How do object reference frames and motion vector decomposition emerge in laminar cortical circuits?

Stephen Grossberg, Jasmin Léveillé, and Massimiliano Versace

Center for Adaptive Systems
Department of Cognitive and Neural Systems
and
Center of Excellence for Learning in Education, Science, and Technology
Boston University, 677 Beacon Street, Boston, MA 02215, USA

Running title: Vector decomposition by laminar cortical circuits

January 13, 2011

Attention, Perception, & Psychophysics, in press

All correspondence should be addressed to:
Professor Stephen Grossberg
Center for Adaptive Systems
Boston University
677 Beacon Street
Boston, MA 02215
Phone: 617-353-7858/7
Fax: 617-353-7755
Email:steve@bu.edu

ABSTRACT

How do spatially disjoint and ambiguous local motion signals in multiple directions generate coherent and unambiguous representations of object motion? Various motion percepts, starting with those of Duncker and Johansson, obey a rule of vector decomposition, whereby global motion appears to be subtracted from the true motion path of localized stimulus components. Then objects and their parts are seen moving relative to a common reference frame. A neural model predicts how vector decomposition results from multiple-scale and multiple-depth interactions within and between the form and motion processing streams in V1-V2 and V1-MST, which include form grouping, form-to-motion capture, figure-ground separation, and object motion capture mechanisms. These mechanisms solve the aperture problem, group spatially disjoint moving object parts via illusory contours, and capture object motion direction signals on real and illusory contours. Inter-depth directional inhibition causes a vector decomposition whereby motion directions of a moving frame at a nearer depth suppress these directions at a farther depth and cause a *peak shift* in the perceived directions of object parts relative to the frame.

Keywords: motion perception, vector decomposition, frames of reference, peak shift, complementary computing, V2, MT, MST

Introduction

How do we make sense of the complex motions of multiple interacting objects and their parts? One required computational step is to represent the various motion paths into an appropriate reference frame. Various ways of defining a reference frame have been proposed, ranging from *retinocentric*, where an object is coded relative to the location of the activity it induces on the retina, to *geocentric*, where objects are represented independent of the observer's viewpoint (Wade and Swanston, 1987). According to an *object-centered* reference frame (Bremner, Bryant and Mareschal, 2005; Wade and Swanston, 1996), objects are perceived relative to other objects. For example, on a cloudy night, the moon may appear to be moving in a direction opposite to that of the clouds. In a laboratory setting, this concept is well illustrated by induced motion experiments, wherein the motion of one object appears to cause opponent motion in another, otherwise static, object (Duncker, 1938).

Frames of reference. From a functional perspective, the creation of perceptual relative frames of reference may be one mechanism evolved by the brain to represent the motion of individual objects in a scene. This ability appears especially important when considering that the meaningfulness of the motion of a particular object can often be compromised by the motion of another object. For example, when looking at a person waving a hand from a moving train, the motion components of the hand and the train become mixed together. By representing the motion of the hand relative to that of the train, the motion component of the train can be removed, and the motion of the hand itself recovered (Rock, 1990). Relative reference frames may also be more sensitive to subtle variations in the visual scene, as suggested by the lower thresholds for motion detection in the presence of a neighboring stationary reference than in completely dark environments (Sokolov and Pavlova, 2006).

Another evolutionary advantage may be that information represented in an object-centered reference frame is partly invariant to changes in viewpoint (Wade and Swanston, 2001). Furthermore, as exemplified by the model presented here, computing an object-centered reference frame does not necessitate a viewer-centered representation (Sedgwick, 1983; Wade and Swanston, 1987), making it an efficient substitute for the latter.

Aperture problem. How does the laminar organization of visual cortex create such a reference frame? The neural model proposed in this article predicts how the form and motion pathways in cortical areas V1, V2, MT and MST accomplish this task using multiple-scale and multiple-depth interactions within and between form and motion processing streams in V1-V2 and V1-MT. These mechanisms have elsewhere been developed to explain data about motion perception by proposing how the brain solves the classical *aperture problem*. Wallach (1935/1996) first showed that the motion of a featureless line seen behind a circular aperture is perceptually ambiguous: no matter what may be the real direction of motion, the perceived direction is perpendicular to the orientation of the line; i.e., the normal component of motion. The aperture problem is faced by any localized neural motion sensor, such as a neuron in the early visual pathway, which responds to a local contour moving through an aperture-like receptive field. In contrast, a moving dot, line end or corner provides unambiguous information about an object's true motion direction (Shimojo, Silverman and Nakayama, 1989). The barber pole illusion demonstrates how the motion of a line is determined by unambiguous signals formed at its terminators, and how these unambiguous signals capture the motion of nearby ambiguous motion regions (Ramachandran and Inada, 1985; Wallach, 1935/1996). The model proposes how such moving visual features activate cells in the brain that compute *feature-tracking signals* which can disambiguate an object's true direction of motion. Our model does not rely on local pooling across motion directions, which has been shown to be unable to account for various data on motion perception (Amano et al., 2009). Instead, a dominant motion direction is determined over successive competitive stages with increasing receptive field sizes, while preserving various candidate motion directions at each spatial position up to the highest model stages, where motion grouping processes make determine the perceived directions of object motion.

The model is here further developed to simulate key psychophysical percepts such as classical motion perception experiments (Johansson, 1950) and the Duncker wheel (Duncker, 1938), and variants thereof, and casts new light on various related experimental findings. In particular, the model makes sense of psychophysical evidence which suggests that properties shared by groups of objects determine a common coordinate frame relative to which the features particular to individual objects are perceived. This process is well summarized in the classical concept of *vector decomposition* (Johansson, 1950).

Vector decomposition. Johansson (1950) showed that the perceived motion of a stimulus can be characterized as a linear combination of motion vectors corresponding to different stimulus parts. Accordingly, the true motion vectors (i.e., the vectors generated by the true motion path of the stimulus) are dissociated into orthogonal components. One component represents the motion of the grouped stimulus, or, in some cases, of a large stimulus element which appears to encompass smaller ones (e.g. the rectangular frame in induced motion experiments). The other component corresponds to the motion of individual objects from which the first component has been subtracted. An example of this vector decomposition process is shown in Figure 1.

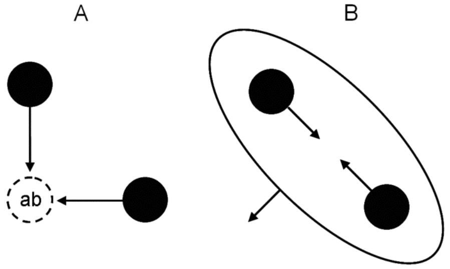

Figure 1. Johansson's experiment 19. (A) The stimulus consists of two dots oscillating in orthogonal directions and meeting periodically at point *ab*. (B) The emergent percept is that of two dots oscillating on a common diagonal axis (represented as an ellipse) which itself oscillates in the orthogonal direction.

Figure 1A depicts the visual stimulus presented to the subject. Here, two dots oscillate in orthogonal directions and meet at one endpoint (point *ab*) of their trajectories. Observers report viewing either the non-rigid motion shown in Figure 1B, or the rigid motion of a bar rotating in depth. The former percept is that of two dots oscillating along a common diagonal axis, denoted by the ellipse, which itself oscillates along the orthogonal direction. In other words, the dots are seen as moving relative to a common reference frame, the diagonal axis. The pertinence of vector decomposition to the stimulus of Figure 1 is shown in greater detail in Figure 2.

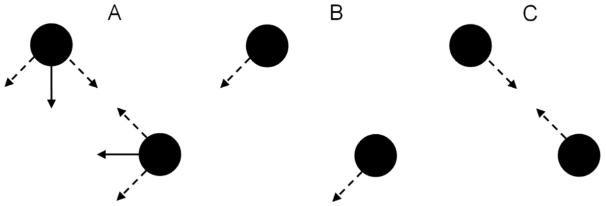

Figure 2. Vector decomposition analysis. (A) The true motion vectors (solid vectors) are cast into an orthogonal basis (dashed vectors). (B) In this basis, the component common to both dots is directed towards the *southwest* corner. (C) The remaining component, specific to each dot, moves along a common axis.

Figure 2A shows vector components into which downward and leftward motions of the individual dots can be decomposed. If the moving frame captures the diagonal direction down-and-left, as in Figure 2B, then the individual dots are left with components that oscillate towards and away from each other, as in Figure 2C. A complete account of vector decomposition requires simultaneously representing common and part motion components. In our model, simultaneous representation of both types of motion is made possible by having cells from different depth planes represent the different motion components. Subtraction of the common motion component is due to inhibition from cells coding for the nearer depth to cells coding for the farther depth. We show below how inter-depth directional inhibition causes a *peak shift* in directional selectivity that behaves like a vector decomposition.

Following Johansson (1950), vector decomposition has been invoked to explain motion perception in multiple experiments employing a variety of stimulus configurations (e.g., Börjesson and von Hofsten, 1972, 1973, 1975, 1977; Cutting and Proffitt, 1982; Di Vita and Rock, 1997; Gogel and MacCracken, 1979; Gogel and Tietz, 1976; Johansson, 1974; Post, Chi, Heckmann and Chaderjian, 1989). The bulk of this work supports the view that vector decomposition is a useful concept in characterizing object-centric frames of reference in motion perception. However, no model has so far attempted to explain how vector decomposition results from the perceptual mechanisms embedded in the neural circuits of the visual system.

The present article fills this gap by further developing the 3D FORMOTION model (Baloch and Grossberg , 1997; Berzhanskaya, Grossberg and Mingolla, 2007; Chey, Grossberg and Mingolla, 1997, 1998; Francis and Grossberg, 1996; Grossberg, Mingolla and Viswanathan, 2001; Grossberg and Pilly, 2008). As the model's name suggests, it proposes how form and motion processes interact to form coherent percepts of object motion in depth and already proposes a unified mechanistic explanation of many perceptual facts, including the barber pole illusion, plaid motion and transparent motion. Form and motion processes, such as those in V2/V4 and MT/MST, occur in the What ventral and Where dorsal cortical processing streams, respectively. Key mechanisms within the What ventral and Where streams seem to obey computationally *complementary* laws (Grossberg, 2000): The ability of each process to compute some properties prevents it from computing other, complementary, properties. Examples of such complementary properties include boundary completion vs. surface filling-in—within the (V1 interblob)-(V2 interstripe) and (V1 blob)-(V2 thin stripe) streams, respectively—and, more relevant to the results herein, boundary orientation and precise depth vs. motion direction and coarse depth—within the V1-V2 and V1-MT streams, respectively. The present article clarifies some of the interactions between form and motion processes that enable them to overcome their complementary deficiencies and to thereby compute more informative representations of unambiguous object motion.

3D FORMOTION model

Figure-ground separation mechanisms play a key role in explaining vector decomposition data. Many data about figure-ground perception have been modeled as part of the *Form-and-Color-and-DEpth* (FACADE) theory of 3D vision (e.g., Fang and Grossberg, 2009; Grossberg, 1994, 1997; Grossberg and Kelly, 1999; Grossberg and McLoughlin, 1997; Grossberg and Pessoa, 1998; Grossberg and Yazdanbakhsh, 2005; 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 V4. Figure-ground separation processes that are needed for the present analysis are predicted to be completed within the pale stripes of cortical area V2. These figure-ground processes help to segregate occluding and occluded objects, along with their terminators, onto different depth planes.

In response to the dot displays of Figure 1, the model clarifies how an illusory contour forms between the pair of moving dots within cortical area V2 and captures motion direction signals in cortical area MT via a form-to-motion, or *formotion*, interaction from V2 to MT. The captured motion direction of this illusory contour causes vector decomposition of the motion directions of the individual dots. Indeed, at the intersection of an illusory contour and a dot, contour curvature is greater in the dot's real boundary than in the illusory contour-completed boundary, since the illusory contour is tangent to the dot boundary. This greater curvature initially results in a weaker representation of the dots' boundaries in area V2. These boundaries are then pushed farther in depth than the grouped illusory contour-completed shape due to interacting processes of orientational competition, boundary pruning and enrichment, which are described and simulated in FACADE theory.

Motion processing is performed in the Where stream whose six levels model dynamics homologous to LGN, V1, MT, and MST (Figure 3). These stages are mathematically defined in the Appendix.

Level 1: Input from LGN. In the 3D FORMOTION model of Berzhanskaya et al. (2007), as in the current model, the boundary input is not depth-specific. Rather, the boundary input models signals that come from retina and LGN into V1 (Xu, Bonds and Casagrande, 2002). This boundary is represented in both ON and OFF channels. After V1 motion processing, described below, the motion signal then goes on to MT and MST. The 3D figure-ground separated boundary inputs in the current model come from V2 to MT and select bottom-up motion inputs from V1 in a depth-selective way. This process clarifies how the visual system uses occlusion clues to segregate moving boundaries into different depth planes, even though the inputs themselves occur within the same depth plane.

Berzhanskaya et al. (2007) showed how a combination of habituative (Appendix Eqs. 7—9) and depth selection (Appendix Eq. 21) mechanisms accomplish the required depth segregation of motion signals in stimuli containing both static and moving components, such as chopstick displays (Lorenceau and Alais, 2001). In particular, habituative preprocessing enables motion cues to trigger the activation of transient cells (model Level 2 in Figure 3), whereas signals due to static elements in the display habituate and become weak over time. As simulated by Berzhanskaya et al. (2007), this mechanism can explain why visible occluders in a chopstick display generate weaker motion signals at all depth planes. Although not necessary in current simulations due to the absence of static elements in the displays, habituative mechanisms in the early stages of the model are included to enable a unified explanation of the data.

