Computational Mechanisms for Perceptual Stability using Disparity and Motion Parallax

Oliver W. Layton¹ and Brett R. Fajen²

¹Department of Computer Science, Colby College, Waterville, Maine 04901, and ²Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, New York 12180

Walking and other forms of self-motion create global motion patterns across our eyes. With the resulting stream of visual signals, how do we perceive ourselves as moving through a stable world? Although the neural mechanisms are largely unknown, human studies (Warren and Rushton, 2009) provide strong evidence that the visual system is capable of parsing the global motion into two components: one due to self-motion and the other due to independently moving objects. In the present study, we use computational modeling to investigate potential neural mechanisms for stabilizing visual perception during self-motion that build on neurophysiology of the middle temporal (MT) and medial superior temporal (MST) areas. One such mechanism leverages direction, speed, and disparity tuning of cells in dorsal MST (MSTd) to estimate the combined motion parallax and disparity signals attributed to the observer's self-motion. Feedback from the most active MSTd cell subpopulations suppresses motion signals in MT that locally match the preference of the MSTd cell in both parallax and disparity. This mechanism combined with local surround inhibition in MT allows the model to estimate self-motion while maintaining a sparse motion representation that is compatible with perceptual stability. A key consequence is that after signals compatible with the observer's self-motion are suppressed, the direction of independently moving objects is represented in a world-relative rather than observer-relative reference frame. Our analysis explicates how temporal dynamics and joint motion parallax-disparity tuning resolve the world-relative motion of moving objects and establish perceptual stability. Together, these mechanisms capture findings on the perception of object motion during self-motion.

Key words: motion; MSTd; MT; object motion; optic flow; self-motion

Significance Statement

The image integrated by our eyes as we move through our environment undergoes constant flux as trees, buildings, and other surroundings stream by us. If our view can change so radically from one moment to the next, how do we perceive a stable world? Although progress has been made in understanding how this works, little is known about the underlying brain mechanisms. We propose a computational solution whereby multiple brain areas communicate to suppress the motion attributed to our movement relative to the stationary world, which is often responsible for a large proportion of the flux across the visual field. We simulated the proposed neural mechanisms and tested model estimates using data from human perceptual studies.

Introduction

Everyday situations demand that humans successfully interact with moving objects. This requires that we accurately perceive their trajectory during self-motion, which is impressive when one considers the complexity of the motion pattern that appears on the mobile observer's eye. The local optical motion only reliably corresponds to the object's movement through the world when

the observer is stationary; eye, head, and full-body self-motion all create global patterns of motion across the visual field that influence the motion corresponding to the object. Therefore, the local optical motion may reflect the observer's self-motion or a combination with the independent movement of objects.

The influence of self-motion could be reconciled by estimating its contribution and suppressing it (Fig. 1*a*). This follows because self-motion additively influences an object's optical motion. An estimate of the global motion induced by the observer's self-motion would allow the visual system to suppress its influence on the optical pattern to resolve the final undetermined component, the motion of the object relative to the stationary world. Crucially, the process of differentiating self-motion and object motion would suppress the global flux throughout the visual field induced by the observer's movement, thereby relating

Received Jan. 4, 2019; revised Sept. 24, 2019; accepted Oct. 7, 2019.

Author contributions: O.W.L. and B.R.F. designed research; O.W.L. and B.R.F. performed research; O.W.L. and B.R.F. analyzed data; O.W.L. and B.R.F. wrote the paper.

This work was supported by the Office of Naval Research (N00014-14-1-0359 and N00014-18-1-2283).

The authors declare no competing financial interests.

Correspondence should be addressed to Oliver W. Layton at oliver.layton@colby.edu.

<https://doi.org/10.1523/JNEUROSCI.0036-19.2019>

Copyright © 2020 the authors

Figure 1. The problem of recovering object motion independent of the observer's movement. **a**, An object moving up-and-left (large red arrow) through the world produces upward (large blue arrow) motion on the eye of a moving observer. This occurs because the optical motion (blue arrow) is the sum of object motion relative to the stationary world (red arrow) and the pattern induced by the observer's self-motion (rightward orange arrow). The world-relative motion (middle) can be recovered by suppressing the self-motion component (orange arrows, right) from the optical pattern (blue arrows, left). **b**, Suppressing an incorrect estimate of the self-motion component results in spurious object motion estimates. For example, if depth is systematically underestimated (right), too much motion would be suppressed, which would yield an incorrect object direction (gray instead of red arrow, middle) and make stationary objects appear to be moving (backward arrows, middle).

perception to the stationary world. This closely relates to perceptual stability, which refers to the perception of a stable environment despite its constantly changing optical projection induced by the observer's movement. In dynamic environments, perceptual stability involves perceiving objects that move with a constant velocity as moving along stable, invariant trajectories, despite changes in the observer's self-motion.

