortigara and Bressan (1991). To understand how the visual system sees this stimulus, it is necessary to consider our model as part of a broader framework of models that perform figure-ground segmentation within the form system and implement both form-to-motion and motion-to-form interactions.

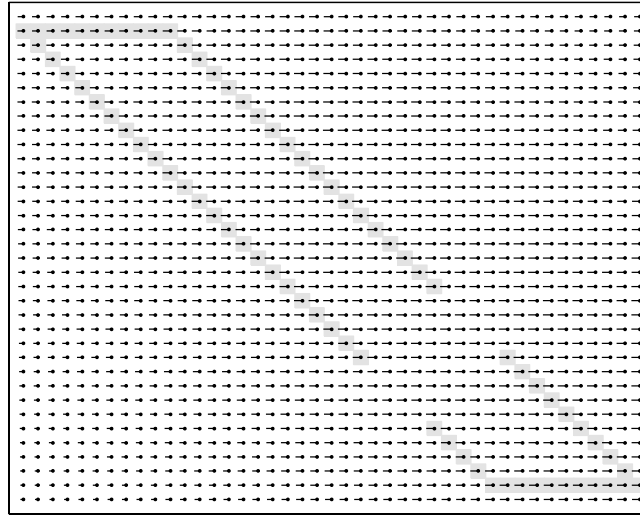
Figure 1 shows the neural pathways and connections that we predict to be involved in providing a complete explanation of the incoherent chopsticks illusion. A complete simulation of this circuit is beyond the scope of the present article, since it would involve simulating the entire figure-ground separation apparatus of FACADE theory and the Formotion BCS, augmented by top-down connections from model area MT to V1. A qualitative explanation can be given, based upon extensive simulations of FACADE (Grossberg and McLoughlin, 1997; Grossberg and Pessoa, 1998, Kelly and Grossberg, 2001), formotion interactions (Baloch and Grossberg, 1997; Francis and Grossberg, 1996b), and top-down connections to V1 (e.g., Grossberg and Raizada, 2000; Raizada and Grossberg, 2000). This qualitative explanation proceeds as follows:

The input motion sequence appears at V1 after retinal and LGN processing. Figure-ground processing between V1 and V2 by FACADE mechanisms detects occlusion events in the form of T-junctions and assigns a depth ordering to object boundaries at the site of an occlusion. This stage, labelled as 1 in Figure 1, represents one source of inputs to the Formotion BCS model; see Level 1 in Figure 4. Form-to-motion signals from V2 to MT enables the motion stream to respond to the figure-ground separated form signals, as indicated by the simulations described above. In particular, the motion system can compute feature tracking signals at the intrinsic line terminators of the

A



B



C

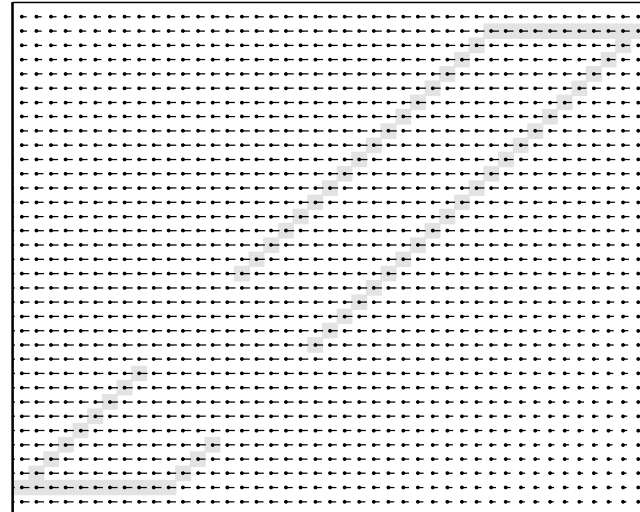


FIGURE 23. Model MST output for the chopsticks illusion. (A) visible occluders. (B) invisible occluders: scale 1. (C) invisible occluders: scale 2. The simulations for panels (a) - (c) were done on a 57 X 33 grid of locations. The leftmost 7 and rightmost 7 columns of the grid were cropped for figure display of (a); the rightmost 14 and leftmost 14 columns of the grids of panels (b) and (c), respectively, were also cropped.

chopsticks, as well as at their intersection. This stage is labelled as 2 in Figure 1.

The grouping and priming MT-MST loop, labelled as 3 in Figure 1, corresponds to Level 5 of the Formotion BCS model. This process detects the lack of a clear directional winner due to the conflicting motion signals from the line terminators. In the MT-MST feedback loop, these conflicting signals propagate from the line terminators to the intersection. At any point of one of the chopsticks, including their intersection, it is assumed that top-down attention in MST randomly or volitionally enhances one of the two chopsticks. As noted in our simulation of motion transparency, even a small asymmetry in activation, whether due to attention or some other internal or external fluctuation, is sufficient to break such a deadlock. For definiteness, let us assume that an attentional fluctuation is the cause. Then attentional enhancement of the motion signals can propagate along the form boundaries of the attended chopstick, just like feature tracking signals do. This top-down attentional priming effect from MST to MT can then propagate to V1 via top-down MT-to-V1 signals, labelled 4 in Figure 1.

The motion-to-form interaction from MT-to-V1 along pathway 4 in Figure 1 is predicted to act like a top-down ART-like attentional prime (Grossberg, 1999a). This proposal is supported by neurophysiological data showing that feedback connections from MT-to-V1 help to differentiate figure from ground (Hupé *et al.*, 1998). Feedback facilitates V1 responses to moving objects in the center and inhibits responses in the surround, as also occurs in the model. Attention amplifies the boundaries formed at the attended chopstick, much as increasing the contrast of that chopstick would do.

Such an activity difference in processing two overlapping figures, in which one figure partially occludes another, is known to cause figure-ground separation (Bregman, 1981; Kanizsa, 1979). FACADE theory explains how such an activity difference can activate figure-ground separation of the boundaries corresponding to the two chopsticks, through V1-V2 interactions (Grossberg, 1997). The boundaries of the two chopsticks are then processed on two different depth planes within the form system. The theory explains how the boundaries of the favored chopstick are processed on the nearer depth plane, leading to a visible, or modal, percept of the occluding chopstick. FACADE also explains how the form system amodally completes the boundaries of the "far" chopstick behind the occluding chopstick. Once the boundaries are separated, they can drive motion processing on different depth planes in MT via a V1-V2-MT interaction. The attentional bias hereby propagates in an MST-MT-V1-V2-MT loop. Once figure-ground separation is initiated, another pass through the model MT-MST interactions, using the separated chopsticks and their motion signals as inputs, can determine the perceived motion directions of the lines at each depth. This second loop is simulated in Figures 23B and 23C, which shows a percept of horizontal incoherent motion of the two chopsticks on two depth planes.

### 3.9 Illusory Contours from Translating Terminators

A related type of experiment can also benefit from a full simulation of the entire formotion system outlined in Figure 1. In the ingenious experiments of Gurnsey and von Grünau (1997), arrays of aligned terminators moving in the direction of their orientation could give rise to either a percept of veridical motion in the real direction of terminator motion, or to a percept of motion in the direction perpendicular to the illusory contours that are formed at the ends of the terminators. Veridical motion was more easily seen when terminators (1) were created in low-frequency carriers, (2) terminated short lines, and (3) moved slowly. In the complementary high-frequency, long line, and fast movement conditions, illusory contour motion was seen. Part of these results can be explained by mechanisms whereby real and illusory boundaries are created in the form processing



stream. In this regard, Gurnsey and von Grünau (1997) cite and build upon the articles by Grossberg and Mingolla (1985) and Grossberg (1987) that introduced the type of "rectified double-filter" model from which many later boundary and texture filter models of other authors grew, and which formed the foundation for the 3D boundary mechanisms of FACADE theory. The rectified double-filter model is not sufficient to explain how illusory contours are formed in response to sparse inducers, but the strength of its output signals do tend to covary with the strength of the illusory contours that may be generated by them, other things being equal.

Properties (1) and (2) are consistent with the hypothesis that increasing the density and length of inducers can strengthen the illusory contours, and thus the probability of perceiving motion perpendicular to the orientation of the illusory contours, other things being equal. The fact that increasing the density and length of inducers can strengthen illusory contours is familiar from studies of stationary illusory contours (e.g., Leshner and Mingolla, 1993; Shipley and Kellman, 1992; Soriano, Spillman, and Bach, 1996) and has been simulated by the FACADE model (Grossberg, Mingolla, and Ross, 1997; Ross, Grossberg, and Mingolla, 2000). With regard to property (3), Gurnsey and von Grünau (1997) note that, on the assumption that "the spatial offset between two filters is proportional to their sizes, then it is natural that [they] should be tuned to faster speeds" (p. 1021). This sort of property is a basic assumption of the Motion BCS (Chey, Grossberg, and Mingolla, 1997, 1998), which shows that a larger response threshold within larger short-range filters (see Figure 4 and Section 2.3) helps to make them speed-sensitive. As a result, larger scales selectively respond to higher speeds. Thus the combination of properties (1)-(3) may be linked to known properties of FACADE illusory contour formation, formation inputs of real and illusory contour signals to the motion system, and known speed-sensitive properties of the Motion BCS.

The rectified double-filter model is insufficient in another way too. Gurnsey and von Grünau (1997) note that, in two conditions called the 75% White and 25% White conditions, when illusory contour motion determines the percept, the illusory contours appear to form part of a 3D occluding surface that moves over a stationary background. This is perceived whether the occluding surface or the background is defined by the array of lines. The double-filter model cannot explain this result. FACADE theory shows how the strongest boundaries form bounding contours of occluding surfaces, and the rest of the scene is perceived at a slightly farther depth.