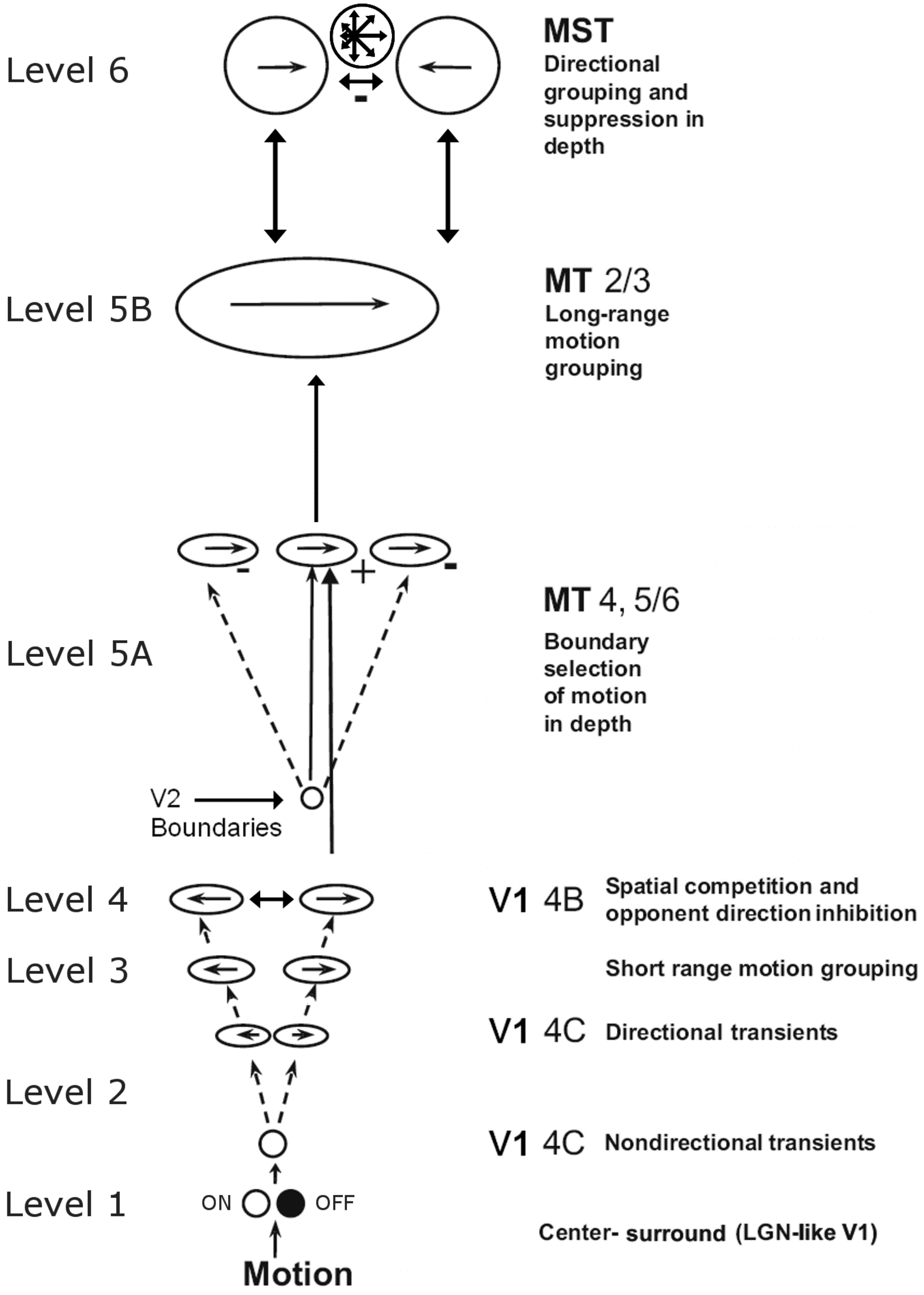
Level 6
MST
Directional grouping and suppression in depth
-
Level 5B
MT 2/3
Long-range motion grouping
-
+
-
Level 5A
MT 4, 5/6
Boundary selection of motion in depth
V2 Boundaries
Level 4
V1 4B
Spatial competition and opponent direction inhibition
Level 3
Short range motion grouping
V1 4C
Directional transients
Level 2
V1 4C
Nondirectional transients
Level 1
ON
OFF
Center- surround (LGN-like V1)
Motion

Figure 3. Motion processing stream of the 3D FORMOTION model. Level 1: ON and OFF input cells. Level 2: transient non-directional and directional cells. Level 3: short-range filter. Level 4: spatial competition and opponent direction inhibition. Level 5A: boundary selection of motion signals at multiple depth planes. Level 5B: long-range spatial filter. Level 6: directional grouping and depth suppression.

The motion selection mechanism separates motion signals in depth by using depth-separated boundary signals from V2 to MT. The model of Berzhanskaya et al. (2007) simulated in greater detail the formation of these depth-separated boundaries. The current model uses algorithmically defined boundaries to simplify simulations. The model shows how these boundaries can capture only the appropriate motion signals onto their respective depth planes in MT. Although the question of how the time-course of boundary formation impacts vector decomposition is not analyzed in detail in the current article, in part because there do not seem to be empirical data on this matter, some of our results nevertheless begin to address this issue, such as the persistence of the perceived motion until a large fraction of the boundary is pruned (see Figure 15).

Both ON and OFF input cells are needed. For example, when a bright dot move downward on a dark background (Figure 4A), ON cells respond to its lower edge (Figure 4B), but OFF cells respond to its upper edge (Figure 4C). Likewise, when the dot reverses direction and starts to move upward, the upper edge now activates ON cells and the lower edge OFF cells. By differentially activating ON and OFF cells in different parts of this motion cycle, these cells have more time to recover from habituation, so that the system remains more sensitive to repetitive motion signals. Model ON and OFF responses are thus relevant to the role played by habituative mechanisms in generating transient cell responses.

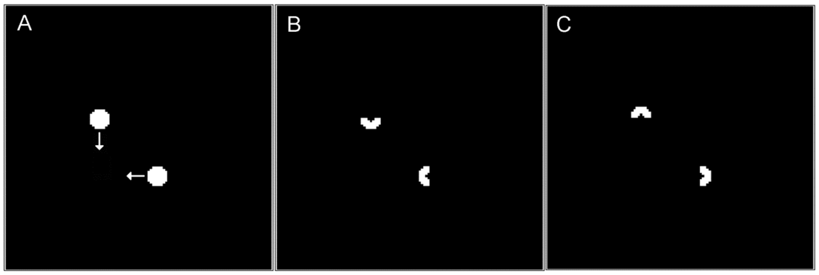

Figure 4. Input to the motion pathway. (A) Motion path of the dots directed toward the lower left corner. The input to the ON input cells corresponds to the leading edge of the dot in (B), whereas the input to the OFF input cells corresponds to the trailing edge (C).

Level 2: Transient cells. The second stage of the motion processing system (Figure 3) consists of non-directional transient cells, inhibitory directional interneurons, and directional transient cells. The non-directional transient cells respond briefly to a change in the image luminance, irrespective of the direction of movement (Appendix Eqs. 7—9). Such cells respond well to moving boundaries and poorly to static objects because of the habituation that creates the transient response. The type of adaptation that leads to these transient cell responses is known to occur at several stages in the visual system, ranging from retinal Y cells (Enroth-Cuggell and Robson, 1966; Hochstein and Shapley, 1976a, 1976b) to cells in V1 and V2 (Abbott, Sen, Varela, and Nelson, 1997; Carandini and Ferster, 1997; Chance, Nelson, and Abbott, 1998; Francis and Grossberg, 1996a, 1996b; Francis, Grossberg and Mingolla, 1994; Varela et al., 1997) and beyond. The non-directional transient cells send signals to inhibitory directional interneurons and directional transient cells, and the inhibitory interneurons interact with each other and with the directional transient cells (Eqs. 10 and 11). A directional transient cell fires vigorously when a stimulus is moved through its receptive field in one direction (called the preferred direction), while motion in the reverse direction (called the null direction) evokes little response (Barlow and Levick, 1965).

The directional inhibitory interneuronal interaction enables the directional transient cells to realize directional selectivity at a wide range of speeds (Chey et al., 1997; Grossberg et al., 2001). Although in the present model directional interneurons and transient cells correspond to cells in V1, this predicted interaction is consistent with retinal data concerning how bipolar cells interact with inhibitory starburst amacrine cells and direction-selective ganglion cells, and how starburst cells interact with each other and with ganglion cells (Fried, Münch, and Werblin, 2002). The possible role of starburst cell inhibitory interneurons in ensuring directional selectivity at a wide range of speeds has not yet been tested. The model is also consistent with physiological data from cat and macaque species showing that directional selectivity first occurs in V1, and that it is due at least in part to inhibition which reduces the response to the null direction of motion (Livingstone, 1998; Murthy and Humphrey, 1999).

Level 3: Short-range filter. A key step in solving the aperture problem is to strengthen unambiguous feature tracking signals relative to ambiguous motion signals. Feature tracking signals are often generated by a relatively small number of moving features in a scene, yet can have a very large effect on motion perception. One process that strengthens feature tracking signals relative to ambiguous aperture signals is the short-range directional filter (Figure 3). Cells in this filter *accumulate 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 selectively strengthens the responses of short-range filter cells to feature tracking signals at unoccluded line endings, object corners, and other scenic features (Appendix Eq. 12). The use of a short-range filter followed by competition at Level 4 eliminates the need for an explicit solution of the *feature correspondence problem* that various other models posit and attempt to solve (Reichardt, 1961; Ullman, 1979; van Santen and Sperling, 1985).

The short-range filter uses multiple spatial scales (Appendix Eq. 16). Each scale responds preferentially to a specific speed range. Larger scales respond better to faster speeds due to thresholding of short-range filter outputs with a self-similar threshold; that is, a threshold that increases with filter size (Appendix Eq. 17). Larger scales thus require "more evidence" to fire (Chey et al., 1998).

Level 4: Spatial competition and opponent direction competition. Two kinds of competition further enhance the relative advantage of feature tracking signals (Figure 3, Appendix Eqs. 18-20). These competing cells are proposed to occur in layer 4B of V1 (Figure 3). Spatial competition among cells of the same spatial scale that prefer the same motion direction boosts the amplitude of feature tracking signals relative to those of ambiguous signals. Feature tracking signals are contrast-enhanced by such competition because they are often found at motion discontinuities, and thus get less inhibition than ambiguous motion signals that lie within an object's interior. Opponent-direction competition also occurs at this processing stage (Albright, 1984; Albright, Desimone, and Gross, 1984) and ensures that cells tuned to opposite motion directions are not simultaneously active.

The activity pattern at this model stage is consistent with data of Pack, Gartland, and Born (2004). In their experiments, V1 cells demonstrate an apparent suppression of responses to motion along visible occluders. A similar suppression occurs in the model due to the adaptation of transient inputs to static boundaries. Also, cells in the middle of a grating respond more weakly than cells at the edge of the grating. Spatial competition in the model between motion signals performs divisive normalization and endstopping, which together amplify the strength of directionally unambiguous feature tracking signals at line ends relative to the strength of aperture-ambiguous signals along line interiors.

Level 5: Long-range filter and formotion selection. Motion signals from model layer 4B of V1 input to model area MT. Area MT also receives a projection from V2 (Anderson and Martin, 2002; Rockland, 1995) that carries depth-specific figure-ground-separated boundary signals whose predicted properties were supported by Ponce, Lomber and Born (2008). These V2 form boundaries select the motion signals (*formotion selection)* by selectively capturing at different depths the motion signals coming into MT from layer 4B of V1 (Appendix Eq. 21).

Formotion selection, or selection of motion signals in depth by corresponding boundaries, is proposed to occur via a narrow excitatory center and broad inhibitory surround projection from V2 to layer 4 of MT. First, in response to the oscillating dot pair, the larger spatial scale at the nearer depth (D1) in V2 allows illusory contours to bridge the two dots. At the same time, ON-center OFF-surround spatial competition inhibits boundaries within the enclosing shape at that depth (Figure 5A). In the smaller spatial scale of farther depth (D2) of V2, no illusory contours bridge the dots. In addition, boundaries at the farther depth are inhibited by corresponding ones at the nearer depth at the corresponding positions. The resulting pruned boundaries are shown in gray in Figure 5B.

Formotion selection from V2 to MT is depth-specific. At the nearer depth D1, V2 boundary signals that correspond to the illusory contour grouping select the larger-scale motion signals (Figure 5A), and suppresses motion signals at other locations in that same depth. At the farther depth D2, V2 boundary signals that correspond to the individual dots (Figure 5B) select motion signals that represent the motion of individual parts of the stimulus.

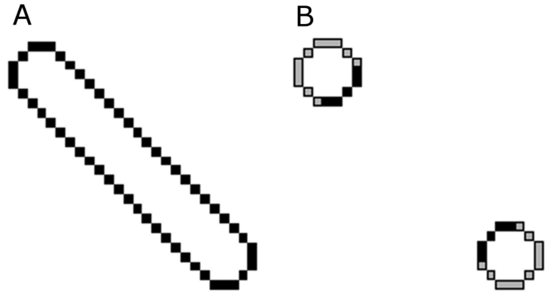

Figure 5. V2 input to MT for the dot configuration of Figure 1. Strong boundaries are represented in black whereas weaker boundaries are represented in gray. (A) Nearer depth (larger scale) input contains FACADE boundaries corresponding to the dots and illusory contours linking the pair. The parts of the dot boundaries that would be located inside the enclosing shape are inhibited due to spatial competition. (B) Farther depth (smaller scale) input contains the boundaries of individual dots. Dot boundaries in that depth and at the same spatial locations as boundaries in the nearer depth are inhibited by the latter, due to near-to-far suppression (see Eq. 27), and are thus shown as being weaker.

Boundary-gated signals from layer 4 of MT are proposed to input to the upper layers of MT (Figure 3; Appendix Eq. 23), where they activate directionally-selective, spatially anisotropic filters via long-range horizontal connections (Appendix Eq. 26). In this long-range directional filter, motion signals coding the same directional preference are pooled from object contours with multiple orientations and opposite contrast polarities. This pooling process creates a true directional cell response (Chey et al., 1997; Grossberg et al., 2001; Grossberg and Rudd, 1989, 1992).

The long-range filter accumulates evidence of a given motion direction using a kernel that is elongated in the direction of that motion, much as in the case of the short-range filter. This hypothesis is consistent with data showing that approximately 30% of the cells in MT show a preferred direction of motion that is aligned with the main axis of their receptive fields (Xiao, Raiguel, Marcar and Orban, 1997). Long-range filtering is performed at multiple scales according to the size-distance invariance hypothesis (Chey et al., 1997; Hershenson, 1999): signals in the nearer depth are filtered at a larger scale and signals in the farther depth are filtered at a smaller scale.

The model hereby predicts that common and part motions are simultaneously represented by different cell populations in MT due to form selection. This type of effect may be compared with the report that some MT neurons are responsive to the global motion of a plaid stimulus, whereas others respond to the motion of its individual sinusoidal grating components (Rust, Mante, Simoncelli and Movshon, 2006; Smith, Majaj and Movshon, 2005).

The long-range filter cells in layer 2/3 of model MT are proposed to play a role in binding together *directional* information that is homologous to the coaxial and collinear accumulation of *orientational* evidence within layer 2/3 of the pale stripes of cortical area V2 for perceptual grouping of form (Grossberg 1999; Grossberg and Raizada, 2000). This anisotropic long-range motion filter allows directional motion signals to be integrated across the illusory contours in Figure 5A which link the pair of dots.

Level 6: Directional grouping, near-to-far inhibition, and directional peak shift. The model processing stages up to now do not fully solve the aperture problem. Although they can amplify feature tracking signals and assign motion signals to the correct depths, they cannot yet explain how feature tracking signals can propagate across space to select consistent motion directions from ambiguous motion directions, without distorting their speed estimates, and at the same time suppress inconsistent motion directions. They also cannot explain how motion integration can compute a vector average of ambiguous motion signals across space to determine the perceived motion direction when feature tracking signals are not present at that depth. The final stage of the model accomplishes this goal by using a motion grouping network (Appendix Eq. 29), interpreted to occur in ventral MST (MSTv), both because MSTv has been shown to encode object motion (Tanaka et al., 1993) and because it is a natural anatomical marker given the model processes that precede and succeed it. We predict that feedback between MT and MST determines the coherent motion direction of discrete moving objects.