Although it is well established that self-motion perception involves visual (Gibson, 1950; Warren et al., 1988) and nonvisual (Bremmer et al., 1999; Cullen, 2012; Fajen and Matthis, 2013) contributions, much less is known about the neural mechanisms involved in establishing perceptual stability. Behavioral (Rushton and Warren, 2005; Matsumiya and Ando, 2009; Warren and Rushton, 2009; Dupin and Wexler, 2013) and preliminary neural (Peltier et al., 2016) studies suggest that the visual system does in fact suppress the self-motion component to establish world-relative object motion perception, which facilitates our interactions with moving objects and explains why objects appear to move along similar trajectories despite potential changes in our self-motion (e.g., starting to run when chasing a fly ball does not alter the perceived trajectory of the ball).

Despite the apparent simplicity of the solution, suppressing self-motion contributions from the optical motion signal requires complex, nuanced considerations. Here, we focus on one in particular, depth throughout the visual field, which the visual system must take into account, otherwise too much or too little would be deducted from the object's optical motion, potentially leading to widespread misperceptions in an object's trajectory. For example, "subtracting out" the self-motion component when depth is systematically underestimated leads to erroneous world-relative object motion and stationary objects appearing as moving (Fig. 1 compare *b*, center with *a*, center).

Neurophysiological studies have begun to identify how cells in areas, such as the middle temporal (MT; Nadler et al., 2008, 2013) and dorsal medial superior temporal (MSTd) areas (Yang et al., 2011), combine motion and depth through disparity during self-motion, and experiments have shown that humans take advantage of depth when judging the direction of moving objects, as if to establish a stable perceptual reference frame (Warren and Rushton, 2007; Matsumiya and Ando, 2009). The mechanisms that underlie and connect these processes are not yet well understood.

The aim of the present study is to formalize the fundamental computational principles involved in using depth through motion parallax and disparity to stabilize visual perception during self-motion. We developed a model to generate predictions of how bottom-up and top-down signals could interact; exploring specifically the hypothesis that feedback from MSTd to MT and MSTv plays a key role in perceptual stability and the recovery of world-relative object motion. Our analysis links surround suppression and joint motion parallax-disparity tuning and compares model behavior with existing behavioral and neural data.

Materials and Methods

Joint motion parallax and disparity model of MT and MST

In this section, we present an overview of the model, depicted in Figure 2a (for mathematical details, see Model equations). The model builds on ViSTARS (Browning et al., 2009a,b) and former versions of the Competitive Dynamics model (Layton et al., 2012; Layton and Browning, 2014; Layton and Fajen, 2016b, 2017), and consists of two interconnected brain areas: MT and MST. It is well established that neurons in area MT demonstrate selectivity to motion direction, speed, and disparity (Born and Bradley, 2005) and receptive fields (RFs) may (MT−) or may not (MT+) have a surround in which motion suppresses the response (Born, 2000; Cui et al., 2013). The properties of neurons in MST differ depending on their anatomical position: cells in MSTd have large RFs and are involved with heading perception (Duffy and Wurtz, 1995); cells in ventral MST (MSTv) have comparatively smaller RFs and respond to small moving objects (Tanaka et al., 1986, 1993).

To anticipate the model's functional organization, we assume that perceived self-motion is reflected in model area MSTd, perceived object motion in model areas MT− and MSTv, and that these two areas interact to recover world-relative object motion. Known connectivity and neurophysiological properties implicate MT and MST for implementing mechanisms like those described here, though this does not rule out the involvement of other areas with similar characteristics. For concreteness, we will specifically reference MT and MST.