Gurnsey and von Grünau (1997) also studied how two arrays of line terminators, with different orientations and moving in different directions, could give rise to the percept of either coherent plaid motion or incoherent component motion. When the two illusory contours were aligned, subjects almost always reported seeing coherent downward motion. As the phase shift between the two illusory contours increased, there was a decrease in the tendency to see coherent motion. The authors note that "this result suggests that the responses are combined so that spatially coincident responses increase the salience of the translating contour" (p. 1023). The authors speculate that the responses to both filters should be combined to yield the desired result and that these responses help to extract occlusion boundaries. In FACADE, the strength of real or illusory contours increases with the cumulative strength of their inducers, a property called *analog coherence* (Grossberg, 1999a), and the strongest boundaries initiate a figure-ground process that tends to make them boundaries of occluding figures.

### 3.10 Adapting Coherent and Incoherent Plaid Motions

Related data can also be qualitatively explained by the Formotion BCS. Von Grünau and Dubé (1993) studied how adaptation to plaids which are seen to be coherent can reduce the time that

coherence is seen relative to incoherent component motion, and conversely. They also showed that adaptation to motion direction per se is not sufficient to explain these results, because adapting to, say, a horizontal component grating moving downwards does not fully adapt the coherent downward plaid motion percept that is derived from two component motions. They state that "the underlying processes are adapted independently" (p. 199) even though the data show a significant amount of adaptation (their Figure 4), but one that is less than complete. Chey, Grossberg, and Mingolla (1997) simulated how adaptation could clarify plaid coherence data showing that greater adaptation is needed to produce incoherent motion for smaller differences between the component orientations. The adaptation in these simulations was proposed to take place from cortical area MT to MST; that is, as part of the motion grouping process. Even with only this adaptation site, incomplete adaptation might occur in the von Grünau and Dubé (1993) experiments if only because the perceived speed of the horizontal motion and of the coherent plaid motion may be different, and would therefore adapt different speed-sensitive MT-to-MST connections. Beyond this precaution, there is also the fact that adaptive sites may exist at multiple levels in the form and motion systems, and have already played a crucial role in simulations of other form, formotion, and motion data; e.g., Baloch and Grossberg (1997), Baloch, Grossberg, Mingolla, and Nogueira (1999), Francis and Grossberg (1996), Grossberg (1987b). As soon as any site prior to the MT-to-MST pathway is made adaptive, incomplete adaptation would prevail, because the directions of the plaid components would not adapt the coherent plaid direction in these pathways.

## 4. Discussion

The Formotion BCS model successfully performs the conflicting tasks of integration and segmentation of motion cues into a unified global percept. Interconnections between neurons in the model (Figure 1) are consistent with, and functionally clarify, currently known data on the connectivity between cortical areas devoted to visual motion processing such as the retina, V1, V2, MT, and MST. The model extracts feature tracking signals from a 2D motion sequence without explicit feature detection or feature matching. The model combines unambiguous motion signals from features with ambiguous signals that arise from the aperture problem. The two types of signals are computed by the same mechanisms. Competition between motion signals from feature tracking regions and other parts of the scene determines the final 3D percept. Simulations show how a range of challenging percepts can be explained by a single model.

### 4.1 The Motion Boundary Contour System

The Motion Boundary Contour System (BCS), which has been further developed in this paper as a Formotion BCS model, was introduced by Grossberg and Rudd (1989, 1992), who simulated data on short-range and long-range apparent motion, including beta, gamma and reverse-contrast gamma, delta, reverse, split, and Ternus and reverse-contrast Ternus motion. Grossberg (1991, 1998) extended this model to explain how a moving target can be tracked when it is intermittently occluded by intervening objects. Grossberg and Mingolla (1993) further extended the model to suggest a solution to the global aperture problem.

Baloch and Grossberg (1997) and Francis and Grossberg (1996) integrated this version of the Motion BCS model with FACADE boundary-formation mechanisms to explain data which depend upon interaction of the form and motion systems. This was the first Formotion BCS model, and it was used to explain and simulate the classical Korte's laws, as well as the line motion illusion, motion induction and transformational apparent motion. This version of the

model did not, however, simulate feature tracking signals or the aperture problem.

To overcome these gaps, Chey, Grossberg, and Mingolla (1998) elaborated the role of transient cells beyond the Grossberg-Rudd model, and added multi-scale dynamics to the model to explain the size-speed correlation and to simulate data on how visual speed perception and discrimination are affected by stimulus contrast, duration, dot density and spatial frequency. Chey, Grossberg, and Mingolla (1997) extended this model to stimulate data about motion integration, notably conditions under which components of moving stimuli cohere into a global direction of motion, as in barberpole and Type 1 and Type 2 plaids. This model also simulated the temporal dynamics of how unambiguous feature tracking signals from line terminators spread to and capture ambiguous signals from line interiors. Baloch *et al.* (1999) showed how adding interactions between ON and OFF cells could simulate both first-order and second-order motion stimuli, including the reversal of perceived motion direction with distance from the stimulus (gamma display), and data about directional judgments as a function of relative spatial phase or spatial and temporal frequency.

This paper extends the model further to perform motion integration as well as motion segmentation by combining figure-ground mechanisms (areas V1 and V2) and formotion interactions (from V2 to MT) with motion mechanisms (areas V1, MT, and MST). Together these mechanisms can distinguish intrinsic vs. extrinsic terminators, and show how feature tracking signals and ambiguous aperture motion signals can influence each other by propagating across space.

It is reasonable to ask whether the Formotion BCS model, in its present form, can simulate all of the data which previous versions of the model have already simulated with a single set of parameters. Such a re-simulation would be an enormous undertaking, which is perhaps best carried out only after the model achieves its final form. One can, however, assert with some confidence that the model can simulate all of these data, for the following reasons. The formotion inputs to the Motion BCS via V2-to-MT connections do not change the mechanisms and parameters with which the Motion BCS responds to motion data via its direct V1-to-MT pathway. This addition does not, therefore, impair the simulations that used the Motion BCS alone.

The Motion BCS, in turn, has been developed in an evolutionary way, such that previous mechanisms are preserved while new mechanisms are added. For example, Grossberg and Rudd (1989, 1992) emphasized the short-range and long-range filters to explain data about long-range apparent motion. Chey, Grossberg, and Mingolla (1997, 1998) refined the transient cell filter that feeds the short-range and long-range filters, but did not disrupt the key properties of these filters that explained the data targeted by Grossberg and Rudd, but also showed how these filters play an important role in amplifying feature tracking signals. Likewise, the Baloch *et al.* (1999) addition of OFF cells to the transient cell filter did not destroy its earlier properties. Taken together, this family of Motion BCS and Formotion BCS models explains an unrivaled set of neural and psychophysical data about motion perception. Additional neurophysiological data that support the model and comparisons with alternative motion models are summarized below.

## **4.2 Neurophysiological evidence**

### **4.2.1 Level 2: Transient Cells**

Directionally sensitive cells, similar to those in Level 2 of the model, have been found both in the retina of rabbit (Barlow, Hill, and Levick, 1964) and in simple and complex cells in V1 (Hubel and Wiesel, 1968), as well as in later stages in the visual processing stream. Barlow and Levick

(1965) first suggested that directional sensitivity in ganglion cells of the rabbit retina is mainly a result of the lateral spread of inhibition in an asymmetric fashion, so that it blocks excitation which subsequently arrives on one side of it, but not on the other. This forward inhibition has a certain rise time and decay and serves to veto cell responses to the null direction. This approach argues against the Reichardt (1961) hypothesis that directional selectivity is achieved by the cross-correlation of a signal with delayed excitation from one side.

The Barlow and Levick (1965) proposal has received considerable support. Pharmacological studies of the retinae and primary visual areas of rabbits, cats and monkeys (Ariel and Daw, 1982; Sato, Katsuyama, Tamura, Hata, and Tsumoto, 1995; Sillito, 1975, 1977; Wyatt and Daw, 1976) conclude that antagonists to the inhibitory neurotransmitter gamma-aminobutyric acid (GABA) abolish or greatly reduce directional selectivity. Ariel and Daw (1982) observed that a potentiator of the excitatory neurotransmitter acetylcholine (ACh) leads to excitation which overcomes or outlasts the null direction GABA inhibition. The spatial extent of GABA inhibition is asymmetric to and larger than the spatial extent of ACh excitation.

Other physiological studies (Emerson, Citron, Vaughn, and Klein, 1987; Emerson and Coleman, 1981; Emerson and Gerstein, 1977; Ganz, 1984; Ganz and Felder, 1984) compared responses to single static flashes at various receptive field locations in either the preferred or the null direction with responses to sequence pairs of static flashes at those same locations. They found that the response to a single bar was smaller when it was preceded by a stimulus from the null side. Hammond and Kim (1994) and Innocenti and Fiore (1974) mapped excitatory and suppressive receptive fields and found that their profiles were spatially offset, especially along the preferred direction such that, for stimuli moving in the non-preferred direction, the inhibition lay ahead of the excitation. Ganz and Felder (1984), Goodwin, Henry, and Bishop (1975a, 1975b) and Heggelund (1984) argued against Hubel and Wiesel's (1959, 1962) hypothesis that directional selectivity can be explained on the basis of a linear combination of responses from adjacent ON and OFF regions of the neuron. Several of these neurophysiological studies (Barlow and Levick, 1965; Emerson, Citron, Vaughn, and Klein, 1987; Emerson and Gerstein, 1977; Ganz, 1984; Ganz and Felder, 1984) agree about the existence of direction-selective subunits distributed across the receptive field and contributing their inputs to a directionally selective neuron.