The motion grouping network works as follows: Cells that code similar directions in MT send convergent inputs to cells in MSTv via the motion grouping network. Unlike the previous 3D Formotion model, where MST cells received input only from MT cells of the same direction, a weighted sum of directions inputs to the motion grouping cells (Appendix Eq. 30). Thus, for example, cells tuned to the southwest direction receive excitatory input not only from cells coding for that direction, but also to a lesser extent from cells tuned to either south or west directions, enabling a stronger representation of the common motion of the two dots.

Directional competition at each position then determines a winning motion direction. This winning directional cell then feeds back to its source cells in MT. This feedback supports the activity of MT cells that code the winning direction, while suppressing the activities of cells that code all other directions. This motion grouping network enables feature tracking signals to select similar directions at nearby ambiguous motion positions, while suppressing other directions there. These competitive processes take place in each depth plane, consistent with the fact that direction-tuned cells in MST are also disparity selective (Eifuku and Wurtz, 1999). On the next cycle of the feedback process, these newly unambiguous motion directions in MT select consistent MSTv grouping cells at positions near them. The grouping process hereby propagates across space as the feedback signals cycle through time between MT and MSTv.

Berzhanskaya et al. (2007), Chey et al. (1997), and Grossberg et al. (2001) have used this motion grouping process to simulate data showing how the present model solves the aperture problem. Pack and Born (2001) have provided supportive neurophysiological data, wherein the responses of MT cells over time to the motion of the interior of an extended line dynamically modulates away from the local direction that is perpendicular to the line and towards the direction of line terminator motion.

Both the V2-to-MT and the MSTv-to-MT signals carry out selection processes using modulatory on-center, off-surround interactions. The V2-to-MT signals select motion signals at the locations and depth of a moving boundary. The MST-to-MT signals select motion signals in the direction and depth of a motion grouping. Such a modulatory on-center, off-surround network was predicted by Adaptive Resonance Theory to carry out attentive selection processes in a manner that enables fast and stable learning of appropriate features to occur. See Raizada and Grossberg (2003) for a review of behavioral and neurobiological data that support this prediction in several brain systems. Direct experiments to test it in the above cases still remain to be done.

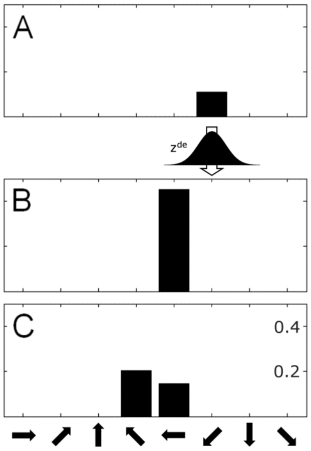

Figure 6. Peak shift in motion perception via depth suppression over the rightward moving dot in Johansson's experiment 19. (A) MST cell activity in nearer depth (larger scale). (B) MST cell activity at the same spatial location but in the farther depth (smaller scale), without depth suppression. (C) MST cell activity in the farther depth (smaller scale), with depth suppression. Each bin represents the activity in one of eight directions.

Near-to-far inhibition and peak shift is the process whereby MST cells that code nearer depth inhibit MST cells that code similar directions and positions at farther depths. In previous 3D FORMOTION models this near-to-far inhibition only involved MST cells of the same direction. Depth suppression in the current model is done via a Gaussian kernel in direction space (Appendix Eq. 32). When this near-to-far inhibition acts, it causes a peak shift in the maximally activated direction at the farther depth. This peak shift causes vector decomposition.

In particular, consider the stimulus in Figure 1. First, note that large-scale MST cells in the near plane inherit the dominant southwest motion direction of the grouped stimulus from MT layer 2/3 cells in the same plane (Figure 6A). For the same reason, MST cells in the far plane inherit the motion direction of single dots from MT layer 2/3 cells in the corresponding depth plane (Figure 6B). Figure 6C illustrates the effect of depth suppression from the direction in Figure 6A on the distribution of directionally specific activity of an MST cell that responds to the dot moving to the left.

If near-to-far depth suppression were disabled, the peak of motion activity would be in the left direction of motion (Figure 6B). With depth suppression however, motion directions close to the southwest direction are strongly inhibited, resulting in a peak shift to the northwest direction of motion (Figure 6C). The same scenario occurs, but in the opposite direction, for the vertical oscillating dot. Thus, vector decomposition occurs because of a *peak shift* in motion direction, which is in turn due to depth suppression and the representation of stimulus motion at various scales and corresponding depths. Empirical evidence supporting predicted model connections is summarized in Table 1.

Table 1. Anatomical connections.

Model connection	Functional interpretation	Selected references
LGN → V1 Layer 4	Strong LGN input; ON and OFF center-surround	Blasdel and Lund (1983); Cai et al. (1997)
V1 Layer 4 non-directional transient cells → directional transient cells	Directional selectivity	De Valois et al. (2000)
V1 Layer 4B → MT Layers 4 and 6	Feedforward local motion input to MT	Anderson et al. (1998)
V2 → MT Layers 4, 5/6	Boundary selection of motion in depth	Gattass et al. (1996), Ponce et al. (2008)
MT Layer 2/3 large receptive fields	Long-range spatial summation of motion	Born and Tootell (1992)
MT Layer 2/3 → MST	Directional motion grouping	Maunsell and van Essen (1983a)
MST → MT Layer 2/3	Selection of consistent motion direction	Maunsell and van Essen (1983a),

Simulation of psychophysical experiments

Symmetrically moving inducers. Johansson (1950, experiment 19) used a stimulus display (Figure 2) wherein each stimulus contributes equally to the common reference frame due to the symmetry in the display. Each frame in the simulation summarized by Figure 7 represents the activity of a different model level at two scales at a single time as the dots move towards the lower left corner.

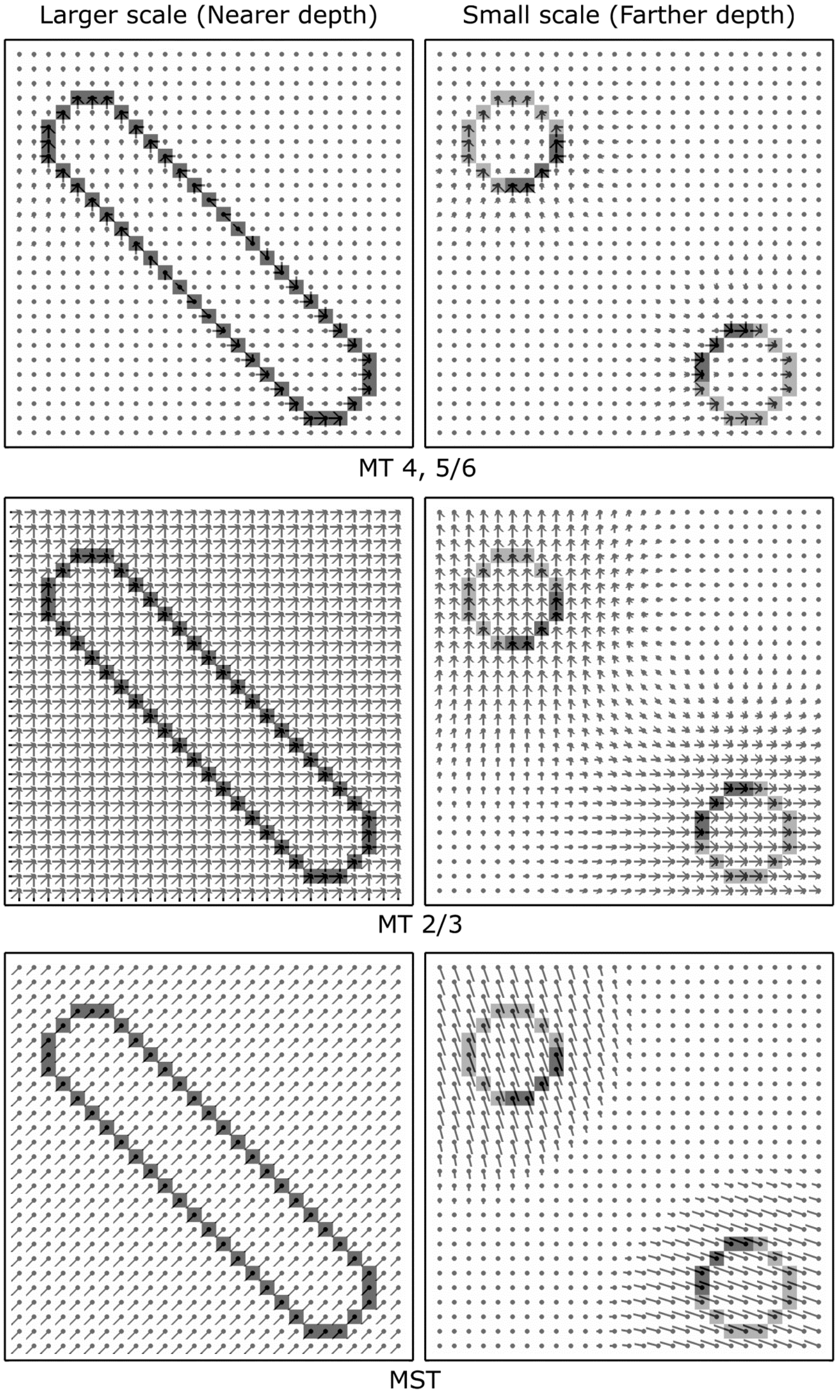

Figure 7. Simulation of Johansson's experiment 19. Each frame represents the activity of one Level at the two scales considered, for a single time-frame as the dots are moving towards the lower left corner (t = 20). V2 form boundaries select signals in MT layers 4 and 5/6 (Figure 3, Level 5A), which enhances the diagonal motion direction in the large scale, and the horizontal/vertical motion directions in the small scale. Long-range filtering in MT layer 2/3 (Figure 3, Level 5B) groups motion signals over the area subtended by the stimulus. Directional competition in MST (Figure 3, Level 6) results in an enhanced diagonal direction of motion in the large scale, which is then subtracted from the corresponding activity in the small scale, resulting in an inward peak shift.

For ease of viewing, network activity is overlaid on top of the V2 boundary input which is depicted in gray. Motion signals selected by V2 boundaries in MT layers 4 and 5/6 are displayed in the top row. The larger scale (left) selects motion signals corresponding to the grouped boundary, whereas the smaller scale (right) selects motion signals corresponding to individual dots. Long-range filtering in MT layer 2/3 (middle row) groups motion signals at each scale. Thus, in the larger scale, the coherent southwest direction is enhanced with respect to what its activity level was at the previous layer. In comparison, the smaller scale maintains the physical motion directions corresponding to each dot. Directional competition in MSTv (bottom row) results in an enhanced diagonal direction of motion in the large scale, which is then subtracted from the corresponding activity in the small scale, resulting in an inward peak shift. Note that the magnitude of the shift reported in Figure 7 is less than the 45° initially reported in Johansson (1950), compatible with results from a more recent instantiation of this paradigm, where angles of 30-40° were reported (Wallach, Becklen and Nitzberg, 1985).

Wallach et al. (1985) explained this result by noting that it corresponds to the average direction which combines the true motion paths and the paths formed by the dots moving relative to each other, a mechanism they called *process combination*. In the model, the magnitude of the shift can be controlled by varying the strength of suppression in depth, which balances the contributions of the real and relative motion paths. Process combination can therefore be interpreted as resulting from the interaction of feedforward mechanisms representing true motion paths and feedback mechanisms responsible for the peak shift in motion direction. Figure 8 shows the MST cell activity in the two scales at the two critical moments of the stimulus cycle: when the dots move toward the left corner (top), and when they move in the reverse direction toward their respective origins (bottom). Note the reversal of motion directions in the small scale, which is again consistent with the percept and obeys the principle of vector decomposition.

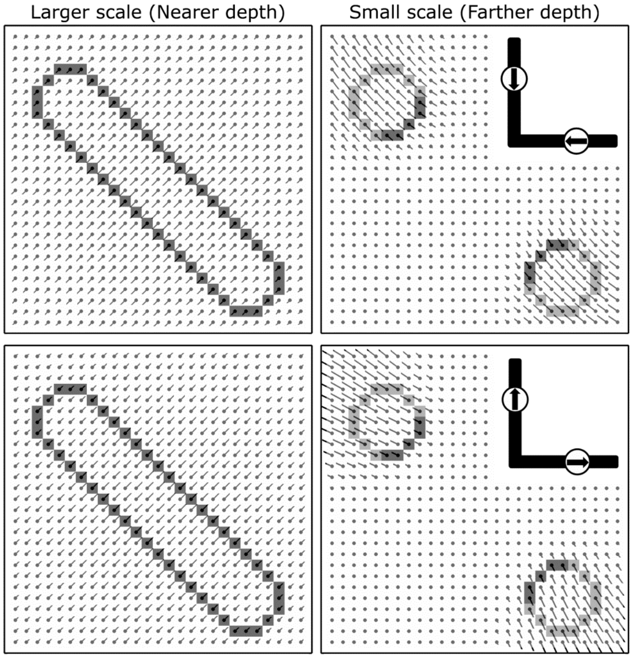

Figure 8. Two different phases of Johansson's experiment 19. (Top) Motion towards the lower left corner causes the dots to be perceived as moving inwardly. (Bottom) Motion towards the outer corner results in an outward motion percept. Insets indicate which phase of the stimulus corresponds to the activity shown.

In his description of this experiment, Johansson (1950, p.89) reported that this motion configuration was not the only one that subjects experienced. The physical motion path of one of the two dots could be recovered with overt attention directed to that dot, in which case the unattended dot was seen as on a sloping path, or even 3D rigid motion of a rotating rod could be perceived. The simulation of Figure 9 was obtained by attending in the nearer depth to the motion direction of the horizontal oscillating dot. As observed by Johansson (1950), attending to the horizontal oscillating dot in the westward direction results in the perception of its real direction of motion in the nearer depth, while the motion of the unattended dot is on a sloped path in the farther depth. Previous explanations of how top-down attention can bias form and motion percepts can also be applied here (Berzhanskaya et al., 2007; Grossberg and Swaminathan, 2004; Grossberg and Yazdanbakhsh, 2005). In Figure 9, the slanted motion direction of the vertical dot results from a peak shift induced by the strong westward motion direction induced in the larger scale by the attended horizontal dot. In the model, top-down attention operates in the motion stream at the level of MST cells (Appendix Eqs. 29 and 31).