Overview

The model takes as input motion vectors $\vec{v}_{x,y,f} = (v_{dx,f}, v_{dy,f}, v_{h,f})$ specifying the motion ($v_{dx,f}, v_{dy,f}$) in degrees/second, and depth ($v_{h,f}$), in centimeters, of points in the world at retinal position (x, y) and frame f of the discrete video-based input sequence. We represented the input in this fashion to focus on the contributions of the novel mechanisms; it could be seen as the output of preprocessing stages that extract motion from a sequence of images (e.g., LGN and V1) and perform stereo matching across binocular inputs (e.g., V1, V2). Direction, speed, and disparity

Figure 2. Diagram of model MT and MST. **a**, The model contains two major pathways: MT+ /MSTd subserves self-motion (top) and MT− /MSTv subserves object motion (bottom). The model is dynamic and estimates self-motion and object motion simultaneously. Units in MSTd model physiologically supported populations tuned to full-field radial patterns, speed (bandpass, speed summing), and non-direction-dependent disparities. Cells in MT− /MSTv have RFs with inhibitory surrounds due to local and feedback-mediated suppression. Feedback comes from dominant MSTd units and suppresses MT− /MSTv units locally tuned to the direction, speed, and disparity that match the estimated self-motion motion pattern. As we will show, feedback-mediated suppression shifts population object motion responses from a retinal to world-relative reference frame. The efficacy of MT+ /MSTd synapses decreases in response to tonic presynaptic activity in MT+. Model MT tuning curves for direction (**b**), speed (**c**), and disparity (**d**).

filters process the input signal and convert the motion and depth vectors into a neural signal by layer 4,6 MT units (hereafter referred to as MT L4,6); units respond to the input based on their tuning curves. We show sample tuning curves for direction, speed, and disparity in Figure 2b–d.

Units in the next stage, layer 2–3 (hereafter referred to as MT L2–3), are jointly tuned to direction, speed, and disparity, but process the output of MT L4,6 differently, depending on whether they possess reinforcing (MT+) or antagonistic (MT−) surrounds (Fig. 2a). This bifurcation reflects the anatomical segregation of MT into pathways that subserve self-motion and object motion (Born and Tootell, 1992; Tanaka et al., 1993; Born, 2000; Yu et al., 2018): MT+ units project to MSTd and MT− project to MSTv. The former pathway matches the MT+ output signals against laminar (Fig. 2a, top) and radial (Fig. 2a, bottom) templates that define the pattern selectivity of MSTd units. We modeled MSTd units that either respond proportionally to retinal speed (“speed summing cells”; Inaba et al., 2007; Perrone, 2012) or respond maximally to different average speeds across the visual field (“bandpass cells”; Duffy and Wurtz, 1997). Every MSTd unit pools over similar MT+ disparity signals to which it is tuned. The disparity tuning of every model MSTd unit is independent of pattern (radial, laminar) and/or speed selectivity, consistent with most cells in MSTd (non-direction-dependent cells; Yang et al., 2011). This motion integration process combined with intra-areal dynamics within MSTd allows the model to estimate the speed gradient and direction pattern that corresponds to the observer’s self-motion (motion parallax) while taking depth through disparity into account.

Units in the MT− /MSTv pathway have surrounds that are tuned to similar directions, speed, and disparities as the RF centers (for anticipated response differences as selectivity varies, see Results). Both areas perform an on-center/off-surround integration of motion and disparity signals, with the exception that MSTv units respond proportionally to retinal speed (Tanaka et al., 1993) like MSTd speed summing cells. We readout the model’s estimate of the object’s motion, usually in area MSTv, from units that have the moving object within the RF at the end of each simulation. We used a population estimate for increased precision when comparing model-derived estimates to human judgments (see Direction readout).

We used the model to explore the prediction that MSTd cells that respond to the pattern of motion generated by the observer’s self-motion send feedback to MT− and MSTv to establish perceptual stability and recover world-relative object motion. We hypothesized that there is a general guiding structure in these feedback signals.