However, another theory for directional selectivity exists (Dean and Tolhurst, 1986; DeAngelis, Ohzawa, and Freeman, 1993a, 1993b; Jagadeesh, Wheat, and Ferster, 1993; Jagadeesh, Wheat, Kontsevich, Tyler, and Ferster, 1997; McLean and Palmer, 1989; McLean, Raab, and Palmer, 1994; Movshon, Thompson, and Tolhurst, 1978; Reid, Soodak, and Shapley, 1987, 1991). This is referred to as *spatiotemporal inseparability* (Adelson and Bergen, 1985). According to this hypothesis, differences in excitatory response timing across the receptive field causes directional sensitivity. A stimulus moving in the preferred direction would activate faster and faster responses which summate optimally if the stimulus speed matches the shift in response time course. In a recent study on alert fixating macaque monkeys, Livingstone (1998) suggested that delayed asymmetric inhibition may contribute to the shifting excitatory response time course. Her data suggest that asymmetric forward inhibition is the major determinant for directionality in V1 cells. She shows how the morphology and connectivity of Meynert cells, that are large, direction-selective, MT-projecting cells in layer 6 of V1, can be used to explain the role of inhibition in direction-selectivity. A Meynert cell has asymmetrical basal dendrites extending in one direction within layer 6. It receives excitatory inputs from its distal dendrites and relatively denser inhibitory inputs from the synapses formed by inhibitory interneurons with its cell body. This structure ensures that the cell receives excitatory and inhibitory inputs from different regions of the visual field. Besides, due to dendritic conduction delays, excitatory inputs from distal dendritic tips

would arrive at the cell body later than the inhibitory inputs from interneurons. These simple properties enable the cell to use asymmetric inhibition to achieve directional selectivity.

#### **4.2.2 Level 4: Spatial Competition and Opponent Direction Inhibition**

Several neurophysiological studies confirm that the opponent direction inhibition used in Level 4 of the model exists in MT but has not been found in V1 (Bradley, Qian, and Andersen, 1995; Heeger, Boynton, Demb, Seidemann, and Newsome, 1999; Qian and Andersen, 1994; Recanzone, Wurtz, and Schwarz, 1997; Snowden, Treue, Erickson, and Andersen, 1991).

#### **4.2.3 Level 5: Long-range Directional Grouping and Attentional Priming**

Several studies show that MT cells are directionally selective (Albright, 1984; Maunsell and van Essen, 1983a; Zeki, 1974a, 1974b). They respond more strongly to moving stimuli, irrespective of direction of contrast, than to static stimuli. Psychophysical evidence using heterogeneous-cue plaids (Stoner and Albright, 1992) shows that motion signals are integrated irrespective of whether they were produced by first-order or second-order form cues. The discovery of two types of MT neuron, those that respond to component motion and those that respond to pattern motion of plaids (Movshon, Adelson, Gizzi, and Newsome, 1985; Rodman and Albright, 1989) supports the hypothesis that MT is the first cortical area in the visual processing stream where motion integration cues occurs.

Outputs from MT feed into MST (Desimone and Ungerleider, 1986; Maunsell and van Essen, 1983b). MST cells are directionally selective and have large receptive fields. The dorsal part of MST, MSTd, responds selectively to expansion, contraction, and clockwise or counterclockwise rotation (Saito *et al.*, 1986) and favors movements of a wide textured field like those caused by observer movements over those of moving objects (Duffy and Wurtz, 1991a, 1991b; Komatsu and Wurtz, 1988; Orban *et al.*, 1992; Tanaka and Saito, 1989). Grossberg, Mingolla, and Pack (1999) modeled how MSTd may control visually-based navigation using optic flow stimuli. The ventral part of MST, MSTv, prefers object movements to whole-field movements. This is the sort of motion processing that we have used in our model of MT-MST directional selection and attentional priming. Pack, Grossberg, and Mingolla (2001) have shown how MSTv cells can represent predicted target speed during smooth pursuit tracking.

Treue and Maunsell (1996, 1999) demonstrated a strong modulatory influence of attention on motion processing in the directionally selective cells of MT and MST in macaque monkeys. Using fMRI on humans subjects, O'Craven *et al.* (1997) found greater activation in MT/MST in the presence of voluntary attention. Further, attention acts as a nonspecific gain control mechanism that enhances responses within the locus of attention without narrowing direction-tuning curves (Treue and Martinez Trujillo, 1999). As noted in Section 2.5, these attentional data are consistent with the predicted relationship between preattentive motion capture and directional attentional priming, but does not directly test this key prediction.

### **4.3 Comparison with other motion models**

Several theories of motion perception have been proposed in the literature. Most of these offer explanations for either motion integration or motion segmentation, but not both, and few of them describe neural mechanisms for all model stages. Although the data about motion integration and segmentation are challenging, since these processes exhibit contradictory yet complementary

goals, it is more difficult to develop a theory that can handle both types of data with the same set of mechanisms. We describe models below that have treated a subset of these data and compare them to our approach. A summary of this analysis is presented in Table 1.

The IOC model of motion integration attempts to explain the perceived motion direction of coherent plaids (Adelson and Movshon, 1982). IOC predicts that observers always see the veridical motion of a coherent plaid pattern. However, a growing body of data suggests that this is not the case (Bowns, 1996; Bressan, Ganis, and Wallortigara, 1993; Cox and Derrington, 1994; Derrington and Ukkonen, 1999; Ferrera and Wilson, 1990, 1991; Rubin and Hochstein, 1993; Vallortigara and Bressan, 1991). Features such as dots, line terminators, object corners and plaid intersections can determine the global direction of motion in both plaid displays (Alais, Burke, and Wenderoth, 1996; Alais, van der Smagt, Verstraten, and van de Grind, 1996; Bowns, 1996; Bressan, Ganis, and Wallortigara, 1993; Burke, Alais, and Wenderoth, 1994; Derrington and Ukkonen, 1999; Vallortigara and Bressan, 1991; Wenderoth *et al.*, 1994) and non-plaid multiple-aperture displays (Alais, van der Smagt, van der Berg, and van der Grind, 1998; Lorenceau and Shiffrar, 1992; Mingolla, Todd, and Norman, 1992).

Paper	Type of model	Type of data simulated
Adelson and Bergen (1985)	spatiotemporal energy	directional and speed sensitivity
Adelson and Movshon (1982)	intersection of constraints (IOC)	motion integration: coherent plaids
Del Viva and Morrone (1998)	feature tracking	motion integration and segmentation
Fennema and Thompson (1979)	gradient	directional and speed sensitivity
Hildreth (1984)	regularization / smoothing	motion integration
Horn and Schunck (1981)	regularization / smoothing	motion integration: optic flow
Jasinschi, Rosenfeld and Sumi (1992)	correlational and IOC	motion integration and segmentation
Jin and Srinivasan (1990)	gradient	directional and speed sensitivity
Johnston, McOwan and Benton (1999)	gradient	motion segmentation: static noise
Johnston, McOwan and Buxton (1992)	gradient	first- and second-order motion
Koch, Wang and Mathur (1989)	regularization / smoothing	motion integration
Lappin and Bell (1972)	correlational	apparent motion
Liden and Pack (1999)	feature tracking	motion integration and segmentation
Loffler and Orbach (1999)	feature tracking	motion integration: coherent plaids
Marr and Ullman (1981)	gradient	directional and speed sensitivity
Marshall (1990)	adaptive learning neural network	motion integration: barber-pole
Nowlan and Sejnowski (1994)	spatiotemporal energy	motion segmentation: transparency
Poggio, Torre and Koch (1985)	regularization / smoothing	motion integration: barber-pole
Qian, Andersen and Adelson (1994)	subtractive and divisive inhibition	motion segmentation: transparency
Reichardt (1961)	correlational	low-level vision
Sachtler and Zaidi (1995)	center-surround shearing	motion segmentation
van Santen and Sperling (1985)	correlational	directional and speed sensitivity
Wang (1997)	adaptive learning neural network	motion integration and segmentation

Paper	Type of model	Type of data simulated
Watson and Ahumada (1985)	spatiotemporal energy	directional and speed sensitivity
Yo and Wilson (1992)	Fourier and non-Fourier channels	motion integration
Yuille and Grzywacz (1988)	regularization / smoothing	motion integration: motion capture
Zemel and Sejnowski (1994)	adaptive learning neural network	motion segmentation

TABLE 1. Comparison of previously presented motion models.

Given that the motion signals from features plays an important role, we are still faced with the problem of how to compute this motion. Correlational models (Lappin and Bell, 1972; Reichardt, 1961; van Santen and Sperling, 1985) suggest that this is done by a pair of receptors separated by some physical distance such that the delayed output of one receptor is multiplied by the output of the other receptor. This matching of corresponding points in succeeding frames can be done at two levels. Feature matching models (Reichardt, 1961; van Santen and Sperling, 1985) detect salient features and match corresponding features to compute image velocities. Global matching models (Lappin and Bell, 1972) perform template matches over larger regions of space by sliding images in subsequent frames to obtain optimal matches. Both kinds of correlational model are susceptible to the *correspondence problem*; namely, how to establish correspondences across successive frames, especially when the similarity of objects in the images suggests that more than one kind of correspondence is possible (Anstis, 1980). Clearly, velocity estimates in the scene depend crucially on which correspondence is chosen. We therefore need a method of computing the motion of features without explicitly detecting and matching features.

Spatiotemporal motion energy models (Adelson and Bergen, 1985; Watson and Ahumada, 1985) are similar to correlational models in that they recover speed and direction estimates from spatiotemporal information in the scene. To do this, they use linear filters whose Fourier transforms are oriented in space-time. Velocity sensitivity is achieved through orientation sensitivity in space-time. Motion energy models are formally equivalent to elaborated Reichardt detectors in that they compute identical outputs for any given input (van Santen and Sperling, 1985). Emerson, Bergen, and Adelson (1992) presented neurophysiological evidence that the responses of directionally selective complex cells in the cat’s striate cortex are consistent neither with correlational models (Reichardt, 1961; van Santen and Sperling, 1985) nor with an opponent combination of motion energy models (Adelson and Bergen, 1985; Watson and Ahumada, 1985).