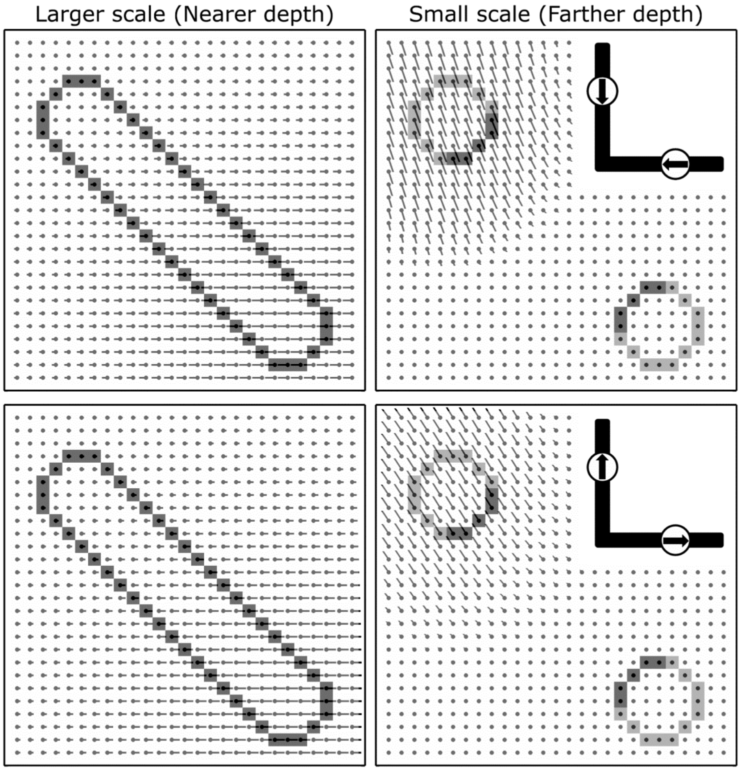

Figure 9. **Simulation of Johansson's experiment 19 with attention top-down to the horizontal moving dot. The true motion direction of the horizontal dot is perceived in the nearer plane while the path of the non-attended dot is seen as moving on a sloping path, as described in Johansson (1950)'s original experiment. (Top) Motion perceived as the dots move towards the lower left corner. (Bottom) Motion perceived as the dots moved towards the outer corner.**

The robustness of the results in Figures 7-9 can be assessed by considering that the network with the same parameters simulates a related experiment, where the dot paths intersect at their midpoint rather than at one end (Johansson, 1950; experiment 20), such that observers report a similar percept as in the previous experiment, with the difference that four phases can be distinguished: when the dots move to the lower left toward the center, away from the center, to the upper right toward the center, and then away from the center. Figure 10 shows the peak shifted activity obtained in the small scale at the four crucial phases.

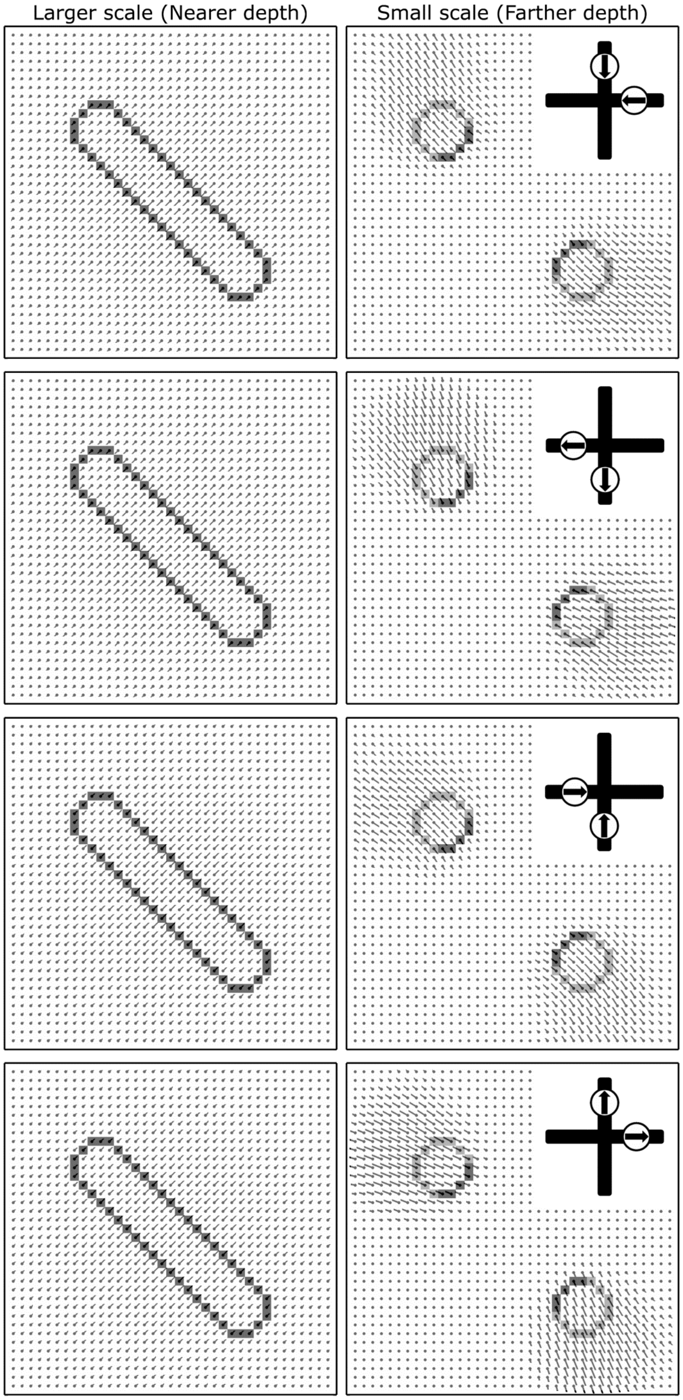

Figure 10. Simulation of Johansson's experiment 20 with attention top-down to the grouped diagonal motion in the nearer plane at the level of MST. Insets indicate which phase of the stimulus corresponds to the activity shown.

Rolling wheel experiment. The rolling wheel experiment of Duncker (1929/1938) demonstrates that not all elements in a display need contribute equally to the emergence of a relative reference frame. The experiment can be described as follows (Figure 11; see Appendix Eq. 4): a single dot moving on the rim of a rolling illusory wheel is perceived to move according to its physical trajectory, in this case a cycloid curve (Figure 11A). If a second dot is added that moves as if on the hub of the same illusory wheel (Figure 11B), then the cycloid is seen as orbiting on a circular path with the hub at its center and translating to the right (Figure 11C).

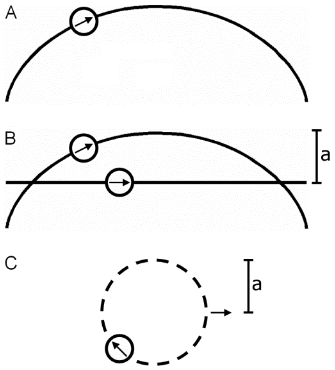

Figure 11. Rolling wheel experiment (Duncker, 1929/1938). (A) When a single dot is seen moving on a cycloid path, which describes the motion of a dot on the rim of a wheel, cycloid motion is seen. (B) When an additional dot moves on the hub of an illusory wheel of radius *a*, the cycloid path is then perceived as rotating on a circular path around the hub (C), and the total stimulus is seen as moving globally to the right.

A proper account of the Duncker wheel experiment must explain the percept of true motion in the case of the single cycloid dot and the rotational motion perceived over the cycloid dot in the two-dot configuration, as well as the global rightward motion in the later configuration. Figure 12 shows that the network is able to represent the true cycloid motion path at the level of MST cells in the single cycloid dot case. Here, each polar histogram shows the distribution of motion directions in MST cells over the area subtended by the cycloid dot at a particular phase of the revolution.

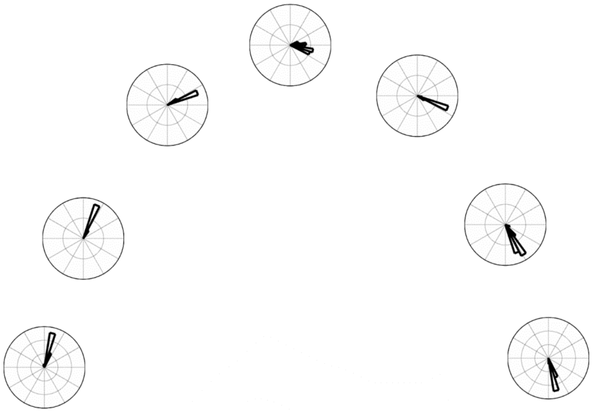

Figure 12. Simulations of a single cycloid dot. Each polar histogram shows the motion activity over the cycloid dot at different phases of one rotation cycle. The presence of multiple bins in a given histogram denotes activation in multiple directions.

Johansson (1974) provided a mathematical explanation of the wheel experiment in terms of vector analysis. As before, if the motion common to both dots is subtracted from the cycloid dot's physical motion, the cycloid dot is seen to move in a circle around the center dot.

It has been suggested that the visual system treats the dot moving with constant velocity as the center of a configuration relative to which the motion of the other dots is perceived (Cutting and Proffitt, 1982; Rubin and Richards, 1988). The successive short-range and long-range directional filtering stages in the 3D FORMOTION model implement this constraint by accumulating directional evidence in the constant rightward motion direction of the hub dot. A strong rightward motion direction in the large scale hereby emerges at the hub and captures the motion of the cycloid dot. Figure 13 shows the activity observed at the level of MST (large scale) over time for the pixels located at the center of the hub (Figure 13A) and the cycloid dot (Figure 13B).

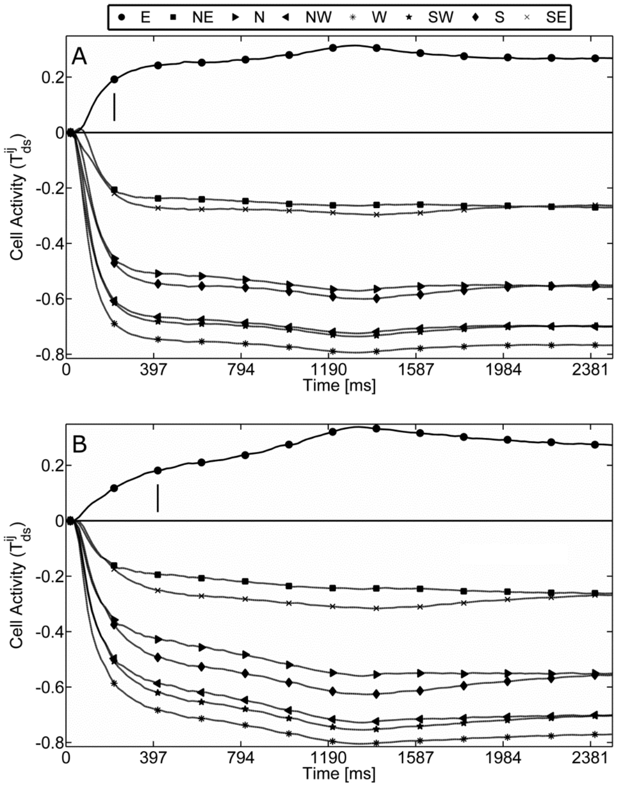

Figure 13. Horizontal motion capture in the large (near) scale over both stimulus dots. Activity in the rightward direction is shown by the curve denoted "E". In each graph, a small vertical line on the x-axis indicates the time at which rightward motion activity reaches 0.18. (A) Activity observed over the hub dot's location. (B) Activity observed over the cycloid dot's location. Notice how the rightward direction quickly becomes dominant and stable over time.

Note the early appearance of the rightward motion direction over the hub as compared to the cycloid. This is made explicit in Figure 13 by a small vertical bar on the horizontal axis of each graph, which marks the time at which corresponding levels of activity are reached for both dots. The rightward motion signal propagates to the cycloid dot over the illusory contours that join them through time. The rightward direction of motion is retained at the position of the cycloid dot even though its position on the y-axis changes throughout the simulation.

The 3D FORMOTION model predicts that elements of a visual display with constant velocity are more likely to govern the emergence of a frame of reference due to the accumulation of motion signals in the direction of motion. A related prediction is that stimuli designed to prevent such accumulation of evidence will not develop a strong object-centered frame of reference. Partial support for this prediction can be found in an experiment by Kolers (1972, cf. array 17 on p. 69), using stroboscopic motion on a display otherwise qualitatively the same as that of Johansson's experiment 19. Subjects' percepts here seemed to reflect the independent motion of the dots rather than motion of a common frame of reference. A related case is that of Ternus-Pikler displays in which one of the moving disks contains a rotating dot. Here, vector decomposition occurs only at the high ISIs that are also necessary to perceive grouped disk motion (Boi et al., 2009).

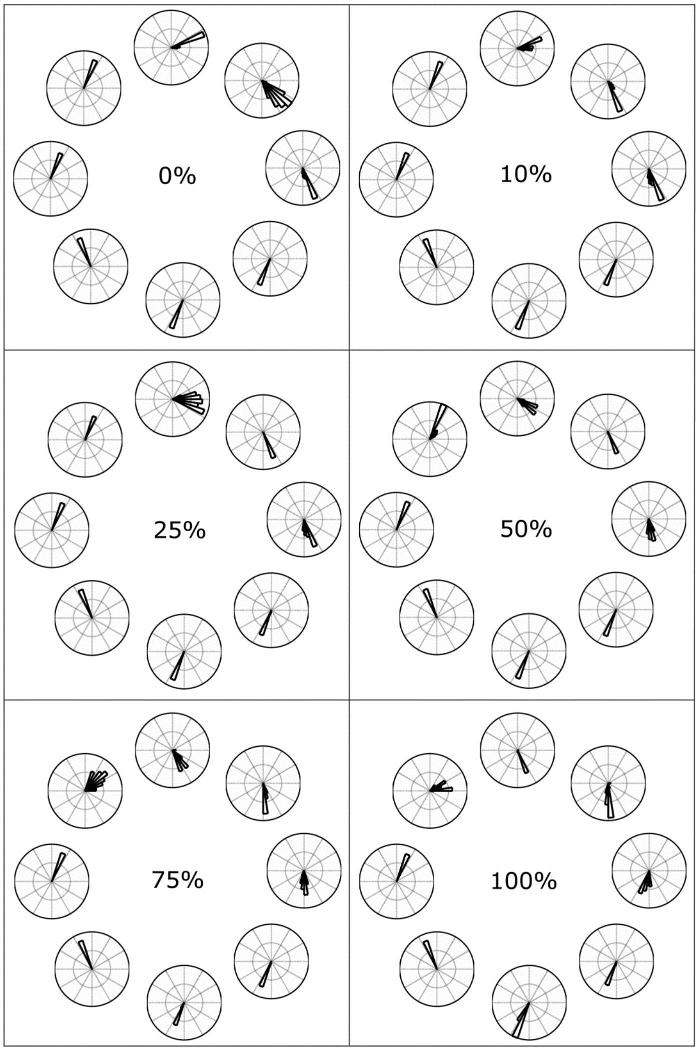

Figure 14. Simulations of the Duncker wheel stimulus, at various levels of boundary pruning. Each polar histogram shows the motion activity in the small (farther) scale over the cycloid dot at different phases of one rotation cycle. The presence of multiple bins in a given histogram denotes activation in multiple directions. For each wheel plotted, the amount of pruning completed is shown as a percentage.