Specifically, we considered the overarching rule whereby MSTd feedback suppresses MT− /MSTv units most compatible with the visual pattern to which the most active MSTd units are tuned. Figure 3 illustrates how this principle not only encompasses direction (i.e., suppressing MT units tuned to directions consistent with the global visual pattern), but also speed and disparity. For example, Figure 3a shows that MT− /MSTv units tuned to far disparities/slow speeds and near disparities/fast speeds (“congruent cells”) receive greater inhibition due to MSTd feedback than units tuned to far disparities/fast speeds and near disparities/slow speeds (“opposite cells”). Note that we define model congruent and opposite

Figure 3. Overview of suppressive feedback mechanism. MT— and MSTv units are inhibited depending on how their tuning properties relate to the corresponding local subregion of the most active MSTd unit's RF. **a**, The strength of the inhibitory feedback from MSTd is speed- and depth-dependent in accordance with motion parallax. When an observer moves forward, the rate of flow is greatest at near depths and least at far depths. Likewise, the MT— and MSTv cells that are inhibited the most are those that are tuned to faster speeds and near disparities, moderate speeds and fixational disparities, and slower speeds and far disparities (i.e., those with “congruent” speed-disparity tuning; see enclosed region along the diagonal in **a**). **b**, The pattern selectivity of the MSTd unit is indicated on the top, right and the RF of some MT units are shown in left superimposed circle. Highest peak curve in plot shows that MT suppression is greatest when the MT direction tuning locally matches (leftward; 0° mismatch) the MSTd RF; units with mismatching direction tuning receive progressively less inhibition. Other curves show that inhibition drops off as speed tuning locally differs from the MSTd RF.

cells based on their disparity and speed tuning, which may differ from cells with same name reported in existing physiological studies. For example, the properties of congruent and opposite cells shown by Nadler et al. (2013) and Kim et al. (2016) are based on the consistency between depth tuning and motion parallax during self-motion with eye movements. Because we do not model eye movements and our viewing geometry differs from that of Nadler et al. (2013) and Kim et al. (2016), we caution against direct comparisons between the cell populations unless evidence is produced to substantiate such a connection.

Figure 3, *a* and *b*, shows that the suppression drops off as tuning in any particular dimension becomes more dissimilar among units that share RFs in the same visuotopic region of the visual field (i.e., in the same MT—/MSTv macrocolumn). For example, suppression decreases in MT—/MSTv units with RFs that coincide with the left side of the horizon and deviate in speed (non-blue curves) and direction (positions along the blue curve that do not correspond to the maximum) tuning from the global pattern depicted in Figure 3*b*.

Direction readout

To estimate the direction of moving objects from the model's neural signals, we relied on a population vector readout technique whereby we weighted the direction preference by the firing rate among units that respond to a moving object. With this approach, activity from the entire population of direction-tuned units factors into the estimate.

$$\tilde{J}(w_d; \theta_d) = \frac{1}{\sum_{d=1}^{\hat{d}} w_d} (J_u, J_v) = \left(\sum_{d=1}^{\hat{d}} w_d \cos \theta_d, \sum_{d=1}^{\hat{d}} w_d \sin \theta_d \right). \quad (1)$$

In analyses, we calculated the population vector for both MSTv cells $\tilde{J}(P_{x,y,d,h}^v; \theta_d)$ and MT— cells $\tilde{J}(M_{x,y,d,s,h}^-; \theta_d)$. We relied on the following equation to extract the object angle signaled by the MT—/MSTv population:

$$\tilde{J} = \arctan 2(J_v, J_u). \quad (2)$$

Finally, we subtracted the object's retinal direction θ_{Obj} from the model estimate $\tilde{J}$ (Eq. 3) to compute the relative difference $\tilde{\theta}_{Obj}$ between the retinal and world relative object directions. A zero valued $\tilde{\theta}_{Obj}$ would

indicate that the neural population maintained the retinal direction of the object, whereas a positive value would indicate a shift in the signal toward the world relative direction.

$$\tilde{\theta}_{Obj} = \tilde{J} - \theta_{Obj}. \quad (3)$$

Simulations

We implemented the model in MATLAB and performed simulations using R2017a on a 4.3 Ghz quadcore desktop machine with 32GB of memory running Microsoft Windows 10. We numerically integrated the model using Euler's method such that each new input frame was integrated 10 time steps. This amounts to a step size of ≈ 3 ms, assuming 30 frames/s. We implemented the divisive, subtractive, and local surround suppression mechanisms in Wolfram Mathematica 11.1.

When simulating the model with monocular inputs, we withheld the input depth signal and left out the disparity related tuning curves and parameters (e.g., $q^{d,6}$, q^{MSTv}), replacing their contribution in equations with unity.

Experimental design and statistical analysis

We performed a series of simulation experiments to test computational mechanisms in

the model. Each experiment involved simulating a different set of visual displays. Given that we often compared model performance with human judgments, some of these displays correspond to those used in psychophysical studies as indicated.