Gradient models (Fennema and Thompson, 1979; Jin and Srinivasan, 1990; Marr and Ullman, 1981) compute velocity by using local spatial and temporal derivatives of the image’s spatiotemporal luminance profile. Speed sensitivity is coded by the magnitudes of the gradients. Since derivatives are computed at single spatial locations, gradient schemes successfully bypass the correspondence problem. However, they succumb to the aperture problem since the expression used to compute velocity in the case of moving 1D bars is ill-conditioned. In an attempt to solve this problem, Johnston and colleagues (Johnston and Clifford, 1995; Johnston, McOwan, and Benton, 1999; Johnston, McOwan, and Buxton, 1992) proposed a model that combines a gradient scheme with the IOC procedure to detect first-order and second-order motion in the presence or absence of static noise. The resulting multi-channel gradient model can detect the motion of a grating superimposed on a static random binary noise pattern. The model is consistent with the data of Lu and Sperling (1995) whose experiments using contrast-modulated noise patterns found no evidence for feature tracking in first-order and second-order motion detection. However, when contrast-modulated sine-wave gratings are substituted for contrast-modulated noise patterns, second-order motion detection is disrupted by the superimposition of a pedestal, thus suggesting that the

motion of contrast envelopes is detected by a mechanism that tracks features (Derrington and Ukkonen, 1999). Although the multi-channel gradient model is well-conditioned for velocity coding, it fails in the same way as IOC in explaining data on Type 2 plaids. The Motion BCS model of Baloch *et al.* (1999), which is consistent with the Formotion BCS model, explains such first-order and second-order motion percepts within the present modeling framework.

Regularization theories (Hildreth, 1984; Horn and Schunck, 1981; Koch, Wang and Mathur, 1989; Poggio, Torre, and Koch, 1985; Yuille and Grzywacz, 1988) minimize a cost function by applying a smoothness constraint to the velocity field. They make the assumption that real-world objects have smooth surfaces, whose projected velocity field is usually smooth. Such techniques are robust to noise and are good for motion integration, but can perform motion segmentation only by explicitly detecting discontinuities in the motion field, such as when the spatial gradient of the velocity field between two neighboring points is larger than some threshold. Further, the iterative minimization of the cost functional is computationally expensive, subject to getting trapped in local minima for non-quadratic functionals, and difficult to interpret biologically.

Marshall (1990) and Wang (1997) presented adaptive neural networks in which weights and connections between neurons are modified during an iterative training phase in which motions of various directions and speeds are presented. However, it remains to be seen whether the perception of motion illusions such as those presented in this paper is the result of adaptive learning.

Other models primarily address the problem of motion segmentation (Nowlan and Sejnowski, 1994; Qian, Andersen, and Adelson, 1994; Sachtler and Zaidi, 1995; Zemel and Sejnowski, 1994). They detect local motion discontinuities and use these to segment the scene. They fail to integrate motion signals across discontinuities that arise from noise in the stimulus.

Computational models of feature tracking have traditionally faced two problems: (1) What constitutes a feature? How should features be detected in a scene? Definitions of features have typically been vague. Dots, line terminators, object corners and plaid intersections are examples of easily detectable features. However, corners of objects formed by subjective contours can also constitute features and these are considerably harder to detect. (2) Even if features can be reliably detected in a scene, how should features in one frame of a motion sequence be matched to features in the next frame? This is the correspondence problem discussed earlier.

Jasinschi, Rosenfeld, and Sumi (1992) proposed a model that combines a feature matching scheme similar to that of correlational models with IOC to explain motion transparency and coherence. The model uses a velocity histogram that combines votes from the velocities of features such as corners and line endings (computed by template matching) with those from the intersections of all possible constraint lines due to the motion of image contours. The model succeeds in explaining motion transparency; namely, how two velocities can be perceived at the same spatial location, as well as the bistability of motion transparency and coherence in plaid displays. However, the use of global correlational matching as well as IOC makes the model susceptible to the drawbacks of both types of scheme.

Del Viva and Morrone (1998) detect features by computing peaks of spatial local energy functions and compute feature velocities using a spatiotemporal motion energy scheme. Such a technique fails to detect features formed by subjective contours. Loffler and Orbach (1999) presented a model of motion integration in coherent plaids which uses two parallel pathways (Fourier and non-Fourier) to perform feature tracking without the explicit use of feature detectors such as end-stopped cells. As noted in Section 2.5, Yo and Wilson (1992) also proposed that two such parallel



pathways exist. However, there is psychophysical evidence against the existence of two pathways (Bowns, 1996; Cox and Derrington, 1994). Moreover, none of the models described so far can explain how the intrinsic-extrinsic classification of features influences the global motion percept. For instance, intrinsic line terminators have unambiguous motion signals while the motion of extrinsic terminators is discounted by the visual system; while the former can block motion grouping across apertures, the latter fail to do so (Lorenceau and Shiffrar, 1992).

Liden and Pack (1999) proposed a neural network model of motion integration and segmentation that consists of two separate but interacting systems of cells, one specialized for integration and the other for segmentation. The model takes into account the relative strengths of intrinsic and extrinsic features by hypothesizing that local motion signals near T-junctions signalling occlusion are masked. In this way, the motion signals generated by extrinsic features are excluded from computations of global motion while those of intrinsic features are preserved. This mechanism predicts the existence of a form-to-motion interaction whereby form cues such as T-junctions inhibit motion signals at nearby locations. The nature of the interaction between the integration and segmentation networks precludes the possibility of two motion velocities being active at the same spatial location. Therefore, the model cannot explain motion transparency.

Our model suggests that a single system is capable of performing the dual tasks of motion integration and segmentation. The model performs neither feature detection nor feature matching, thus circumventing both the problems faced by most feature tracking models. Nevertheless, we can reliably compute feature tracking signals by accumulating evidence at short-range and long-range spatial filters and through the use of competitive mechanisms. For a motion signal at a given spatial location to be attributed to the motion of a feature, it is sufficient that the signal be consistent and have few competitors both across direction at the same spatial location and across space from similar directions. Model dynamics then ensure that these signals are made strong enough to dominate the final percept. Our model differs from that of Liden and Pack (1999) in that only form cues are inhibited at T-junctions, leaving motion cues intact. The use of multiple spatial scales makes it possible for distinct motion velocities to be active at the same spatial location but at different scales, thus allowing an explanation of depth segregation due to motion transparency.

#### 4.4 Model Complexity and Robustness

It is sometimes claimed that neural models of vision "contain a lot of parameters". Counting such parameters does not make a lot of sense, since even a well-known and simple neural mechanism, like an on-center off-surround network, uses several parameters. Rather, it makes sense only to count the number of *mechanisms* or *processing stages*; to assess whether removal of any stage prevents the explanation of key data; to survey experimental evidence for the neural existence of these stages; to test whether the mechanisms that realize the stages are robust within a conceptually meaningful parameter range; and to make predictions that test these properties.

In the case of the Formotion BCS model, all of these criteria were realized. In particular, the model was found to be robust within parameter ranges in which its main mechanisms had the functional effects for which they were included. For example, if the short-range filter is not big enough to amplify feature tracking signals, then motion capture will not occur. If the off-surround within the top-down MST-to-MT feedback pathway is not strong enough to inhibit ambiguous aperture signals from the long-range filter, then motion capture will not occur. And so on. Each of these mechanisms has a clear conceptual and functional interpretation. This is often not the case in purely formal models of perception, for which issues about whether one is "just" fitting data with functionally rather meaningless parameters or form factors is a very real issue.

As to predictions of the Formotion BCS model, every one of its processing stages, the mechanisms used to realize them, and its predicted role in generating motion percepts constitutes a series of predictions. Here we wish to focus on the particularly exciting prediction that the feedback interaction within MT-MST that is predicted to realize *preattentive motion capture* is the same circuit by which the brain achieves *attentive directional priming*. This prediction suggests that cooling ventral MST will prevent MT cells from exhibiting motion capture in the aperture-ambiguous interiors of moving objects. It also predicts that an attentive directional prime can reorganize the preattentive motion capture process. A third prediction derives from the fact that the top-down feedback is predicted to carry out ART matching (Carpenter and Grossberg, 1987; Grossberg, 1980, 1999b), which clarifies how directional receptive fields can develop and maintain themselves. The model predicts that pharmacological inhibition of the MT-to-MST bottom-up adaptive weights can prevent directional MST cells from developing, and inhibition of the MST-to-MT adaptive weights can destabilize learning in these bottom-up adaptive weights.

Grossberg (1999a) also predicted how top-down ART attention is realized within the laminar circuits of cortical areas from V2-to-V1, and by extension from MST-to-MT. Arguing by analogy from the V2-to-V1 situation, we predict that an attentional pathway may exist from layer 6 of ventral MST to layer 6 of MT (possibly by a multi-synaptic pathway from layer 6 of MST to layer 1 apical dendrites of layer 5 MT cells that project to layer 6 MT cells) followed by activation of a modulatory on-center off-surround network from layer 6-to-4 of MT. Thus, preattentive motion capture signals, as well as directional attentional priming signals, from MST are predicted to strongly activate layer 6 of MT, but to only modulate excitation within the on-center of layer 4 MYcells, while strongly inhibiting layer 4 cells in the off-surround. Without such a detailed neural model, such predictions would be inconceivable, and the means whereby the brain gives rise to visual behaviors would remain an impenetrable mystery.