As noted previously, the common motion direction is subtracted from part motion via near-to-far suppression in depth, which gives rise to a wheel-like percept over the cycloid dot, as the simulations of Figure 14 shows, using the same polar histogram representation as in Figure 12, and for various levels of pruning (indicated as percentages) of the farther V2 boundaries. Although these results could be improved with a finer sampling of the direction space, they are sufficient to demonstrate a predicted role of MSTv interactions in generating a peak shift in motion direction that leads to the observed vector decomposition.

In order to quantify the degradation of the percept as a function of the amount of boundary pruning, the motion directions obtained at each time-step over the cycloid dot were correlated with that of an ideal rotating wheel according to the measure ("s") defined in Appendix Eqs. 34-36. The measure is defined so as to be bounded in [-1, 1], where s = -1 corresponds to a wheel rotating in the opposite direction, and s = 1 corresponds to a perfectly represented wheel. Figure 15A shows the results obtained using this similarity measure for Duncker wheel simulations with increasing amounts of pruning completed. Figure 15B shows the result obtained for the simulation of the cycloid dot only, in which there is no boundary pruning. Comparing Figures 15A and 15B is sufficient to see that Duncker wheel simulations yielded more wheel-like activation in MST than the cycloid simulation, at all levels of boundary pruning.

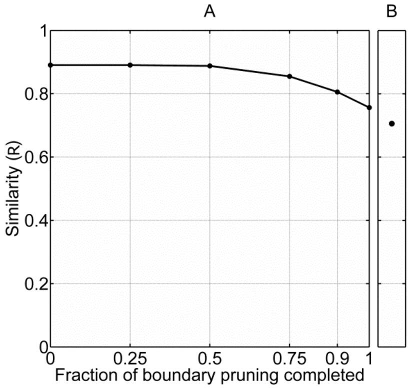

Figure 15. Effect of boundary pruning on MST activity evaluated using similarity measure *R*. Directional activity in MST perfectly consistent with wheel-like rotation in the rightward direction should yield *R* = 1, whereas non-rotational motion should lead to a smaller value of *R*. (A) As the percentage of pruning completed increases, MST activity observed in the Duncker wheel simulations becomes less wheel-like. Nevertheless, motion is always more circular than that observed in cycloid dot simulations (B).

Discussion

The 3D FORMOTION model predicts that the creation of an object-centric frame of reference is driven by interacting stages of the form and motion streams of visual cortex: form selection of motion-in-depth signals in area MT, and inter-depth directional inhibition in MST to cause a vector decomposition whereby the motion directions of a moving frame at a nearer depth suppress these directions at a farther depth and thereby cause a *peak shift* in the perceived directions of object parts moving with respect to the frame. In particular, motion signals predominant in the larger scale, or nearer depth, induce a peak shift of activity in smaller scales, or farther depths. The model qualitatively clarifies relative motion properties as manifestations of how the brain groups motion signals into percepts of moving objects, and quantitatively explains and simulates data about vector decomposition and relative frames of reference.

The model also qualitatively explains other data about frame-dependent motion coherence. Tadin, Lappin, Blake and Grossman (2002) presented observers with a display consisting of an illusory pentagon circularly translating behind fixed apertures, with each side of the pentagon defined by an oscillating Gabor patch. The location of the apertures and of the corners of the pentagon never overlapped, such that the latter were kept hidden during the entire stimulus presentation. Subjects had to judge the coherence of motion of the Gabor patches belonging to the different sides of the pentagon. Crucially, when the apertures were present, subjects reported seeing the patches as forming the shape of a pentagon, whereas when the apertures were absent, the patches did not seem to belong to the same shape. Results showed that motion coherence estimates were much better when apertures were present than not. According to the FACADE mechanisms in the form stream, the presence of apertures triggers the formation of illusory contours linking the contours of the Gabor patches into a single pentagon behind the apertures (see Berzhanskaya et al., 2007). Subsequent form selection and long-range filtering in MT leads to a representation of the pentagon's motion at a particular scale. This global motion direction is then subtracted from local motion signals of individual patches, thereby leading to better coherence judgments. In the absence of the apertures, form selection followed by long-range filtering of motion signals did not occur, such that the motion of individual patches mixed the common and part motion vectors, making coherence judgments difficult.

In displays where the speed of the moving reference frame and of a smaller moving target can be decoupled, the perceived amount of vector decomposition has been shown to be proportional to the speed of the frame (Gogel, 1979; Post, Chi, Heckmann and Chaderjian, 1989). This can be interpreted by noting that the firing rate of a MT cell in response to motion stimuli is proportional to the speed tuning of the cell (Raiguel et al., 1999). A frame of reference moving at a higher speed should therefore lead to higher cortical activation in the larger scales of MT and MST and thus to a more pronounced motion direction peak shift, reflecting the stronger percept of vector decomposition (Gogel, 1979; Post, Chi, Heckmann and Chaderjian, 1989). For the same reason, the model also predicts that the amount of shift in the perceived direction of the moving target is inversely proportional to target speed: a stronger peak in the motion direction distribution in the smaller scale (before subtraction) will be shifted less by subtraction from the large scale. Another prediction is that vector decomposition mechanisms occur mainly through MT-MST interactions.

The simulations shown here were conducted using a minimum number of scales to explain the experimental results. However, the model can be generalized to include a finer sampling of scale space, perhaps with depth suppression occurring as a transitive relation across scale. Such an arrangement of scales would then be able to account for experimental cases where vector decomposition must be applied in a hierarchical manner, such as in biological motion displays (Johansson, 1973). Accordingly, residual motion of the knee is obtained after subtraction of the common motion component of the hip and knee, whereas residual motion of the ankle is obtained after subtraction of the common motion component of the knee and ankle. Similar decompositions occur for upper limb parts. Such vector decompositions would require the use of spatial scales roughly matched to the lengths of the limbs with depth suppression occurring from larger scales coding for limb motion to smaller scale coding for joint motion.

The current model explains cases of vector analysis in which all dots in the stimulus display are moving, as opposed to some being static. The model would need to be refined to account for induced motion displays using an oscillating rectangle to induce an opposite perceived motion direction in a static dot (Duncker, 1938). The suggestion that additional mechanisms are needed to explain induced motion is supported by experimental evidence highlighting differences between induced motion and vector decomposition, as summarized by Di Vita and Rock (1997). For example, induced motion is typically not observed when the reference frame's physical speed is above the threshold for motion detection, whereas the vector decomposition stimuli analyzed here are robust to variations in speed. Also, in induced motion, the motion of the frame is underestimated or not perceived at all, whereas the common motion component in vector decomposition stimuli is perceived simultaneously to that of the parts.

Acknowledgement

This work was partially supported by CELEST, a NSF Science of Learning Center [SBE-0354378] and by the DARPA SyNAPSE program [HR0011-09-C-0001].

Appendix

All stages of the model were numerically integrated using Euler's method. All motion sequences are given to the network as series of static 2D frames representing black-and-white image snapshots at the consecutive moments of time (see next section). All model equations are membrane, or shunting, equations of the form:

$$C_m \frac{dX}{dt} = -[X - E_{leak}]g_{leak} - [X - E_{excit}]g_{excit} - [X - E_{inhib}]g_{inhib}. \tag{1}$$

(Grossberg, 1968; Hodgkin and Huxley, 1952). In this equation, g_{leak} is a leakage conductance, whereas g_{excit} and g_{inhib} represent excitatory and inhibitory inputs. Parameters E_{leak}, E_{excit}, and E_{inhib} are reversal potentials for leakage, excitatory and inhibitory conductances, respectively. All conductances contribute to the divisive normalization of the equilibrium membrane potential, X:

$$X = \frac{E_{leak}g_{leak} + E_{excit}g_{excit} + E_{inhib}g_{inhib}}{g_{leak} + g_{excit} + g_{inhib}}. \tag{2}$$

Reversal potentials in the following simulations were for simplicity set to E_{leak} =0, E_{excit} =1, and E_{inhib} = -1 unless noted otherwise. When the reversal potential of the inhibitory channel, E_{inhib}, is close to the resting potential, the inhibitory effect is pure "shunting"; i.e., it decreases the effect of excitation only through an increased membrane conductance. By abstracting away some of the details of the Hodgkin-Huxley neuron, the model in Eq.1 allows us to bridge the temporal gap between the dynamics of perception and of neuronal populations and networks in a parsimonious way. Although using the full range of Hodgkin-Huxley dynamics would likely require some model refinements in order to handle issues such as fast synchronization, recent work on converting rate into spiking neural networks has clarified that the network organizational principles and architecture remain the same, even as finer dynamical and structural details that are compatible with this architecture are revealed (Cao and Grossberg, 2010; Grossberg and Versace, 2008; Léveillé, Grossberg and Versace, 2010).

Depending on a layer's functionality, activities at each position (*i,j*) are represented as x_{ij}^{p}, where $p \in \{1,2\}$ indicates whether the cell (population) belongs to the ON or OFF stream; as x_{ij}^{d}, where $d \in \{0,...,7\}$ indicates directional preference within a single spatial scale; or else as x_{ij}^{ds} where $d \in \{0,...,7\}$ indicates motion directional preference, and $s \in \{1,2\}$ indicates spatial scale, with s = 1 indicating a farther scale (D2) and s = 2 a nearer scale (D1). The values used for all parameters are summarized in Tables 2 and 3.

Table 2. Spatial displacements.

Direction (*d*)	Δ_x^d	Δ_y^d
0	1	0
1	1	1
2	0	1
3	-1	1
4	-1	0
5	-1	-1
6	0	-1
7	1	-1

Table 3. Parameter values.

Level	Parameter	Value	Equation number
2	A_1	100	8
	B_1	5	8
	C_1	1	8
	A_2	0.01	9
	K_2	20	9
	A_3	50	10
	B_3	2	10
	C_3	10	10
	K_3	20	10
	A_4	50	12
	B_4	6	12
	C_4	10	12
	K_4	20	12
3	A_5	20	13
	σ_x^1	3	15
	σ_y^1	2	15
	σ_x^2	2	15
	σ_y^2	1	15
	G	5	15
	θ_1	0.0002	16
	θ_2	0.0001	16
4	A_6	5	17
	C_6	1	17
	D_6	10	17
	σ_x	2.5	18
	σ_y	0.5	18
	J	2	18
	σ_k	5.5	19
	K	2	19
5A	A_7	100	20
	K_E	1	20
	K_I	0.12	20
	K_B	10	20
	σ_R	6	21
	R	9	21
5B	A_8	50	22
	D_8	0.1	22
	α	2	22
	θ_n^1	0.001	24
	θ_n^2	0.001	24
	λ_x^1	50	25
	λ_y^1	35	25
	L^1	20	25
	λ_x^2	8	25
	λ_y^2	4	25
	L^2	20	25
	κ_1	20	26
	κ_2	8	26
	w^{Max}	2	27
6	A_9	80	28
	B_9	1	28
	C_9	6	28
	D_9	6.5	28
	V	1	29
	σ_V	1.2	29
	σ_O	10	30
	A	5	30
	Z	0.5	31
	σ_Z	2	31

All simulations were implemented in C++ on a dual 2Ghz AMD Opteron (AMD, 2003) workstation with 4Gb of RAM running Microsoft Windows XP x64 (Microsoft, 2003).

Convolution kernels separable along the horizontal and vertical axes (directions $d \in \{0, 2, 4, 6\}$) were implemented as one horizontal 1D convolution followed by one vertical 1D convolution in order to speed up computations (Haralick and Shapiro, 1992). Comparable speedups were obtained for non-separable kernels (directions $d \in \{1, 3, 5, 7\}$) by applying the convolution theorem with the FFTW library (Frigo and Johnson, 2005). Additional speedups were obtained by using OpenMP to assign convolutions at each model layer to different processors (Chapman et al., 2007). Computation time for one integration step was roughly 100ms for the Johansson (1950) stimuli (120x120 frames) and 1.2s for the rolling wheel experiment (170x350 frames).

Level 1: input

Inputs, J_{ij}^p, to the motion system are provided by 3-cell wide boundaries in separate ON and OFF channels, p = 1, 2. Oscillating dots are created by generating trajectories indexed by the position of a single point per shape for each time frame and then convolving the stimulus shape (a circle, square, or parallelogram) with the obtained frames. Input to the motion system is generated by subtracting the stimulus of the preceding time frame from the stimulus at the current time frame, and convolving the result with a 2x2 uniform mask in order to yield motion boundaries 3 cells wide, denoted by I_{ij}^p in Eq. 3. The convolved shapes are filled-in, positive values corresponding to inputs to the ON system. Negative values correspond to inputs to the OFF system. All obtained values are constrained to be composed of 1's or 0's only by computing:

$$J_{ij}^p = \begin{cases} 1 & if\ I_{ij}^p > 0\ and\ p = 1 \\ 1 & if\ I_{ij}^p < 0\ and\ p = 2 \\ 0 & otherwise \end{cases}. \tag{3}$$

Given the simplicity of experimental vector decomposition displays (all white boundaries on a dark background), the scheme used here to define motion inputs is sufficient to demonstrate key perceptual properties. The model's front-end could be further extended to process more natural scenes (e.g., as in Browning, Grossberg and Mingolla, 2009). For the Johansson (1950) stimuli, the trajectories of the dots are both rectilinear, one vertical and one horizontal. Figure 4 shows typical input to the motion stream generated with the above procedure. The position and direction of the dots at one particular time is indicated in Figure 4A. Corresponding ON and OFF inputs are displayed in B and C, respectively. For the rolling wheel stimulus, the trajectories of both the cycloid and hub dots are given by Eq. 4:

$$\begin{aligned} x &= a\varphi - b\sin\varphi \\ y &= a - b\cos\varphi \end{aligned}, \tag{4}$$

where φ represents scaled time or instantaneous phase, $a = 40$ is the radius of the wheel, and b is the distance between the peripheral dot and the center of the wheel. The trajectory of the dot on the spoke is obtained by setting $a = b$, whereas the trajectory of the central dot is obtained by setting $b = 0$. The equations above are computed for $\varphi \in [0, 2\pi]$, which corresponds to one revolution of the wheel. The resulting coordinates are rounded to the nearest integer (so that each value corresponds to a discrete pixel).