Full, global, local conditions

To test interactions between the suppressive feedback and local inhibitory computational mechanisms, we generated optic flow displays (Full, Global, and Local conditions) that correspond to those used by Warren and Rushton (2009). The Full condition contained full-field motion. The Global condition was similar, except motion within a circular mask centered on the object was removed. There was also a mask in the Local condition, except motion was masked on the exterior. In all conditions, the simulated observer moved at 59 cm/s along a central straight-ahead heading through a 3D volume of dots. To maintain the dot density over time, we regenerated dots at a random relative depth when they exited the field-of-view. The virtual 3D dot cloud contained 4000 dots, up from the 300 used by Warren and Rushton (2009) to account for our use of stereo (they used monocular displays). The dots spanned relative depths between 50 and 150 cm from the observer. The object moved along the depth axis at the same speed as the observer (59 cm/s) so as to maintain a constant depth of 100 cm. It also moved vertically at a speed of 0.65°/s from an initial position on the horizontal midline at a retinal eccentricity of 6.5° to the right of center. We generated 14 frames of input, presenting each to the model for 30 ms.

We also simulated monocular versions of the Full, Global, and Local conditions to explore the contributions of disparity. These were identical to their stereo counterparts, except the input depth signal was withheld from the model.

Translation and rotation conditions. To explore the influence of disparity tuning in MT/MST on model mechanisms, we simulated the stereo displays of Warren and Rushton (2007). The simulated observer translated rightward at 2 cm/s in the Translation condition or performed a 0.75°/s rightward eye movement in the Rotation condition. In each self-motion scenario, the simulated environment either consisted of a 3D dot cloud 100 cm deep (Full condition) or a 20 cm cross-section thereof (Variable Depth conditions). The Full and Variable Depth conditions in the present study correspond to the displays used by Warren and Rushton (2007) in their Experiments 1–2 and 3, respectively. Our implementation deviated from theirs in that our scene consisted of a 3D cloud of 2000 dots rather than textured wireframe cubes. The object was posi-

Figure 4. The influence of divisive and subtractive suppression of self-motion on object motion signals. **a**, Simulated forward self-motion toward a wall (purple radial pattern) in the presence of a moving object that moves upward on the eye (blue “retinal” arrow). Top, Right, Pattern preference of the MSTd unit that responds best to the radial optical motion pattern on the eye. Bottom, Schematic local direction signals integrated by MT neurons (purple Gaussian profiles) in two specific regions: the object (bottom, left) and lower-right portion of the visual field (bottom, right). The purple Gaussian profiles (Retinal motion; +) show the feedforward input motion signal to MT units. The orange profiles show the suppressive feedback signal, which locally targets MT units consistent with the pattern preference of the MSTd cell sending the feedback (top, right). Motion MT responses emerge through an interaction between the top-down (orange) and bottom-up (purple) signals. **b**, Shift in the object direction represented by the peak MT population response, compared with the optical direction, from a mechanism that models MT–MSTd interactions through divisive interactions. The x-axis represents the degree of similarity between the feedforward and feedback signals: 0° indicates perfect overlap (**a**, bottom, right) and 180° indicates small degrees of overlap. The different curves show the influence that the broadness (σ) of each signal (i.e., variance of the orange and purple direction distributions in **a**) has on the shift. **c**, Same as **b** except generated by a mechanism that models subtractive interactions between MT–MSTd signals. **d**, SD (broadness) of the resulting object direction distribution after MT–MSTd interactions as predicted by the divisive interaction model. The top thin dashed line indicates the broadness of the MT population activity without suppression in the case of the red curve ($\sigma = 25^\circ$). **e**, Same as **d** except generated by the subtractive interaction model.

tioned 85, 105, or 125 cm away from the observer, at 6.5° eccentricity right of the horizontal midline. It moved upward at 0.9 cm/s through the world. In the Variable Depth conditions, the 3D cloud was centered at each of the object depths (85, 105, or 125 cm).

Direction threshold object detection conditions. We performed several experiments that compared the model’s moving object detection capabilities with human thresholds. In some cases, we simulated a subset of the optic flow displays used by Royden and Connors (2010) in their Experiment 4, whereby the simulated observer translated forward at 100 cm/s through a virtual environment consisting of stationary discs and a moving object, which appeared at 5° eccentricity and moved at a range of retinal speeds. In our implementation, we used a 3D cloud of 500 dots rather than discs spanning relative depths of 70 and 140 cm. Moreover, we created a stereo version of the displays wherein the moving object started 100 cm in depth relative to the observer.