## 5. Appendix: Model Equations

We first describe the symbols and notations used in the network equations. Each cell activity is denoted by a variable whose letter indicates the cell type. Subscripts indicate the spatial position of the cell. Superscripts indicate the directional tuning and scale of the cell. For example,  $F_{ij}^{ds}$  indicates the activity of a thresholded short-range filter cell at spatial location  $(i,j)$ , directional preference  $d$  and scale  $s$ . The notation  $[w]^+ = \max(w, 0)$  stands for half-wave rectification. Similarly,  $[w - t]^+$  denotes rectification with threshold at  $t$ . The outputs of every level of the model are rectified before being fed into the next level. The notation  $||S||$  indicates the size of the set  $S$ . Some equations involve interactions between opponent directions. We compute the direction  $D^O$  exactly opposite to the direction  $d$  as follows:

$$D^O = \left(d + \frac{ND}{2}\right) \bmod(ND) \quad (\text{A1})$$

where  $ND$  is the total number of discrete directions used in the simulation and  $\bmod$  is the modulo operator. All simulations use 8 directions, so  $ND = 8$ . The motion transparency and chopsticks simulations use 2 scales; all others use a single scale. These two simulations are different from the others in that they require interscale competition. Other than this difference, all simula-

tions used the same parameters. Only the inputs are varied between simulations.

## 5.1 Level 1: Input

The input consists of a series of static frames each of which represents a time slice of a motion sequence. As mentioned in Section 2.1, the boundary representations at the farther depth, computed by FACADE at each frame of the sequence, serve as the inputs,  $I_{ij}$ , to the Formotion BCS Model. Input dimensions for each simulation are listed in Table 2.

Simulation	Display Width (in pixels)	Display Height (in pixels)	No. of frames in the motion sequence	Other input specific parameters
Classic Barber Pole	60	30	15	No. of horizontal terminators = 4 No. of vertical terminators = 2
Motion Capture	60	30	15	No. of horizontal terminators = 4 No. of vertical terminators = 2 No. of dots = 4
Spotted Barber Pole	60	30	15	No. of horizontal terminators = 4 No. of vertical terminators = 2 No. of dots = 4
Line Capture	71	71	10	None
Triple Barber Pole	60	90	15	No. of horizontal terminators = 4 No. of vertical terminators = 6
Translating Square:				None
Visible Rectangular Occluders	33	33	15	
Invisible Rectangular Occluders	33	33	15	
Invisible Jagged Occluders	37	37	15	
Motion Transparency	20	20	15	No. of dots = 20
Chopsticks	57	35	15	None

TABLE 2. Input dimensions for all simulations.

## 5.2 Level 2: Transient cell network

Undirectional transient cell activities,  $b_{ij}$ , are computed by:

$$b_{ij} = x_{ij}z_{ij}, \quad (A2)$$

where simple cell activities,  $x_{ij}$ , perform leaky integration of their inputs as follows:

$$\frac{dx_{ij}}{dt} = 10[-x_{ij} + (2 - x_{ij})I_{ij}] \quad (A3)$$

and  $z_{ij}$  are habituated transmitter gates defined by:

$$\frac{dz_{ij}}{dt} = 0.03[1 - z_{ij} - 100x_{ij}z_{ij}]. \quad (\text{A4})$$

The constants outside the brackets in (A3) and (A4) depict the rates of change of the simple cell activities,  $x_{ij}$ , and the transmitter gates,  $z_{ij}$ , respectively. In (A3), the constant 2 represents the maximum value that the simple cell activities can reach. The term  $1 - z_{ij}$  in (A4) signifies that the transmitter gate  $z_{ij}$  can reach a maximum value of 1. The term  $-100x_{ij}z_{ij}$  in this equation says that transmitter habituates in proportion to the strength of the signal passing through the gate with 100 being the constant of proportionality. Thus,  $z_{ij}$  accumulates at a constant rate to a finite maximum value and habituates, or is inactivated, at a rate proportional to the strength  $x_{ij}$  of the signal. The unidirectional transient cell responses,  $b_{ij}$ , in (A2) are the gated signals of (A4). These cell activities correspond to the lowest layer of cells in Fig. 8.

Directional interneurons,  $c_{ij}^d$ , perform a time-average of unidirectional transient cell activities:

$$\frac{dc_{ij}^d}{dt} = -c_{ij}^d + b_{ij} - 10[c_{XY}^{D^O}]^+. \quad (\text{A5})$$

Each cell acquires a preferred direction as follows: Each cell receives excitatory input,  $b_{ij}$ , from the unidirectional transient cell at the same spatial location, and inhibitory input,  $c_{XY}^{D^O}$ , from the directional interneuron tuned to the opposite direction  $D^O$  at a location  $(X, Y)$  that is spatially offset from  $(i, j)$  by one unit along the preferred direction,  $d$ . For example, a directional interneuron tuned to leftward motion at location  $(i, j)$  receives inhibitory input from the directional interneuron one unit to its left and tuned to rightward motion (see Fig. 8). The inhibition is stronger than the excitation; cf., coefficient 10 in (A5).

The dynamics of directional transient cell activities,  $e_{ij}^d$ , are similar to those of directional interneurons. These cells receive excitatory input from unidirectional transient cells,  $b_{ij}$ , and inhibitory input from directional interneurons,  $c_{XY}^{D^O}$ :

$$\frac{de_{ij}^d}{dt} = 10(-e_{ij}^d + b_{ij} - 10[c_{XY}^{D^O}]^+). \quad (\text{A6})$$

In Equations (A5) and (A6), direction  $D^O$  is the direction opposite to direction  $d$  and is computed by (A1). The output of Level 2 is rectified before being sent to Level 3:  $E_{ij}^d = [e_{ij}^d]^+$ .

Equations (A5) and (A6) implement a vetoing mechanism through spatially asymmetric inhibition. The need for inhibitory directional interneurons is not only biologically motivated, as discussed in Sections 2.2 and 4.2.1, but is also functionally essential. A veto mechanism based solely on inhibitory connections between neighboring transient cells is insufficient because vetoed tran-

sient cells are incapable of further vetoing their neighbors. This problem is solved by introducing inhibitory interneurons that are capable of maintaining their activities independently of the transient cells that they veto. Besides, interneurons can operate over a time scale different from that of the transient cells. Vetoing can thus be performed robustly at a variety of speeds. Mutual inhibition between interneurons is necessary to construct transient cells that respond preferentially to a range of directions of motion and whose response is essentially invariant with input speed and to preserve the speed tuning of the short-range filters at higher stimulus speeds.

### 5.3 Level 3: Short-range filter network

The short-range filter cell activities,  $f_{ij}^{ds}$ , perform space- and time-averaging of directional transient cell responses. Each activity,  $f_{ij}^{ds}$ , receives excitatory input from directional transient cells tuned to the same direction and within a Gaussian receptive field,  $G_{ijXY}^{ds}$ , that is oriented along the preferred direction,  $d$ , of the cell. The scale,  $s$ , of each cell determines the size of its receptive field:

$$\frac{df_{ij}^{ds}}{dt} = 4 \left[ -f_{ij}^{ds} + \sum_{(X,Y)} E_{XY}^d G_{ijXY}^{ds} \right]. \quad (A7)$$

The Gaussian kernel,  $G_{ijXY}^{ds}$ , for upward and downward motion is:

$$G_{ijXY}^{Ds} = \exp \left[ -\frac{1}{2} \left( \left( \frac{X-i}{\sigma_{G_x}} \right)^2 + \left( \frac{Y-j}{\sigma_{G_y}^s} \right)^2 \right) \right], \quad (A8)$$

where  $D = 1$  or  $5$ ,  $\sigma_{G_x} = 0.5$  and  $\sigma_{G_y}^s = s + 0.5$ . The kernels for the other motion directions are obtained by rotating kernel (A8) and aligning it with the current motion direction. Short-range filter cell outputs,  $F_{ij}^{ds}$ , result from a self-similar threshold applied to  $f_{ij}^{ds}$ . This threshold increases linearly with filter size. Each scale is then activated by a different speed range that increases with scale size:

$$F_{ij}^{ds} = \left[ f_{ij}^{ds} - \frac{s}{4} \right]^+. \quad (A9)$$

### 5.4 Level 4: Competition network

Competition cell activities,  $h_{ij}^{ds}$ , implement spatial competition within each direction and opponent directional inhibition within each scale. Shunting gain-controls cell responses:

$$\frac{dh_{ij}^{ds}}{dt} = 20 \left[ -h_{ij}^{ds} + (1 - h_{ij}^{ds}) \left( \sum_{(X,Y)} F_{XY}^{ds} J_{ijXY}^d \right) - 10(h_{ij}^{ds} + 0.1) \left( \sum_{(X,Y)} F_{XY}^{ds} K_{ijXY}^d \right) - 50h_{ij}^{ds} F_{ij}^{D^o s} \right].$$

(A10)

Direction  $D^O$  is the direction opposite to direction  $d$ . The excitatory and inhibitory Gaussian kernels,  $J_{ijXY}^d$  and  $K_{ijXY}^d$ , for upward motion ( $d = 1$ ) are:

$$J_{ijXY}^1 = \left( \frac{1}{2\pi\sigma_{J_X}\sigma_{J_Y}} \right) \exp \left[ -\frac{1}{2} \left( \left( \frac{X-i}{\sigma_{J_X}} \right)^2 + \left( \frac{Y-j}{\sigma_{J_Y}} \right)^2 \right) \right] \quad (\text{A11})$$

and

$$K_{ijXY}^1 = \left( \frac{1}{2\pi\sigma_{K_X}\sigma_{K_Y}} \right) \exp \left[ -\frac{1}{2} \left( \left( \frac{X-i}{\sigma_{K_X}} \right)^2 + \left( \frac{Y-(j-1)}{\sigma_{K_Y}} \right)^2 \right) \right]. \quad (\text{A12})$$