Input from V2 to the motion system (B_{ij}^s in Eq. 20) is provided by m-cell wide boundaries in separate depth planes, where m = 1 and 3 for the Johansson displays and the Duncker wheel, respectively. Using m = 3 in the Duncker wheel simulations was necessary to reduce spurious spatial aliasing which occurs when simulating a rotating stimulus in low resolution input frames

(170x350 pixels). The shape and strength of V2 boundaries is designed based on the following FACADE mechanisms (Grossberg, 1994). In nearer depth D1 (s = 2), bipole cells quickly group the collinear boundaries between the two dots and spatial competition within that depth inhibits the portions of the dot boundaries located within the emerging enclosing shape, thereby yielding a representation of the global shape of the object, shown in Figure 5A (cf. Grossberg and Mingolla, 1985; Grossberg and Raizada, 2000). At the same time, in farther depth D2 (s = 1), the smaller scale bipole cells group the boundaries of each dot individually, while newly emerged boundaries in the nearer depth start to inhibit the emerging boundaries in farther depth that are at the same position. Inhibited boundaries in the farther depth are shown in gray in Figure 5B. Since this temporally-extended process – termed boundary pruning – occurs as the stimulus is in motion, inhibition of the farther boundaries by the nearer ones may not be complete at a given time frame. There does not seem to be any psychophysical data available to indicate the proper amount of pruning that may occur at each time frame. Simulations were thus conducted assuming various amounts of V2 boundary pruning (specifically, 0, 25, 50, 75, 90 and 100% pruning complete). The amount of pruning did not affect the Johansson stimulus simulations, while it led to a graceful degradation of the Duncker wheel simulation (Figure 15).

For the Johansson (1950) stimuli, V2 input to the near plane is generated by convolving the trajectory points with half shape boundaries (instead of full shapes) and then linking the two half shapes with straight lines (Figure 5A). The use of half-shape boundaries removes those boundaries which would otherwise be contained in the interior region of the grouped stimulus of Figure 5A. V2 input to the far plane is generated by convolving the trajectory points with dot boundaries at the various amounts of pruning considered above (Figure 5B). In both cases, the value of a V2 boundary at a particular spatial location is set to 0, 0.1 0.25, 0.5, 0.75 or 1, depending on the amount of pruning.

For the rolling wheel stimulus, the rotating grouped boundary is generated for each time step (i.e., for each angle φ and global translation (t_x, t_y)) by applying the following affine transform to the coordinates of the pixels on the boundaries of an initially horizontal grouping, shown in Figure 16:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \tag{5}$$

In order to reduce aliasing and increase input strength, the resulting boundaries are filtered with a 2x2 uniform mask, and pixel values are clipped as for the other stimuli.

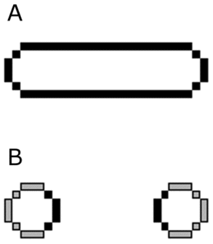

Figure 16. Base shapes used as V2 boundary input to MT. These shapes are made to follow the path described by a rolling wheel by applying the affine coordinate transform in Eq. 5. (A) Grouped boundaries. (B) Individual dot boundaries. Unlike the V2 boundaries defined for Johansson's experiments (Figure 5), which have a constant 45º orientation and where stimulus size changes as the dots move toward and away from each other, the orientation of the V2 boundaries here rotates due to Eq. 5, while stimulus size remains constant.

Spatial and temporal characteristics of the input were determined as follows. In all cases, each pixel in the simulations is assumed to represent roughly 1/10 of a degree of visual angle. For the Johansson (1950) experiments, the length of each dot's path is 34 pixels, that is, approximately 3.4 degrees of visual angle. The speed of the dots is set so as to take 5 second for one complete cycle of the stimulus. It was found in psychophysical experiments that these parameters yielded the desired effect. In comparison, in Johansson's experiment, the observer was placed 75cm away from the display, the dots had a diameter of 3mm and made a single 20mm-wide oscillation in approximately 1.5 second. This represents angular sizes much smaller than the ones simulated here, but he also reported that the effects were robust to variations on these parameters. The stimuli move by a distance of exactly 1 pixel (along their respective oscillatory axes) over successive input frames. The diameter of the dots is 7 pixels (<1 degree of visual angle). The size of each input frame is 120x120 pixels.

For the Duncker wheel stimulus, the length of the horizontal translation of the central dot is 251 pixels (25.1° of visual angle), the radius of the wheel is approximately 40 pixels (4°), the diameter of the dots is 13 pixels (1.3°), the spoke rotates 0.025 degrees per frame, and the wheel performs one revolution every five seconds. The size of the simulated display is 170x350 pixels.

Based on the settings above, the number of (Euler) integration steps performed on each frame is given by Eq. 6:

$$\#\,Euler\ steps = \frac{frame\ duration}{dt} = \frac{Cycle\ duration}{\#\,frames \cdot dt} = \frac{1}{Cycles/s \cdot \#\,frames \cdot dt} \qquad (6)$$

where, consistent with previous 3D FORMOTION simulations (Berzhanskaya et al., 2007), dt = 0.001. In the Johansson (1950) stimuli, the number of frames per cycle is 68 (34 towards the southwest corner, 34 towards the northeast corner). Since it takes 5 second for one cycle, the number of cycles per second is approximately 0.2. Thus, the number of Euler steps per frame for these simulations is $1/(0.2 \cdot 68 \cdot 0.001) \approx 74$. In the rolling wheel experiment, the number of frames is 252. Since it takes 5 seconds for one revolution, the number of cycles per second is 0.2. Thus the number of Euler steps per frame for these simulations is $1/(0.2 \cdot 252 \cdot 0.001) \approx 20$.

Level 2: transient cells

At the first stage of V1, non-directional transient cell activities b_{ij} are computed as a sum of ON (p = 1) and OFF (p = 2) channels:

$$b_{ij} = \sum_p x_{ij}^p z_{ij}^p\,, \qquad (7)$$

where input cell activities, x_{ij}^p, perform leaky integration on their inputs J_{ij}^p:

$$\frac{dx_{ij}^p}{dt} = A_1\left(-B_1 x_{ij}^p + (C_1 - x_{ij}^p) J_{ij}^p\right). \qquad (8)$$

Non-zero activation x_{ij}^p results in slow adaptation of a habituative presynaptic transmitter gate, or postsynaptic membrane sensitivity, z_{ij}^p:

$$\frac{dz_{ij}^p}{dt} = A_2(1 - z_{ij}^p - K_2 x_{ij}^p z_{ij}^p) \qquad (9)$$

(Abbott et al., 1997; Grossberg, 1972, 1980). In Eq. 8, $-A_1B_1x_{ij}^p$ is the rate of passive decay and C_1 is the maximum activity x_{ij}^p can reach. For non-zero inputs J_{ij}^p, x_{ij}^p approaches C_1 with a rate proportional to $(C_1 - x_{ij}^p)$ while it decays with the rate proportional to $-A_1B_1x_{ij}^p$. In Eq. 9, when a non-zero input x_{ij}^p is presented, z_{ij}^p is inactivated or habituates, at the rate $-A_2K_2x_{ij}^p z_{ij}^p$ as it tries to recover to 1 at rate A_2.

Input activity x_{ij}^p combined with transmitter gate z_{ij}^p results in transient non-directional cell activities b_{ij} that model activity of the non-directionally selective cells in layers 4Ca with circular receptive fields (Livingstone and Hubel, 1984). ON and OFF inputs summate at this stage. For visual inputs with a short dwell time, such as moving boundaries, activities b_{ij} respond well. A static input, on the other hand, produces only a weak response after an initial presentation period, because of the habituation (Muller, Metha, Krauskopf, and Lennie, 2001).

The next two cell layers provide a directional selectivity mechanism that can retain its sensitivity in response to variable speed inputs (Chey et al., 1997). As noted above, index *d* denotes the directional preference of a given cell. First, directional interneuron activities c_{ij}^d integrate transient cell inputs b_{ij}:

$$\frac{dc_{ij}^d}{dt} = A_3\left(-B_3c_{ij}^d + C_3b_{ij} - K_3\left[c_{XY}^D\right]^+\right). \tag{10}$$

A directional inhibitory interneuron c_{ij}^d receives excitatory input from a transient non-directional cell activity b_{ij} at the same position, and suppression from a directional interneuron c_{XY}^D of opposite direction preference *D* at the position (*X,Y*) offset by 1 cell in the direction *d*. For example, for the direction of motion 45°, *X* = *i*+*1*, *Y* = *j*+*1*, and *D* = 135°.

Activity c_{ij}^d increases proportionally to the transient input b_{ij} at rate A_3C_3 and passively decays to zero with rate $-A_3B_3c_{ij}^d$. The strength of opponent inhibition is $-A_3K_3\left[c_{XY}^D\right]^+$, where

$$[w]^+ = \max(w,0), \tag{11}$$

defines an output threshold. Inhibition is stronger than excitation (see Table 2) and "vetoes" a directional signal if the stimulus arrives from the null direction. Thus, a bar arriving from the left into the rightward directional interneuron receptive field would activate it well, while a bar arriving from the right would inhibit it even if activation b_{ij} is non-zero.

Directional transient cell activities e_{ij}^d at the next level combine transient input b_{ij} with inhibitory interneuron activity c_{ij}^d. Their dynamics are similar to those of c_{ij}^d:

$$\frac{de_{ij}^d}{dt} = A_4\left(-B_4e_{ij}^d + C_4b_{ij} - K_4\left[c_{XY}^D\right]^+\right). \tag{12}$$

As in Eq. 10, activity e_{ij}^d increases proportionally to transient input b_{ij}, passively decays at a fixed rate, and is inhibited by an inhibitory interneuron tuned to the opponent direction. Computation at Level 2 results in multiple directions activated in response to a moving line, which is consistent with the ambiguity caused by the aperture problem due to the limited size of V1 receptive fields.

Level 3: Short-range motion filter

Short-range filter activities, f_{ij}^{ds}, accumulate motion in each direction *d:*

$$\frac{df_{ij}^{ds}}{dt} = A_5\left(-f_{ij}^{ds} + \sum_{XY} E_{XY}^{d} G_{ijXY}^{ds}\right). \tag{13}$$

In Eq. 13,

$$E_{XY}^{d} = \left[e_{xy}^{d}\right]^{+} \tag{14}$$

is the rectified output of the directional transient cell e_{xy}^{d} from Level 2, and G_{ijXY}^{ds} is a Gaussian receptive field that depends both on direction *d* and scale *s*:

$$G_{ijXY}^{ds} = G\exp\left(-0.5\left(\left(\frac{x-i}{\sigma_x^s}\right)^2 + \left(\frac{y-i}{\sigma_y^s}\right)^2\right)\right). \tag{15}$$

Kernel G_{ijXY}^{ds} is elongated in the direction of motion. Scale *s* determines receptive field size, and therefore the extent of spatiotemporal integration of lower-level motion signals. Larger receptive fields respond selectively to larger speeds, smaller receptive fields to smaller speeds; cf., Chey et al. (1998). While in our simulations speed did not vary much, in more motion-rich environments speed-depth correlations can help to assign an approximate depth order to the moving objects. Output of the short-range filter is thresholded and rectified according to Eq. 16,

$$F_{ij}^{ds} = \left[f_{ij}^{ds} - \theta_s\right]^{+}, \tag{16}$$

with thresholds $\theta_s = \theta s$, where s = 1, 2. The thresholds are thus scale-specific. If they were the same, the larger scale would always activate more strongly. With a larger threshold, the larger scale prefers larger speeds. See Chey et al. (1998).

Level 4: Spatial competition and opponent direction inhibition

The spatial competition and opponent direction inhibition activities, h_{ij}^{ds}, are determined according to the following membrane equation:

$$\frac{dh_{ij}^{ds}}{dt} = A_6\left(-h_{ij}^{ds} + (1-h_{ij}^{ds})\sum_{XY} F_{XY}^{ds} J_{ijXY}^{ds} - (0.1+h_{ij}^{ds})\left[C_6 \sum_{XY} F_{XY}^{ds} K_{ijXY}^{ds} + D_6 F_{ij}^{Ds}\right]\right), \tag{17}$$

where F_{XY}^{ds} is the output (see Eq. 15) of a Level 3 cell at spatial location *XY*, direction *d*, and scale *s*. Eq. 17 defines a spatial competition within one motion direction *d* with inhibition from the opponent motion direction *D* at the same location. The on-center kernel J_{ijXY}^{ds} of the spatial competition is elongated in the direction of motion:

$$J_{ijXY}^{ds} = \frac{J}{2\pi\sigma_x\sigma_y}\exp\left(-0.5\left(\left(\frac{x-i}{\sigma_x}\right)^2 + \left(\frac{y-i}{\sigma_y}\right)^2\right)\right), \tag{18}$$

whereas the off-surround K_{ijXY}^{ds} is spatially isotropic:

$$K_{ijXY}^{ds} = \frac{K}{2\pi\sigma_k^2}\exp\left(-0.5\left(\frac{(x-(i-\Delta_x^d))^2 + (y-(j-\Delta_y^d))^2}{\sigma_k^2}\right)\right). \tag{19}$$

The center of the inhibitory kernel K_{ijXY}^{ds} is offset from the (*i,j*) position by one cell in the direction opposite to the preferred direction *d*, as determined by Δ_x^d and Δ_y^d (see Table 3). This arrangement results in a on-center off-surround recurrent spatial competition network wherein inhibition trails excitation. Signal F_{ij}^{Ds} is the output of a Level 3 cell at spatial location *ij*, and opponent direction $D = d + \pi$. The strength of spatial competition is determined by parameter C_6, and that of opponent inhibition by D_6.

Level 5: Formotion capture and long-range filter

Rectified motion output signals, $H_{ij}^{ds} = \left[h_{ij}^{ds}\right]$, from V1 (model Level 4) are selected by form boundary signals, B_{ij}^s, from V2 in the input layers 4 and 6 of MT. The activities, q_{ij}^{ds}, of these MT cells combine motion and boundary signals via a membrane equation:

$$\frac{dq_{ij}^{ds}}{dt} = A_7\left(-q_{ij}^{ds} + (1-q_{ij}^{ds})H_{ij}^{ds}(K_E + K_B B_{ij}^s) - K_I(1+q_{ij}^{ds})\sum_{XY} B_{XY}^s R_{ijXY}^s \right). \tag{20}$$

In Eq. 20, input from the V1 motion stream $H_{ij}^{ds}K_E$ is positively modulated by boundaries $K_B B_{ij}^s$ in the excitatory term of the equation. Activity B_{ij}^s in Eq. 20 codes an idealized m-cell wide boundary that simulates output from V2. In the case of the Johansson (1950) stimuli, m = 1. In the rolling wheel experiment, m = 3 to reduce aliasing effects due to the rotation of the wheel on the discrete input grid. In addition, these boundaries inhibit unmatched motion signals via term $\sum_{XY} B_{XY}^s R_{ijXY}^s$. This modulatory on-center, off-surround network allows boundaries to select motion signals at their positions and depths. Parameter K_E determines the strength of feedforward inputs H_{ij}^{ds}, and K_B the strength of modulation by V2 boundaries. The V2 boundary projection to MT is stronger than the bottom-up motion projection; that is, $K_E << K_B$. The strength of V2 boundary inhibition $\sum_{XY} B_{XY}^s R_{ijXY}^s$ is scaled by the coefficient K_I, and its spatial range is determined by inhibitory Gaussian kernel R_{ijXY}^s:

$$R_{ijXY}^s = \frac{R}{2\pi\sigma_I^2}\exp\left(-0.5\left(\frac{(x-i)^2 + (y-j)^2}{\sigma_R^2} \right)\right). \tag{21}$$

The modulatory on-center and driving off-surround in Eq. 20 could be implemented in the brain in various ways after direct excitatory inputs from V2-to-MT are registered in MT. We interpret this network to be built up in much the same way as seems to occur in primary visual cortex; namely, with a layer 4 on-center and inhibitory interneurons from cortical layer 6 to 4 (Ahmed et al., 1997; Callaway, 1998; McGuire et al., 1984; Stratford et al., 1996). When no boundary is provided and $B_{ij}^s = 0$ everywhere (for example, the parvocellular stream is inactivated), motion signals can still activate MT via the term $H_{ij}^{ds}K_e$ in Eq. 19. In this case, no inhibition is present as well. In the presence of boundary input, motion signals at boundary positions are strong, whereas those outside of the boundary position are suppressed.