To compute model moving object detection thresholds, we increased the object’s angular deviation in 1° increments from the radial background pattern until the MSTd cell whose RF was centered on the object exceeded a firing rate threshold of 0.154 after 14 frames of input.

Disparity-based object detection conditions. We also performed an experiment that tested the extent to which the model can detect a moving object based on disparity. We created displays with a 3D volume of 5000 dots 110–140 cm away from the observer wherein a small moving object moved as if it were part of the stationary background, except we independently controlled its depth (70–120 cm from the observer). The object moved with a direction and speed like a dot positioned at the forefront of the 3D dot volume (110 cm), but we adjusted the z position in simulation. At the selected horizontal offsets of 15, 35, and 55 cm the object appeared at approximate average eccentricities of 10°, 20°, and 30°, respectively.

Model equations

The following two sections present computational mechanisms for perceptual stability (Figs. 4, 5). These mechanisms focus on direction alone, ignoring speed and disparity to focus on the principles; afterward, we present equations for the more elaborate model that builds on these mechanisms to account for interactions between motion parallax and disparity that evolve over time.

Computational mechanisms for perceptual stability

Self-motion and object motion interaction. Local motion signals that surround a moving object cannot alone account for human judgments about its direction (Warren and Rushton, 2008, 2009); strong evidence implicates a mechanism that relies on the global pattern of motion throughout the visual field. We propose that this global mechanism emerges in the process of establishing perceptual stability based on an estimate of the visual pattern corresponding to the observer’s self-motion. This singular estimate obtained by integrating the motion across the visual field must suppress different matching local motion signals, depending on their originating position on the retina. We considered how this task could be accomplished through interactions between neurons small enough to locally signal the direction of moving and stationary objects in the environment (e.g., in MT) and feedback from neurons with large enough RFs to integrate global optic flow patterns that arise during self-motion (e.g., in MSTd). We examined how two plausible interaction models would impact the direction signaled by a population of neurons that respond to moving object and share a common spatial RF (e.g., in a MT macrocolumn). The first is a divisive inhibition model (Grossberg, 1973; Reynolds and Heeger, 2009) that satisfies a Naka–Rushton equation (Naka and Rushton, 1966):

$$x_d = \frac{E_d}{A + F_d}, \quad (4)$$

where x_d represents the response of neuron x tuned to direction d , E_d is the excitatory (feedforward) input signal, F_d is the direction tuned feedback derived inhibitory signal, and the constant A represents the cell’s decay rate. When considering the inputs to the entire direction tuned population of units with spatially overlapping RFs, the excitatory and inhibitory inputs become the vectors $\vec{E}$ and $\vec{F}$, respectively. We modeled these directional inputs with the following Gaussians distributions:

$$E_d = e^{-\left(\frac{d - \mu_1}{\sqrt{2}\sigma_1}\right)^2}, \quad (5)$$

$$F_d = e^{-\left(\frac{d - \mu_1}{\sqrt{2}\sigma_1}\right)^2}, \quad (6)$$

Figure 5. Effects of local center-surround interactions on object direction signals. **a**, Scenario used in simulations: forward, straight-ahead simulated self-motion and rightward moving object that changed positions along an iso-eccentricity circle. Among units that signal object motion, the central direction remains fixed, whereas the surround direction varied systematically (difference represented by x -axis in **b** and **c**). **b**, Effect of broadness of the center and surround RF direction tuning curves (e.g., broadness of the Gaussians in Fig. 4a, bottom) on the shift in the object's direction toward the world-relative direction. Color curves show simulations for different relative ratios between the center/surround tuning bandwidth (σ_c/σ_s). Positive shifts are in the world-relative direction. The baseline tuning width is $\sigma_c = 20^\circ$. **c**, Effect of gain in the surround direction tuning curve on the object direction signal. The baseline center gain is $g_c = 1$. **d**, Output of the center (top) and surround (middle) direction tuning curves for a unit maximally tuned to the object's rightward motion (0°). The local center-surround interactions induce a shift in the output distribution in the world-relative direction (bottom).

where (μ_i, μ_j) indicate the direction preference of the cells most stimulated and inhibited, respectively, and (σ_i, σ_j) control the extent to each input influences units with dissimilar direction preferences. In simulations summarized in Figure 4, we fixed the excitatory input ($\mu_i = 0^\circ, \sigma_i = 25^\circ$) and varied which cells received the inhibition, adjusting $\mu_j \in (0^\circ, 180^\circ)$. We fixed $A = 0.2$, 20% of the maximum mean firing rate and set σ_j as labeled in Figure 4.