The excitatory kernel,  $J_{ijXY}^1$ , is spatially anisotropic with  $\sigma_{J_X} = 0.5$  and  $\sigma_{J_Y} = 2.5$ . The inhibitory kernel,  $K_{ijXY}^1$ , is spatially isotropic with  $\sigma_{K_X} = \sigma_{K_Y} = 4$ , but it is offset from the cell's spatial location  $(i, j)$  by one unit in the direction opposite to the preferred direction of the cell; that is, by one unit in the downward direction. Thus inhibition spatially lags behind excitation along the preferred direction. As with (A8), kernels for the remaining motion directions are computed by aligning the kernels in (A11) and (A12) parallel to the desired direction. The simulations in this paper all use 8 directions. The kernels for north-east motion ( $d = 2$ ) are obtained by rotating kernels (A11) and (A12) clockwise by  $45^\circ$ . Level 4 activity is rectified before outputting to Level

$$5: H_{ij}^{ds} = [h_{ij}^{ds}]^+.$$

## 5.5 Level 5: Long-range Directional Grouping and Attentional Priming

The long-range filter summates competition cell outputs over large spatial extents:

$$N_{ij}^{ds} = \sum_{(X,Y)} (H_{XY}^{ds})^2 L_{ijXY}. \quad (\text{A13})$$

In (A13),  $L_{ijXY}$  is an isotropic Gaussian kernel centered at position  $(i, j)$  and defined by

$$L_{ijXY} = \exp \left[ -\frac{1}{2} \left( \left( \frac{X-i}{\sigma_{L_X}} \right)^2 + \left( \frac{Y-j}{\sigma_{L_Y}} \right)^2 \right) \right], \quad (\text{A14})$$

where  $\sigma_{L_X} = \sigma_{L_Y} = 20$ . Each model MT cell activity,  $m_{ij}^d$ , receives bottom-up excitation from the long-range filter and top-down inhibition from model MST cells,  $n_{ij}^{Ds}$ , tuned to all directions  $D$  other than the preferred direction  $d$  of the cell:

$$\frac{dm_{ij}^{ds}}{dt} = \left( -m_{ij}^{ds} + (1 - m_{ij}^{ds}) N_{ij}^{ds} - (1 + m_{ij}^{ds}) \sum_{D \neq d} [n_{ij}^{Ds}]^+ \right). \quad (\text{A15})$$

The output  $M_{ij}^{ds} = [m_{ij}^{ds}]^+$ .

### 5.5.1 Case I: Without Interscale Competition

Except for motion transparency and the chopsticks illusion, all simulations used only one scale without interscale competition. Here MST cells obey:

$$\frac{dn_{ij}^{ds}}{dt} = \left( -n_{ij}^{ds} + (1 - n_{ij}^{ds})M_{ij}^{ds} - 5 \sum_{D \neq d} [n_{ij}^{Ds}]^+ \right). \quad (\text{A16})$$

By (A16), each model MST cell activity,  $n_{ij}^{ds}$ , receives excitation,  $M_{ij}^{ds}$ , from model MT cells and lateral inhibition from model MST cells tuned to all directions  $D$  other than the preferred direction  $d$  of the cell. Competition between model MST cells chooses a winning direction which boosts activities in model MT cells tuned to the same direction, via Equation (A15).

### 5.5.2 Case II: With Interscale Competition: Motion Transparency and Chopsticks

The motion transparency and chopsticks simulations use two scales that compete with each other. In addition to the competition in Equation (A16), the equation for model MST cell activities,  $n_{ij}^{ds}$ , includes asymmetric inhibition from smaller to larger scales:

$$\frac{dn_{ij}^{ds}}{dt} = \left( -n_{ij}^{ds} + (1 - n_{ij}^{ds})(m_{ij}^{ds} + A_{ij}^{ds}) - 0.01 \sum_{(X,Y)} \sum_{D \neq d} Z^{Dd} [n_{ij}^{Ds}]^+ P_{ijXY} - \sum_{S < s} n_{ij}^{dS} \right). \quad (\text{A17})$$

In (A17),  $P_{ijXY}$  is an isotropic Gaussian kernel defined by

$$P_{ijXY} = \exp \left[ -\frac{1}{2} \left( \left( \frac{X-i}{\sigma_{P_X}} \right)^2 + \left( \frac{Y-j}{\sigma_{P_Y}} \right)^2 \right) \right], \quad (\text{A18})$$

where  $\sigma_{P_X} = \sigma_{P_Y} = 20$ .  $Z^{Dd}$  is a kernel that ensures that inhibition between opponent directions is greater than that between any other two directions:

$$Z^{Dd} = \begin{cases} 2 & D = D^O \\ 1 & \text{otherwise} \end{cases}. \quad (\text{A19})$$

$A_{ij}^{ds}$  is attentional enhancement that is specific to both direction and scale and directed to a given region of space. No attentional enhancement was used for the chopsticks simulation. For the motion transparency simulation, attention was directed to a particular direction, say  $D^A$ , and a specific scale, say  $S$ , within a given rectangular region of space centered at the center of the display  $(C_X, C_Y)$ , and with half-width  $R_X = 5$  and half-height  $R_Y = 5$ . Direction  $D^A$  is the direction for which the total activity in the long-range filter in the rectangular region is maximum. We

assume that attention is always allocated to the closest depth; i.e., the smallest scale, so,  $S = 1$  :

$$A_{ij}^{ds} = \begin{cases} 0.01 & |i - C_X| \leq R_X, |j - C_Y| \leq R_Y, d = D^A, s = S \\ 0 & \text{otherwise} \end{cases}. \quad (\text{A20})$$



## 6. References

- Adelson, E. H., and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2), 284-299.
- Adelson, E., and Movshon, A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300, 523-525.
- Alais, D., Burke, D., and Wenderoth, P. (1996). Further evidence for monocular determinants of perceived plaid direction. *Vision Research*, 36(9), 1247-1253.
- Alais, D., van der Smagt, M. J., van den Berg, A. V., and van de Grind, W. A. (1998). Local and global factors affecting the coherent motion of gratings presented in multiple apertures. *Vision Research*, 38(11), 1581-1591.
- Alais, D., van der Smagt, M. J., Verstraten, F. A., and van de Grind, W. A. (1996). Monocular mechanisms determine plaid motion coherence. *Visual Neuroscience*, 13(4), 615-626.
- Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, 52(6), 1106-1130.
- Albright, T. D. (1992). Form-cue invariant motion processing in primate visual cortex. *Science*, 255(5048), 1141-1143.
- Albright, T.D., Desimone, R., and Gross, C.G. (1984). Columnar organization of directionally sensitive cells in visual area MT of the macaque. *Journal of Neurophysiology*, 51, 16-31.
- Anstis, S. M. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society of London B*, 290, 153-168.
- Anstis, S. M. (1990) Imperceptible intersections: The chopstick illusion. In: A. Blake, and T. Troscianko, *AI and the Eye*, Chapter 5 (pp105-117). New York: John Wiley and sons.
- Ariel, M., and Daw, N. W. (1982). Pharmacological analysis of directionally sensitive rabbit retinal ganglion cells. *Journal of Physiology*, 324, 161-185.
- Baloch, A. A., and Grossberg, S. (1997). A neural model of high-level motion processing: Line motion and formotion dynamics. *Vision Research*, 37(21), 3037-3059.
- Baloch, A. A., Grossberg, S., Mingolla, E., and Nogueira, C. A. (1999). Neural model of first-order and second-order motion perception and magnocellular dynamics. *Journal of the Optical Society of America A*, 16(5), 953-978.
- Barlow, H. B., Hill, R. M., and Levick, W. R. (1964). Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *Journal of Physiology*, 173, 377-407.
- Barlow, H. B., and Levick, W. R. (1965). The mechanism of directionally selective units in the rabbit's retina. *Journal of Physiology*, 178, 477-504.
- Bowns, L. (1996). Evidence for a feature tracking explanation of why type II plaids move in the vector sum direction at short durations. *Vision Research*, 36(22), 3685-3694.
- Bradley, D. C., Qian, N., and Andersen, R. A. (1995). Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature*, 373(6515), 609-611.
- Bregman, A.L. (1981). Asking the "what for" question in auditory perception. In M. Kubovy and J.R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum Associates, 99-118.
- Bressan, P., Ganis, G., and Vallortigara, G. (1993). The role of depth stratification in the solution of the aperture problem. *Perception*, 22(2), 215-228.
- Burke, D., Alais, D., and Wenderoth, P. (1994). A role for a low level mechanism in determining plaid coherence. *Vision Research*, 34(23), 3189-3196.
- Carpenter, G., A., and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computational Vision Graphics Image Processing*, 37, 54-115.
- Cavanagh, P. (1992). Attention-based motion perception. *Science*, 257, 1563-1565.
- Chey, J., Grossberg, S., and Mingolla, E. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America A*, 14(10), 2570-2594.
- Chey, J., Grossberg, S., and Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, 38, 2769-2786.
- Cox, M. J., and Derrington, A. M. (1994). The analysis of motion of two-dimensional patterns: Do Fourier components provide the first stage? *Vision Research*, 34(1), 59-72.
- Dean, A. F., and Tolhurst, D. J. (1986). Factors influencing the temporal phase of response to bar and grating stimuli for simple cells in the cat striate cortex. *Experimental Brain Research*, 62(1), 143-151.

- DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1993a). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, 69(4), 1091-1117.
- DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1993b). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *Journal of Neurophysiology*, 69(4), 1118-1135.
- Del Viva, M. M., and Morrone, M. C. (1998). Motion analysis by feature tracking. *Vision Research*, 38(22), 3633-3653.
- Derrington, A. M., and Ukkonen, O. I. (1999). Second-order motion discrimination by feature-tracking. *Vision Research*, 39(8), 1465-1475.
- Desimone, R., and Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *Journal of Computational Neurology*, 248(2), 164-189.
- DeYoe, E. A., and Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, 11(5), 219-226.
- Duffy, C. J., and Wurtz, R. H. (1991a). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65(6), 1329-1345.
- Duffy, C. J., and Wurtz, R. H. (1991b). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *Journal of Neurophysiology*, 65(6), 1346-1359.
- Emerson, R. C., Bergen, J. R., and Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, 32(2), 203-218.
- Emerson, R. C., Citron, M. C., Vaughn, W. J., and Klein, S. A. (1987). Nonlinear directionally selective subunits in complex cells of cat striate cortex. *Journal of Neurophysiology*, 58(1), 33-65.
- Emerson, R. C., and Coleman, L. (1981). Does image movement have a special nature for neurons in the cat's striate cortex? *Investigative Ophthalmology and Visual Sciences*, 20(6), 766-783.
- Emerson, R. C., and Gerstein, G. L. (1977). Simple striate neurons in the cat. II. Mechanisms underlying directional asymmetry and directional selectivity. *Journal of Neurophysiology*, 40(1), 136-155.
- Enroth-Cugell, C., and Robson, J. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London)*, 187, 517-552.
- Fennema, C., and Thompson, W. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, 9, 301-315.
- Ferrera, V., P., and Wilson, H., R. (1990). Perceived direction of moving two-dimensional patterns. *Vision Research*, 30, 273-287.
- Ferrera, V., P., and Wilson, H., R. (1991). Perceived speed of moving two-dimensional patterns. *Vision Research*, 31, 877-893.
- Field, D.J., Hayes, A. and Hess, R.F. (1993). Contour integration by the human visual system: Evidence for a local "Association Field". *Vision Research*, 33, 173-193.
- Francis, G. and Grossberg, S. (1996a). Cortical dynamics of boundary segmentation and reset: Persistence, afterimages, and residual traces. *Perception*, 35, 543-567.
- Francis, G., and Grossberg, S. (1996b). Cortical dynamics of form and motion integration: Persistence, apparent motion, and illusory contours. *Vision Research*, 36(1), 149-173.
- Ganz, L. (1984). Visual cortical mechanisms responsible for direction selectivity. *Vision Research*, 24(1), 3-11.
- Ganz, L., and Felder, L. (1984). Mechanism of directional selectivity in simple neurons of cat's visual cortex analyzed with stationary flash sequences. *Journal of Neurophysiology*, 51, 294-324.
- Giersch, A., and Lorenceau, J. (1999). Effects of a benzodiazepine, lorazepam, on motion integration and segmentation: An effect on the processing of line-ends? *Vision Research*, 39(11), 2017-2025.
- Goodwin, A. W., Henry, G. H., and Bishop, P. O. (1975a). Direction selectivity of simple striate cells: Properties and mechanism. *Journal of Neurophysiology*, 38(6), 1500-1523.
- Goodwin, A. W., Henry, G. H., and Bishop, P. O. (1975b). Direction selectivity of complex cells in a comparison with simple cells. *Journal of Neurophysiology*, 38(6), 1524-1540.
- Groner, R., Hofer, D. and Groner, M. (1986). The role of anticipation in the encoding of motion signals-sensitization or bias. In F. Klix and H. Hagendorf (Eds.) *Human memory and cognitive capabilities*. Amsterdam, Elsevier.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 1, 1-51.
- Grossberg, S. (1987a). Cortical dynamics of three-dimensional form, color, and brightness perception, I: Monocular theory, *Perception and Psychophysics*, 41, 87-116.
- Grossberg, S. (1987b). Cortical dynamics of three-dimensional form, color, and brightness perception, II: Binocular theory. *Perception and Psychophysics*, 41, 117-158.

- Grossberg, S. (1997b). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychological Review*, 104(3), 618-658.
- Grossberg, S. (1991). Why do parallel cortical systems exist for the perception of static form and moving form? *Perception and Psychophysics*, 49(2), 117-141.
- Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, 55, 48-120.
- Grossberg, S. (1998). How is a moving target continuously tracked behind occluding cover? In: T. Watanabe, *High Level Motion Processing*, pp3-52. Cambridge, MA: MIT Press.
- Grossberg, S. (1999a). How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision*, 12(2), 163-185.
- Grossberg, S. (1999b). The link between brain learning, attention, and consciousness. *Consciousness and Cognition*, 8, 1-44.
- Grossberg, S., and Kelly, F. (1999). Neural dynamics of binocular brightness perception. *Vision Research*, 39, 3796-3816.
- Grossberg, S., and McLoughlin, N. P. (1997). Cortical dynamics of three-dimensional surface perception: Binocular and half-occluded scenic images. *Neural Networks*, 10(9), 1583-1605.
- Grossberg, S., and Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92(2), 173-211.
- Grossberg, S., and Mingolla, E. (1993). Neural dynamics of motion perception: direction fields, apertures, and resonant grouping. *Perception and Psychophysics*, 53(3), 243-278.
- Grossberg, S., Mingolla, E., and Pack, C. (1999). A neural model of motion processing and visual navigation by cortical area MST. *Cerebral Cortex*, 9(8), 878-895.
- Grossberg, S., Mingolla, E., and Ross, W. D. (1997). Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neurosciences*, 20(3), 106-111.
- Grossberg, S. and Williamson, J.R. (2000). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cerebral Cortex*, in press.
- Grossberg, S., and Pessoa, L. (1998). Texture segregation, surface representation and figure-ground separation. *Vision Research*, 38, 2657-2684.
- Grossberg, S., and Raizada, R. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40, 1413-1432.
- Grossberg, S., and Rudd, M. E. (1989). A neural architecture for visual motion perception: Group and element apparent motion. *Neural Networks*, 2, 421-450.
- Grossberg, S., and Rudd, M. E. (1992). Cortical dynamics of visual motion perception: Short-range and long-range apparent motion. *Psychological Review*, 99(1), 78-121.
- Gurnsey, R., and von Grunau, M. (1997). Illusory contour-motion arising from translating terminators. *Vision Research*, 37(8), 1007-1024.
- Hammond, P., and Kim, J. N. (1994). Spatial correlation of suppressive and excitatory receptive fields with direction selectivity of complex cells in cat striate cortex. *Proceedings of the Royal Society of London B*, 257(1349), 179-184.
- Heeger, D. J., Boynton, G. M., Demb, J. B., Seidemann, E., and Newsome, W. T. (1999). Motion opponency in visual cortex. *Journal of Neuroscience*, 19(16), 7162-7174.
- Heggelund, P. (1984). Direction asymmetry by moving stimuli and static receptive field plots for simple cells in cat striate cortex. *Vision Research*, 24(1), 13-16.
- Hildreth, E. (1984). *The Measurement of Visual Motion*. Cambridge, MA: MIT Press.
- Hochstein, S., and Shapley, R. M. (1976a). Quantitative analysis of retinal ganglion cell classifications. *Journal of Physiology (London)*, 262(2), 237-264.
- Hochstein, S., and Shapley, R. M. (1976b). Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *Journal of Physiology (London)*, 262(2), 265-284.
- Horn, B. K. P., and Schunck, B. G. (1981). Determining optic flow. *Artificial Intelligence*, 17, 185-203.
- Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148, 574-591.
- Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106-154. 49

- Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, 195(1), 215-243.
- Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394(6695), 784-787.
- Innocenti, G. M., and Fiore, L. (1974). Post-synaptic inhibitory components of the responses to moving stimuli in area 17. *Brain Research*, 80(1), 122-126.
- Jagadeesh, B., Wheat, H. S., and Ferster, D. (1993). Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex. *Science*, 262(5141), 1901-1904.
- Jagadeesh, B., Wheat, H. S., Kontsevich, L. L., Tyler, C. W., and Ferster, D. (1997). Direction selectivity of synaptic potentials in simple cells of the cat visual cortex. *Journal of Neurophysiology*, 78(5), 2772-2789.
- Jasinschi, R., Rosenfeld, A., and Sumi, K. (1992). Perceptual motion transparency: The role of geometrical information. *Journal of the Optical Society of America*, 9, 1865-1879.
- Jin, Z. F., and Srinivasan, M. V. (1990). Neural gradient models for the measurement of image velocity. *Visual Neuroscience*, 5(3), 261-271.
- Johnston, A., and Clifford, C. W. (1995). Perceived motion of contrast-modulated gratings: Predictions of the multi-channel gradient model and the role of full-wave rectification. *Vision Research*, 35(12), 1771-1783.
- Johnston, A., McOwan, P.W., and Benton, C. P. (1999). Robust velocity computation from a biologically motivated model of motion perception. *Proceedings of the Royal Society of London B*, 266(1418), 509-518.
- Johnston, A., McOwan, P. W., and Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proceedings of the Royal Society of London B*, 250(1329), 297-306.
- Kanizsa, G. (1979). *Organization in vision: Essays in Gestalt Perception*. New York: Praeger Press.
- Kelly, F., and Grossberg, S. (2001). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception and Psychophysics*, 62(18), 1596-1618.
- Kim, J. and Wilson, H.R. (1993). Dependence of plaid motion coherence on component grating directions. *Vision Research*, 33, 2479-2490.
- Koch, C., Wang, H. T., and Mathur, B. (1989). Computing motion in the primate's visual system. *Journal of Experimental Biology*, 146, 115-139.
- Komatsu, H., and Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *Journal of Neurophysiology*, 60(2), 580-603.
- Lappin, J. S., and Bell, H. H. (1972). Perceptual differentiation of sequential visual patterns. *Perception and Psychophysics*, 12, 129-134.
- Leshner, G.W. and Mingolla, E. (1993). The role of edges and line-ends in illusory contour formation. *Vision Research*, 33, 2253-2270.
- Liden, L., and Pack, C. (1999). The role of terminators and occlusion cues in motion integration and segmentation: A neural network model. *Vision Research*, 39, 3301-3320.
- Lindsey, D. T., and Todd, J. T. (1996). On the relative contributions of motion energy and transparency to the perception of moving plaids. *Vision Research*, 36(2), 207-222.
- Livingstone, M. S. (1998). Mechanisms of direction selectivity in macaque V1. *Neuron*, 20(3), 509-526.
- Loffler, G., and Orbach, H. S. (1999). Computing feature motion without feature detectors: A model for terminator motion without end-stopped cells. *Vision Research*, 39(4), 859-871.
- Lorenceau, J., and Shiffrar, M. (1992). The influence of terminators on motion integration across space. *Vision Research*, 32(2), 263-273.
- Lu, Z. L., and Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research*, 35(19), 2697-2722.
- Marr, D., and Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*, 211, 151-180.
- Marshall, J. A. (1990). Self-organizing neural networks for perception of visual motion. *Neural Networks*, 3(1), 45-74.
- Maunsell, J. H., and van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49(5), 1127-1147.