Next, model MT cell activities, m_{ij}^{ds}, in layer 2/3 receive MT signals, N_{ij}^{ds}, from layer 4 via a long-range filter and top-down matching signals, T_{ij}^{ds}, from MST:

$$\frac{dm_{ij}^{ds}}{dt} = A_8\left(-m_{ij}^{ds} + (1-m_{ij}^{ds})N_{ij}^{ds}(1+\alpha\left[T_{ij}^{ds}\right]^+) - D_8(1+m_{ij}^{ds})\sum_{e,XY} w^{de}\left[T_{XY}^{es}\right]^+ P_{ijXY}^{s}\right). \tag{22}$$

To compute the long-range filter inputs, N_{ij}^{ds}, the MT input activities, q_{ij}^{ds}, are rectified:

$$Q_{ij}^{ds} = \left[q_{ij}^{ds}\right]^+, \tag{23}$$

and squared to generate output signals before being anisotropically filtered by a long-range filter L_{ijXY}^{ds}, thresholded, and rectified again:

$$N_{ij}^{ds} = \left[\sum_{XY}\left(Q_{XY}^{ds}\right)^2 L_{ijXY}^{ds} - \theta_n^s\right]^+. \tag{24}$$

In Eq. 24, the long-range filter L_{ijXY}^{ds} is defined by an anisotropic Gaussian kernel:

$$L_{ijXY}^{ds} = \frac{L^s}{2\pi\lambda_x^s\lambda_y^s}\exp\left(-0.5\left(\left(\frac{x-i}{\lambda_x^s}\right)^2 + \left(\frac{y-i}{\lambda_y^s}\right)^2\right)\right) \tag{25}$$

that is elongated in the direction of preferred motion. This long-range filter accumulates evidence for motion in its preferred direction over time and space. The anatomical basis for such integration can be provided by long-range horizontal projections in layers 2/3 of MT. The squaring operation gives higher preference to larger signals, which leads to winner-take-all dynamics in competitive recurrent networks (Grossberg, 1973, 1980, 1988).

Due to the locality of the winner-take-all dynamics, multiple directions of motion in different spatial positions and depth planes can, in principle, be simultaneously represented in MT and further projected to MST. However, the evidence that is accumulated at one position may be similar to that accumulated at nearby positions, leading to the same winner at these positions. The long-range filter is not, however, sufficient to realize the kind of motion capture that can solve the aperture problem and impart a global perceived motion direction on an entire object. This is accomplished by positive feedback between the long-range grouping process in MT and the directional grouping process in MST. This combination of properties has elsewhere been shown capable of simulating properties of motion transparency at different depths (Berzhanskaya et al., 2007).

As in the case of the V2-to-MT projection, MST-to-MT feedback is defined by a modulatory on-center, off-surround network. Excitatory feedback $\alpha\left[T_{ij}^{ds}\right]^+$ in Eq. 22 from MST (See Eq. 28) is modulatory in nature and its strength is determined by coefficient α. Thus, top-down input T_{ij}^{ds} is only effective if bottom-up input N_{ij}^{ds} is positive. The strength of MST off-surround feedback $\sum_{e,XY} w^{de}\left[T_{XY}^{es}\right]^+ P_{ijXY}^{s}$ is determined by coefficient D_8. The spatial extent of the off-surround is determined by the isotropic kernel P_{ijXY}^{s}:

$$P_{ijXY}^{s} = \frac{1}{2\pi\kappa_s^2}\exp\left(-0.5\left(\frac{(x-i)^2+(y-j)^2}{\kappa_s^2}\right)\right). \tag{26}$$

Off-surround inhibition is from all directions except *d*. This is controlled by the inhibitory weight w^{de} between a given direction *d* and another direction *e*:

$$w^{de} = w^{Max} \frac{|d-e|}{\pi}, \tag{27}$$

where d and $e \in \{-3\pi/4, -\pi/2, ..., \pi\}$ denote the direction preferences of the cells. The kernel in Eq. 27 is maximal when *d* and *e* are of opposite direction, and zero when *d* = *e*. Because excitatory input N^{ds}_{ij} is from the preferred direction, this directionally asymmetric suppression effectively amplifies *d* and suppresses other motion directions. Although various neurophysiological studies are consistent with directionally-selective receptive fields in MT (e.g., Livingstone, Pack, and Born, 2001; Xiao, Marcar, Raiguel, and Orban, 1997; Xiao, Raiguel, Marcar, and Orban, 1997), we are not aware of direct anatomical data concerning the validity the synaptic kernel defined in Eq.27. Such an inhibitory sharpening mechanism is compatible with reports that blockage of GABAergic transmission in area MT weakens direction selectivity (Thiele, Distler, Korbmacher and Hoffmann, 2004). Motion from unambiguous feature tracking signals propagates to ambiguous motion positions through the large kernel P^{s}_{ijXY}.

Level 6: Directional grouping and suppression in depth

The MT-MST directional grouping circuit acts in a winner-take-all mode, selecting a single direction of motion at each point. MST activity T^{ds}_{ij} is described by

$$\frac{dT^{ds}_{ij}}{dt} = A_9\left(-T^{ds}_{ij} + (1-T^{ds}_{ij})\sum_{e} v^{de} M^{es}_{ij}(1+O^{ds}_{ij}) - (B_9 + T^{ds}_{ij})\left[D_9 \sum_{e,XY} w^{de} \left[T^{es}_{XY}\right]^+ P^{s}_{ijXY} + C_9 \sum_{s<S,e} z^{de} \left[T^{eS}_{ij}\right]^+ \right] \right). \tag{28}$$

The activity T^{ds}_{ij} decays at rate $-A_9$. Bottom-up input $M^{es}_{ij} = \left[m^{es}_{ij}\right]^+$ is the rectified MT output. A Gaussian kernel v^{de} determines the magnitude of input from different directions:

$$v^{de} = V \exp\left(-0.5\left(\frac{(e-d)^2}{\sigma_v^2} \right) \right). \tag{29}$$

Bottom-up excitation is modulated by attention via term O^{ds}_{ij}. Such a modulatory term has been shown to be able to account for the effect of spatial attention on the activity of direction-selective neurons in area MT (cf. Eq. 2 in Womelsdorf, Anton-Erxleben and Treue, 2008). If attention focuses on features in the near depth plane, then this modulation would help one motion direction to win in the near depth. The suggestion that attention directed to a particular direction of motion may enhance the activity of cells selective for that motion direction is corroborated by physiological data in both MT and MST (Treue and Martínez Trujillo, 1999; Treue and Maunsell, 1996).

Attention was used only in simulations of Figures 8, 9 and 10 to show that attention directed to the dominant direction of motion of the grouped stimulus can bias the vector decomposition observed over the stimulus parts. Attention was applied as a single Gaussian "spot" in the near depth (*s* = 1) and along the southwest-northeast diagonal axis for the simulations of Figures 8 and 10 (*d* = 5 or 1, depending on the current direction of the grouped stimulus), and along the horizontal axis for the simulation of Figure 9 (*d* = 0 or 4, depending on the current direction of the tracked dot):

$$O_{ij}^{ds} = A\exp\left(-0.5\left(\frac{(x_0-i)^2+(y_0-j)^2}{\sigma_O^2}\right)\right). \tag{30}$$

Here x_0 and y_0 are the coordinates of the center of the attentional spotlight and are designed to follow the middle of the grouped stimulus for the simulations of Figures 8 and 10 or of one particular dot for the simulation of Fig. 9. This bias is similar to the one used in the case of transparent motion in Grossberg et al. (2001) and Berzhanskaya et al. (2007) and allows a single motion signal to win in the near depth D1.

Inhibition in Eq. 28 takes the form of directional competition and suppression in depth. All inhibitory terms are gated by shunting term $-(B_9+T_{ij}^{ds})$, where $B_9 > 0$. Directional competition is implemented by recurrent connections within MST in the term $D_9\sum_{e,XY} w^{de}\left[T_{XY}^{es}\right]P_{ijXY}^{s}$. Its strength is determined by coefficient D_9, and its spatial extent by the kernel P_{ijXY}^{s}, where XY and ij represent the spatial locations of the presynaptic and postsynaptic cells, respectively, and s is the scale. The weighting coefficient w^{de} and surround suppression kernel P_{ijXY}^{s} are the same as in Eqs. 26 and 27. MST also includes direction-specific suppression, $C_9\sum_{s<S,e} z^{de}\left[T_{ij}^{eS}\right]$, from the near depth (D1, S = 2) to the far depth (D2, s = 1), which is important for the proposed mechanism of vector decomposition. Kernel z^{de} determines the magnitude of depth suppression across directions and is computed as:

$$z^{de} = Z\exp\left(-0.5\left(\frac{(e-d)^2}{\sigma_z^2}\right)\right). \tag{31}$$

If the motion in the direction d wins in D1, this direction will be suppressed in D2. This allows the model to avoid a single motion direction being represented in both depths. In the case of transparent motion, suppression of one direction in D2 would allow another direction to win there. The kernel in Eq. 31 also implies that suppression from the larger (nearer) scale to the smaller (farther) scale is strongest for the same direction $e = d$, and weakest for opposite directions $e = d+\pi$. This prediction is consistent with experimental data in which lesions to cortical areas including MT and MST resulted in weaker activation of superior colliculus neurons – which receive feedback form MT – to a small target when it was moving in the same direction as a textured background but not when it was moving in the opposite direction (Joly and Bender, 1997).

Vector summation

The output of MST cells (Level 6) is displayed as a vector summation according to the following equation:

$$\vec{v}_{ij}^{s} = \sum_d T_{ij}^{ds}u^d , \tag{32}$$

where T_{ij}^{ds} is a scalar representing the activity of the MST cell at location ij, and direction d. The variable u^d is a unit vector representing direction d. For example, for the eastward direction, $u^E = (1, 0)$ and for the northeast direction, $u^{NE} = (\sqrt{2}/2, \sqrt{2}/2)$.

Similarity estimate for Duncker wheel

In order to calculate the influence of pruning on the path of the cycloid (Figure 15), a similarity estimate was defined as follows. Using Eq. 32, let $\vec{v}^{1}_{c_x(t)c_y(t)}$ be the 2D vector representing the velocity of the MST cell in scale s = 1 whose coordinates are located at the center (*c_x(t)*, *c_y(t)*) of the cycloid dot at time step *t*. Furthermore, let $v(t) = (v_x(t), v_y(t))$ be the orthogonal projection of $\vec{v}^{1}_{c_x(t)c_y(t)}$ on the x- and y-axis, respectively. These components are compared to the theoretically-derived velocity components for a perfectly represented wheel. The latter is defined as the derivative of Eq. 4 from which common motion is subtracted:

$$v^T(t) = \begin{bmatrix} v^T_x(t) \\ v^T_y(t) \end{bmatrix} = \begin{bmatrix} -b\cos t \\ \sin t \end{bmatrix} \tag{33}$$

The difference between $v(t)$ and $v^T(t)$ is calculated as a normalized inner product:

$$r(t) = \frac{v^T(t) \cdot v(t)}{\left\| v^T(t) \right\| \left\| v(t) \right\|}, \tag{34}$$

which takes a value of 1 if the two vectors are perfectly aligned, and -1 if they are of opposite orientations. The similarity measure is given by integrating across all time frames and dividing by the number of frames:

$$R = \frac{1}{N_t} \sum_t r(t), \tag{35}$$

where N$_t$ is the number of time frames. It follows that $R \in [-1,1]$, where *R* = 1 indicates a perfectly represented wheel and where *R* = -1 indicates wheels rolling in opposite directions. In order to ensure that *r(t)* is always well-defined in Eq. 35, it is set to 0 when $\vec{v}^{6}_{c_x(t)c_y(t)} = 0$, which occurs in the first few time-frames of the wheel when the cycloid dot has not accumulated enough motion activity. Note that this does not bias the estimate *R* in any direction.

References

Abbott, L.F., Sen, K., Varela, J.A., and Nelson, S.B. (1997). Synaptic depression and cortical gain control. *Science*, 275, 220-222.

Ahmed, B., Anderson, J. C., Martin, K. A. C., & Nelson, J. C. (1997). Map of the synapses onto layer 4 basket cells of the primary visual cortex of the cat. *Journal of Comparative Neurology*, 380, 230–242.

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., 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.

AMD. (2003). AMD. *Sunnyvale, CA 94088*.

Amano, K., Edwards, M., Badcock, D.R. and Nishida, S. (2009). Adaptive pooling of visual motion signals by the human visual system revealed with a novel multi-elements stimulus. *Journal of Vision*, 9, 1-25.

Anderson, J.C., Binzegger, T., Martin, K.A. and Rockland, K.S. (1998). The connection from cortical area V1 to V5: a light and electron microscopy study. *Journal of Neuroscience*, 18, 10525-10540.

Anderson, J.C. and Martin, K.A. (2002). Connection from cortical area V2 to MT in macaque monkey. *Journal of Comparative Neurology*, 443(1), 56-70.

Baloch, A.A. and Grossberg, S. (1997). A neural model of high-level motion processing: Line motion and formotion dynamics. *Vision Research*, 37, 3037-3059.

Barlow, H.B. and Levick, W.R. (1965). The mechanism of directionally selective units in rabbit's retina*.* *Journal of Physiology*, 178, 477-504.

Berzhanskaya, J., Grossberg, S. and Mingolla, E. (2007). Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spatial Vision*, 20(4), 337-395.

Blasdel G.G. and Lund J.S. (1983). Termination of afferent axons in macaque striate cortex. *Journal of Neuroscience*, 3,1389–1413.