For the subtractive model, we used the following equation:

$$x_d = [E_d - F_d]^+, \quad (7)$$

where the inputs E_d and F_d are defined according to Equations 5 and 6, respectively, and $[\cdot]^+$ represents thresholding to maintain positivity in the firing rate.

In both cases, we quantified the influence of suppression on the direction signaled by the cell population $\tilde{x}$ (Fig. 4b,c), we computed the shift s (in degrees) in the activity peak induced by the divisive or subtractive interactions:

$$s(\tilde{x}, \tilde{E}) = \operatorname{argmax}_d \tilde{x} - \operatorname{argmax}_d \tilde{E}. \quad (8)$$

Surround suppression from local motion signals. In this section, we present a simple model to quantify how the well established on-center/off-surround organization of MT RFs could influence object motion responses due to surround suppression. We used the following Gaussian filter to model the MT unit RF center and surround:

$$G_{x,d}(x_0, d_0, \sigma_x, \sigma_d, g) = \frac{g}{2\pi\sigma_x\sigma_d} e^{-\left(\left(\frac{x-x_0}{\sqrt{2}\sigma_x}\right)^2 + \left(\frac{d-d_0}{\sqrt{2}\sigma_d}\right)^2\right)}, \quad (9)$$

where (x_0, d_0) describe the spatial position and direction, respectively, that yield the maximum filter response, σ_x controls the sensitivity across space, and g refers to the filter gain. The parameter σ_d controls the degree to which the center or surround is direction tuned. For example, large values of σ_d produce an untuned surround (Rust et al., 2006; Cui et al., 2013), whereas small values make the filter highly selective for a particular direction. Equation 9 assumes a one-dimensional (1D) representation of space to keep things simple, but the filter could be readily extended to include a second spatial dimension. We defined the center region as $G_{x,d}(x_0, d_0, \sigma_x, \sigma_d, g_c)$, assuming that the RF is centered at the origin ($x_0 = 0$), has unit gain ($g_c = 1$), and the unit is tuned to rightward motion ($d_0 = 0^\circ$). The latter assumption is arbitrary and made for concreteness because we are concerned with the relative tuning of the surround.

We defined the surround as $G_{x,d}(x_0, d_0, \sigma_x, \sigma_d, g_s)$, with the assumption that its spatial extent is broader than the center ($\sigma_s > \sigma_x$). Specifically, we fixed $\sigma_x = 0.42^\circ$ and $\sigma_s = 4.25^\circ$ to define 1° and 10° center and surround regions, measured by the full-width at half-maximum (FWHM). By default, we equated the center/surround directional pooling extent to a broad value $\sigma_{ds} = \sigma_{dc} = 40^\circ$ ($\sim 95^\circ$ FWHM; Born, 2000) and set $g_s = 30$.

We modified σ_{ds} according to the ratio σ_{ds}/σ_{dc} for the simulations summarized in Figure 5b to test how the strength of surround direction tuning impacted object motion signals: larger values σ_{ds} weaken the surround direction selectivity (i.e., surround pools over a greater range) and smaller values strengthen it. In Figure 5c, we show the influence of g_s .

For the simulations reported in Figure 5, we considered a scenario wherein the observer moves straight forward along a central heading toward a distant wall in the presence of a small object that moves rightward in the upper periphery of the visual field (Fig. 5a). In separate simulations, we repositioned the object at the same eccentricity to assess the impact of different directions in the surround, while keeping the object centered on the model unit's RF. That is, rightward motion appears in the RF center, and both rightward and the background motion directions (e.g., upward when the object appears on the middle upper region of the visual field) appear in the surround. Assuming an MT unit with a 1° center and surround 10° as wide (Born and Bradley, 2005), we defined the model input as follows:

$$I_{x,d} = \begin{cases} 1 & |x| < 0.5^\circ, d = d_0 \\ 1 & 0.5^\circ \leq |x| < 5^\circ, d = \tilde{d} \\ 0 & \text{otherwise} \end{cases}, \quad (10)$$

where $\tilde{d}$ specifies the direction present in the surround.