- Maunsell, J. H., and van Essen, D. C. (1983b). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience*, 3(12), 2563-2586.
- McLean, J., and Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Research*, 29(6), 675-679.
- McLean, J., Raab, S., and Palmer, L. A. (1994). Contribution of linear mechanisms to the specification of local motion by simple cells in areas 17 and 18 of the cat. *Visual Neuroscience*, 11(2), 271-294.
- Mingolla, E., Todd, J., T., and Norman, J., F. (1992). The perception of globally coherent motion. *Vision Research*, 32, 1015-1031.
- Movshon, J. A., Adelson, E. A., Gizzi, M. S., and Newsome, W. T. (1985). The analysis of moving visual patterns. In: C. Chagas, R. Gattass, and C. Gross, *Pattern Recognition Mechanisms*, Vol. 11, Experimental Brain Research Supplementum (pp. 117-151). Berlin: Springer-Verlag.
- Movshon, J. A., Thompson, I. D., and Tolhurst, D. J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology (London)*, 283, 53-77.
- Nakayama, K., Shimojo, S., and Silverman, G., H. (1989). Stereoscopic depth: Its relation to image segmentation grouping and the recognition of occluded objects. *Perception*, 18, 55-68.
- Nakayama, K., and Silverman, G. H. (1988a). The aperture problem. I: Perception of nonrigidity and motion direction in translating sinusoidal lines. *Vision Research*, 28(6), 739-746.
- Nakayama, K., and Silverman, G. H. (1988b). The aperture problem. II: Spatial integration of velocity information along contours. *Vision Research*, 28(6), 747-753.
- Nowlan, S. J., and Sejnowski, T. J. (1994). Filter selection model for motion segmentation and velocity integration. *Journal of the Optical Society of America A*, 11(12), 3177-3200.
- O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., and Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, 18(4), 591-598.
- Orban, G. A., Lagae, L., Verri, A., Raiguel, S., Xiao, D., Maes, H., and Torre, V. (1992). First-order analysis of optical flow in monkey brain. *Proceedings of the National Academy of Sciences USA*, 89(7), 2595-2599.
- Pack, C., Grossberg, S., and Mingolla, E. (2001). A neural model of smooth pursuit control and motion perception by cortical area MST. *Journal of Cognitive Neuroscience*, 13(1), 102-120.
- Poggio, T., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317(6035), 314-319.
- Qian, N., and Andersen, R. A. (1994). Transparent motion perception as detection of unbalanced motion signals. II. Physiology. *Journal of Neuroscience*, 14(12), 7367-7380.
- Qian, N., Andersen, R. A., and Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals. III. Modeling. *Journal of Neuroscience*, 14(12), 7381-7392.
- Raizada, R.D.S. and Grossberg, S. (2000). Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition*, in press.
- Ramachandran, V., S., and Inada, V. (1985). Spatial phase and frequency in motion capture of random-dot patterns. *Spatial Vision*, 1(1), 57-67.
- Recanzone, G. H., Wurtz, R. H., and Schwarz, U. (1997). Responses of MT and MST neurons to one and two moving objects in the receptive field. *Journal of Neurophysiology*, 78(6), 2904-2915.
- Reichardt, W. (1961). Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In: W. A. Rosenblith. *Sensory Communication*. New York: Wiley.
- Reid, R. C., Soodak, R. E., and Shapley, R. M. (1987). Linear mechanisms of directional selectivity in simple cells of cat striate cortex. *Proceedings of the National Academy of Sciences USA*, 84(23), 8740-8744.
- Reid, R. C., Soodak, R. E., and Shapley, R. M. (1991). Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *Journal of Neurophysiology*, 66(2), 505-529.
- Rodman, H. R., and Albright, T. D. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research*, 75(1), 53-64.
- Ross, W.D., Grossberg, S., and Mingolla, E. (2000). Visual cortical mechanisms of perceptual grouping: Interacting layers, networks, columns, and maps. *Neural Networks*, 13, 571-588.
- Rubin, N., and Hochstein, S. (1993). Isolating the effect of one-dimensional motion signals on the perceived direction of moving two-dimensional objects. *Vision Research*, 33, 1385-1396.
- Sachtler, W. L., and Zaidi, Q. (1995). Visual processing of motion boundaries. *Vision Research*, 35(6), 807-826.

- Sato, H., Katsuyama, N., Tamura, H., Hata, Y., and Tsumoto, T. (1995). Mechanisms underlying direction selectivity of neurons in the primary visual cortex of the macaque. *Journal of Neurophysiology*, 74(4), 1382-1394.
- Saito, H., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., and Iwai, E. (1986). Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience*, 6(1), 145-157.
- Sekuler, R. and Ball, K. (1977). Mental set alters visibility of moving targets. *Science*, 198, 60-62.
- Shiffrar, M., Li, X., and Lorenceau, J. (1995). Motion integration across differing image features. *Vision Research*, 35(15), 2137-2146.
- Shimojo, S., Silverman, G. H., and Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29(5), 619-626.
- Shipley, T.F. and Kellman, P.J. (1992). Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception and Psychophysics*, 52, 97-106.
- Sillito, A. M. (1975). The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat. *Journal of Physiology (London)*, 250(2), 305-329.
- Sillito, A. M. (1977). Inhibitory processes underlying the directional specificity of simple, complex and hypercomplex cells in the cat's visual system. *Journal of Physiology*, 271, 699-720.
- Snowden, R. J., Treue, S., Erickson, R. G., and Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience*, 11(9), 2768-2785.
- Soriano, M., Spillman, L., and Bach, M. (1996). The abutting grating illusion. *Vision Research*, 36, 109-116.
- Stelmach, L.B., Herdman, C.M., and McNeil, R. (1994). Attentional modulation of visual processes in motion perception. *Journal of Experimental Psychology*, 20, 108-121.
- Stone, L.S., Watson, A.B., and Mulligan, J.B. (1990). Effect of contrast on the perceived direction of a moving plaid. *Vision Research*, 30, 1049-1067.
- Stoner, G. R., and Albright, T. D. (1992). Motion coherency rules are form-cue invariant. *Vision Research*, 32(3), 465-475.
- Stoner, G. R., Albright, T. D., and Ramachandran, V. S. (1990). Transparency and coherence in human motion perception. *Nature*, 344(6262), 153-155.
- Tanaka, K., and Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, 62(3), 626-641.
- Tommasi, L., and Vallortigara, G. (1999). Figure ground segregation modulates perceived direction of ambiguous moving gratings and plaids. *Vision Research*, 39(4), 777-787.
- Treue, S., and Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399(6736), 575-579.
- Treue, S., and Maunsell, J., H., R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature (London)*, 382, 539-541.
- Treue, S., and Maunsell, J., H., R. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *Journal of Neuroscience*, 19(17), 7591-7602.
- Vallortigara, G., and Bressan, P. (1991). Occlusion and the perception of coherent motion. *Vision Research*, 31(11), 1967-1978.
- van Santen, J. P., and Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, 2(2), 300-321.
- Viswanathan, L., Grossberg, S. and Mingolla, E. (1999). Neural dynamics of motion grouping across apertures. *Investigative Ophthalmology and Visual Science*, 40(4), p. S422.
- von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224, 1260-1262.
- Wallach, H. (1935). On the visually perceived direction of motion. *Psychologische Forschung*, 20, 325-380.
- Wang, R. (1997). A network model of motion processing in area MT of primates. *Journal of Computational Neuroscience*, 4(4), 287-308.
- Watson, A. B., and Ahumada, A. J. Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 2(2), 322-341.
- Wenderoth, P., Alais, D., Burke, D., and van der Zwan, R. (1994). The role of the blobs in determining the perception of drifting plaids and their motion aftereffects. *Perception*, 23(10), 1163-1169.

- Wuerger, S., Shapley, R., and Rubin, N. (1996). "On the visually perceived direction of motion" by Hans Wallach: 60 years later. *Perception*, 25, 1317-1367.
- Wyatt, H., and Daw, N. W. (1976). Specific effects of neurotransmitter antagonists on ganglion cells in rabbit retina. *Science*, 191, 204-205.
- Yo, C. and Wilson, H.R. (1992). Perceived direction of moving two-dimensional patterns depends on duration, contrast, and eccentricity. *Vision Research*, 32, 135-147.
- Yuille, A. L., and Grzywacz, N. M. (1988). A computational theory for the perception of coherent visual motion. *Nature*, 333(6168), 71-74.
- Zeki, S. M. (1974a). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *Journal of Physiology (London)*, 236(3), 549-573.
- Zeki, S. M. (1974b). Cells responding to changing image size and disparity in the cortex of the rhesus monkey. *Journal of Physiology (London)*, 242(3), 827-841.
- Zemel, R. S., and Sejnowski, T. J. (1998). A model for encoding multiple object motions and self-motion in area MST of primate visual cortex. *Journal of Neuroscience*, 18(1), 531-547.