Boi, M., Öğmen, H., Krummenacher, J., Otto, T.U. and Herzog, M.H. A (fascinating) litmus test for human retino- vs. non-retinotopic processing. *Journal of Vision*, 9, 1-11.

Börjesson, E. and von Hofsten, C. (1972). Spatial determinants of depth perception in two-dot motion patterns. *Perception and Psychophysics*, 11, 263-268.

Börjesson, E. and von Hofsten, C. (1973). Visual perception of motion in depth: Application of a vector model to three-dot motion patterns. *Perception and Psychophysics*, 2, 169-179.

Börjesson, E. and von Hofsten, C. (1975). A Vector model for perceived object rotation and translation in space. *Psychological Research*, 38, 209-230.

Börjesson, E. and von Hofsten, C. (1977). Effects of different motion characteristics on perceived motion in depth. *Scandinavian Journal of Psychology*, 18, 203-208.

Born, R.T. and Tootell, R.B.H. (1992). Segregation of global and local motion processing in primate middle temporal visual area. *Nature*, 357, 497-499.

Bremner, A.J., Bryant, P.E. and Mareschal, D. (2005). Object-centred spatial reference in 4-month-old infants. *Infant Behaviour and Development*, 29, 1-10.

Browning, N.A., Grossberg, S. and Mingolla, E. (2009). Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance. *Neural Networks*, 22, 1383-1398.

Cai D., DeAngelis G.C. and Freeman R.D. (1997). Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *Journal of Neurophysiology*, 78,1045-61.

Callaway, E. M. (1998). Local cicrcuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, 21, 47–74.

Cao, Y. and Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, 18, 515-578.

Cao, Y. and Grossberg, S. (2010). Stereopsis and 3D surface perception by spiking neurons in laminar cortical circuits: A method for converting neural rate models into spiking models. *Neural Networks*, submitted for publication.

Carandini, M. and Ferster, D. (1997). Visual adaptation hyperpolarizes cells of the cat striate cortex. *Science*, 276, 949.

Chance, F.S., Nelson., S.B., and Abbott, L.F. (1998). Synaptic depression and the temporal response characteristics of V1 cells. *Journal of Neuroscience*, 18(12), 4785-4799.

Chapman, B., Jost, G. and van der Pas, R. (2007). Using OpenMP. Portable shared memory parallel programming. Boston: MIT Press.

Chey, J., Grossberg, S. and Mingolla, E. (1998). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America*, 14, 2570-2594.

Chey, J., Grossberg, S., and Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research,* 38, 2769-2786.

Cutting, J.E. and Proffitt, D.R. (1982). The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology*, 14, 211-246.

De Valois, R., Cottaris, N.P., Mahon, L.E., Elfar, S.D. and Wilson, J.A. (2000). Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision Research*, 40, 3685-3702,

Di Vita, J.C. and Rock, I. (1997). A belongingness principle of motion perception. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1343-1352.

Duncker, K. (1938). Induced motion. In W.D. Ellis (Ed.), *A sourcebook of Gestalt psychology*. London: Routledge & Kegan Paul, 1938. (Originally published in German, 1929).

Eifuku, S. and Wurtz, R.H. (1999). Response to motion in extrastriate area MSTl: Disparity sensitivity. *Journal of Neurophysiology*, 82, 2462-2475.

Enroth-Cugell, C. and Robson, J. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London),* 187, 517-552.

Fang, L., and Grossberg, S. (2009). From stereogram to surface: How the brain sees the world in depth. *Spatial Vision*, 22, 45-82.

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, 149-173.

Francis, G., Grossberg, S. and Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research*, 34, 1089-1104.

Fried, S.I., Münch, T.A. and Werblin, F.S. (2002). Mechanisms and circuitry underlying directional selectivity in the retina. *Nature*, 420, 411-414.

Frigo, M. and Johnson, S.G. (2005). The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2), 216-231.

Gattass, R., Sousa, A.P.B., Mishkin, M. and Ungerleider, L.G. (1997). Cortical projections of area V2 in the macaque. *Cerebral Cortex*, 7, 110-129.

Gogel, W.C. (1979). Induced motion as a function of the speed of the inducing object, measured by means of two methods. *Perception*, 8(2), 255-262.

Gogel, W.C. and Mac Cracken, P.J. (1979). Depth adjacency and induced motion. *Perceptual and Motor Skills*, 48, 343-350.

Gogel, W.C. and Tietz, J.D. (1976). Adjacency and attention as determiners of perceived motion. *Vision Research*, 16, 839-845.

Grossberg, S. (1968). Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Sciences*, 60, 758-765.

Grossberg, S. (1972). A neural theory of punishment and avoidance, II: Quantitative theory. *Mathematical Biosciences*, 15, 253-285.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 213-257.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87(1), 1-51.

Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, 1, 12-61.

Grossberg, S. (1991). Why do parallel cortical systems exist for the perception of static form and moving form? *Perception and Psychophysics*, 49, 117-141.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, 55, 48-121.

Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two dimensional pictures. *Psychological Review*, 104, 618-658.

Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision,* 12**,** 163-185.

Grossberg, S. (2000). The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Science*, 4(6), 233-46.

Grossberg, S. and Kelly, F. (1999). Neural dynamics of binocular brightness perception. *Vision Research*, 39, 3796-3816.

Grossberg, S. and Levine, D. (1976). On visual illusions in neural networks: Line neutralization, tilt aftereffect, and angle expansion. *Journal of Theoretical Biology*, 61, 477-504.

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 perceptual grouping: textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, 38, 141-171.

Grossberg, S., Mingolla, E., and Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, 41, 2351-2553.

Grossberg, S. and Pessoa, L. (1998). Texture segregation, surface representation and figure-ground separation. *Vision Research*, 38, 2657-2684.

Grossberg, S. and Pilly, P. (2008). Temporal dynamics of decision-making during motion perception in the visual cortex. *Vision Research*, 48, 1345-1373.

Grossberg, S. and Raizada, R.D. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40(10-12),1413-32.

Grossberg, S. and Rudd, M. (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.

Grossberg, S. and Swaminathan, G. (2004). A laminar cortical model of 3D perception of slanted and curved surfaces and of 2D images: development, attention and bistability. *Vision Research*, 44, 1147-1187.

Grossberg, S. and Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Research*, 1218, 278-312.

Grossberg, S. and Yazdanbaksh, A. (2005). Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vision Research*, 45, 1725-1743.

Haralick, R.M. and Shapiro, L.G. (1992). Computer and robot vision (vol.1). Boston: Addison-Wesley Longman Publishing Co., Inc.

Hershenson, M. (1999). Visual Space Perception. Cambridge: MIT Press.

Hirsch, J. A. and Gilbert, C. D. (1991). Synaptic physiology of horizontal connections in the cat's visual cortex. *Journal of Neuroscience*, 11, 1800-1809.

Hochstein, S. and Shapley, R.M. (1976a). Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *Journal of Physiology*, 262(2), 265-84.

Hochstein, S. and Shapley R.M. (1976b). Quantitative analysis of retinal ganglion cell classifications. *Journal of Physiology*, 262(2), 237-64.

Hodgkin, A.L., & Huxley, A.F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology, 117*, 500-544.

Johansson, G. (1950). *Configurations in event perception*. Uppsala: Almqvist and Wiksell.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14, 201-211.

Johansson, G. (1974). Vector analysis in visual perception of rolling motion. *Psychologische Forschung*, 36, 311-319.

Johansson, G. (1985). Vector analysis and process combinations in motion perception: a reply to Wallach, Becklen and Nitzberg (1985). *Journal of Experimental Psychology: Human Perception and Performance*, 11(3), 367-371.

Joly, T.J. and Bender, D.B. (1997). Loss of relative-motion sensitivity in the monkey superior colliculus after lesions of cortical area MT. *Experimental Brain Research*, 117, 43-58.

Kelly, F. and Grossberg, S. (2001). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception & Psychophysics*, 62(18), 1596-1618.

Kolers, P.A. (1972). *Aspects of motion perception*. Oxford: Pergamon, 1972.

Léveillé, J., Versace, M. and Grossberg, S. (2010). Running as fast as it can: How spiking dynamics form object groupings in the laminar circuits of visual cortex. *Journal of Computational Neuroscience*, 28, 323-346.

Livingstone, M.S. and Conway, B.R. (2003). Substructure of direction-selective receptive fields in macaque V1. *Journal of Neurophysiology*, 89, 2743-2759.

Livingstone, M.S. and Hubel, D.H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, 4(1), 309-356.

Livingstone, M.S., Pack, C.C. and Born, R.T. (2001). Two-dimensional substructure of MT receptive fields. *Neuron*, 30, 781-793.

Livingstone, M.S. (1998). Mechanisms of direction selectivity in macaque V1. *Neuron*, 20, 509-526.

Lorenceau, J. and Alais, D. (2001). Form constraints in motion binding. *Nature Neuroscience*, 4, 745-751.

Maunsell, J.H. and van Essen, D.C. (1983a). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience*, 3, 2563-2586.

McGuire, B. A., Hornung, J. P., Gilbert, C. D., & Wiesel, T. N. (1984). Patterns of synaptic input to layer 4 of cat striate cortex. *Journal of Neuroscience*, 4(12), 3021–3033.

Microsoft. (2003). Microsoft Corporation. *Redmond, WA 98052*.

Muller, J.R., Metha, A.B., Krauskopf, J. and Lennie, P. (2001) Information conveyed by onset transients in responses of striate cortical neurons. *Journal of Neuroscience*. 21(17), 6987-6990.

Murthy, A. and Humphrey, A.L. (1999). Inhibitory contributions to spatiotemporal receptive-field structure and direction selectivity in simple cells of cat area 17. *Journal of Neurophysiology*, 81, 1212-1224.

Pack, C.C., Gartland, A.J., and Born, R.T. (2004). Integration of contour and terminator signals in visual area MT of alert macaque. *Journal of Neuroscience,* 24(13), 3268-80.

Ponce, C. R., Lomber, S. G. and Born, R. T. (2008) Integrating motion and depth via parallel pathways. *Nature Neuroscience*, 11(2), 216-23.

Post, R.B., Chi, D., Heckmann, T. and Chaderjian, M. (1989). A reevaluation of the effect of velocity on induced motion. *Perception and Psychophysics*, 45 (4), 411-16.

Raiguel, S.E., Xiao, D.-K., Marcar, V.L., and Orban, G.A. (1999). Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. *Journal of Neurophysiology*, 82, 1944-1956.

Raizada RD. and Grossberg S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex,* 13(1), 100-13.

Ramachandran, V., S., and Inada, V. (1985). Spatial phase and frequency in motion capture of random-dot patterns. *Spatial Vision*, 1(1), 57-67.

Reichardt W. in *Sensory Communication*, W.A. Rosenblith, ed. (Wiley, New York, 1961), p303.

Rock, I. The frame of reference. In *The legacy of Solomon Asch*, I. Rock, Ed. (Lawrence Erlbaum Associates, Hillsdale, NJ, 1990), p.243-268.

Rockland, K.S. (1995). Morphology of individual axons projecting from area V2 to MT in the macaque. *Journal of Comparative Neurology*, 355(1), 15-26.

Rubin, J. and Richards, W.A. (1988). Visual perception of moving parts. *Journal of the Optical Society of America*, 5(12), 2045-2049.

Rust, N.C., Mante, V., Simoncelli, E.P. and Movshon, J.A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, 9(11), 1421-31.

Sedgwick, H.A. (1983). Environment-centered representation of spatial layout: Available visual information from texture and perspective. In J. Beck, B. Hope & A. Rosenfeld (eds.), *Human and Machine Vision*. Elsevier.

Shimojo, S., Silverman, G.H. and Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29(5), 619-26.

Smith, M.A., Majaj, N.J. and Movshon, J.A. (2005). Dynamics of motion signaling by neurons in macaque area MT. *Nature Neuroscience*, 8, 220-228.

Sokolov, A., and Pavlova, M. (2006). Visual motion detection in hierarchical spatial frames of reference. *Experimental Brain Research*, 174, 477-486.

Stratford, K. J., Tarczy-Hornoch, K., Martin, K. A. C., Bannister, N. J., & Jack, J. J. B. (1996). Excitatory synaptic inputs to spiny stellate cells in cat visual cortex. *Nature*, 382, 258–261.

Tadin, D., Lappin, J.S., Blake, R. and Grossman, E.D. (2002). What constitutes an efficient reference frame for vision? *Nature Neuroscience*, 5(10), 1010-1015.

Tanaka, K., Sugita, Y., Moriya, M. and Saito, H. (1993). Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *Journal of Neurophysiology*, 69, 128-142.

Thiele, A., Distler, C., Korbmacher, H. and Hoffmann, K.-P. (2004). Contribution of inhibitory mechanisms to direction selectivity and response normalization in macaque middle temporal area. *Proceedings of the National Academy of Sciences*, 101, 9810-9815.

Treue, S. and Martínez Trujillo, J.C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399, 575-579.

Treue, S. and Maunsell, J.H.R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, 382, 539-541.

Ullman, S. (1979). *The interpretation of visual motion*. Boston: MIT Press.

van Santen, J.P. and Sperling, G. (1985) Elaborated Reichardt detectors. *Journal of the Optical Society of America A.,* 2(2), 300-21.

Varela, J.A., Sen, K., Gibson, J., Fost, J., Abbott, L.F., and Nelson., S.B. (1997). A quantitative description of short-term plasticity at excitatory synapses in Layer 2/3 of rat primary visual cortex. *Journal of Neuroscience*, 17(20), 7926-7940.

Wade, N.J. and Swanston, M.T. (1987). The representation of nonuniform motion: Induced movement. *Perception*, 16(5), 555-571.

Wade, N.J. and Swanston, M.T. (1996). A general model for the perception of space and motion. *Perception*, 25, 187-194.

Wade, N.J. and Swanston, M.T. (2001). *Visual perception: An introduction*, 2nd edition. London: Psychology Press.

Wallach, H. (1935/1996). On the visually perceived direction of motion. *Psychologische Forschung*, 20, 325-380.

Wallach, H., Becklen, R. and Nitzberg, D. (1985). Vector analysis and process combination in motion perception. *Journal of Experimental Psychology: Human Perception and Performance*, 11(1), 93-102.

Womelsdorf, T., Anton-Erxleben, K. and Treue, S. (2008). Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. *The Journal of Neuroscience*, 28, 8934-8944.

Xiao, D.-K., Marcar, V.L., Raiguel, S.E. and Orban, G.A. (1997). Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. *European Journal of Neuroscience*, 9, 956-964.

Xiao, D.K., Raiguel, S., Marcar, V. and Orban, G.A. (1997). The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cerebral Cortex,* 7(7), 662-77.

Xu, X., Bonds, A.B., and Casagrande, VA. (2002). Modeling receptive-field structure of koniocellular, magnocellular, and parvocellular LGN cells in the owl monkey (Aotus trivigatus). *Visual Neuroscience*, 19(6), 703-11.