The stimulus-dependent center (C_{x_0,d_0}) and surround (S_{x_0,d_0}) filter output is given by the following:

$$C_{x_0,d_0} = \int_{-180^\circ}^{180^\circ} \int_{-\infty}^{\infty} G_{x,d}(x_0, d_0, \sigma_{x_c}, \sigma_{d_c}) I_{x,d} dx dd, \quad (11)$$

$$S_{x_0,d_0} = \int_{-180^\circ}^{180^\circ} \int_{-\infty}^{\infty} G_{x,d}(x_0, d_0, \sigma_{x_c}, \sigma_{d_c}) I_{x,d} dx dd. \quad (12)$$

We used the divisive model specified by Equation 4 to generate the MT unit response:

$$y_{x_0,d_0} = \frac{C_{x_0,d_0}}{A + S_{x_0,d_0}}, \quad (13)$$

setting $A = 0.1$. After evaluating Equation 13, we set $x_0 = 0$, because we are only interested in units whose RFs coincide with the origin (centered on the object). As in the previous section, we quantified the influence of surround suppression on the direction carried by the object motion signal using the following equation, analogous to Equation 8:

$$s(y_{0,d_0}, d_0) = \underset{d_0}{\operatorname{argmax}} y_{0,d_0} - d_0. \quad (14)$$

Equation 14 computes the difference between the direction signaled by the most active MT unit and the direction of the object present in the RF center.

Joint motion parallax and disparity model of MT and MST

Here, we present the model of MT and MST that integrates mechanisms described in the previous sections into a unified system that takes into account speed and disparity, not only direction. Additionally, the model estimates the motion pattern corresponding to the observer's self-motion in area MSTd and dynamically influences object motion signals, even as the object moves and the self-motion estimate updates over time. The model extends the Competitive Dynamics model (Layton and Fajen, 2016b) and builds on the VISTARS model (Browning et al., 2009a,b).

Notation. When defining functions, we use the notation $F(x_1, \dots, x_n; p_1, \dots, p_m)$, where F is the function name, $x_1, \dots, x_n$ are the independent variables, and $p_1, \dots, p_m$ are parameters. In expressions and equations, we use the notation $F_{x_1, \dots, x_n}^a$, where F is the name, $x_1, \dots, x_n$ are the independent variables, and a refers to the relevant model area, if applicable. Parameters that appear with a hat refer to ordinal count parameters (e.g., $\hat{d}$ refers to the 24 MT tuning directions). Specific parameter values used in Gaussian or von Mises tuning curves appear in Table 1, values that relate to neural dynamics appear in Table 2.

MT layer 4,6. Activity within the input layers of MT is obtained by filtering the input motion and depth signal ($\tilde{v}_{x,y,f}$) with direction, speed, and disparity tuning curves to define the joint sensitivity of each unit. We filtered input signals along each of these dimensions independently because these tuning characteristics are largely separable (Smolyanskaya et al., 2013). We defined MT direction tuning according to a von Mises distribution:

$$V(a; \mu, \kappa) = \frac{\exp(\kappa \cos(a - \mu))}{\exp(\kappa)}, \quad (15)$$

and evaluated it at $V(\Theta_{x,y,f}; \theta_d, b^{4,6})$ for all x, y , and input frame f . We extract the angle of the motion vector at every position ($\Theta_{x,y,t}$) using the two-argument form of arctangent:

$$\Theta_{x,y,f} = \arctan 2(v_{dy,f}, v_{dx,f}). \quad (16)$$

Equation 15 matches this angle with the preferred direction θ_d of each unit, where the parameter $b^{4,6}$ controls the direction tuning width. We

normalized Equation 15 so that it outputs 1 when the peak tuning of the unit matches the input.

MT L4,6 units process the input signal for speed by computing the following:

$$\rho_{x,y,f} = \sqrt{(v_{dy,f})^2 + (v_{dx,f})^2} \quad (17)$$

and evaluating the following Gaussian tuning curve at $G(\rho_{x,y,f}; \rho_s^{4,6}, \sigma_s^{4,6})$ for all x, y , and f :

$$G(a; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{a-\mu}{\sqrt{2}\sigma}\right)^2}, \quad (18)$$

where $\sigma_s^{4,6}$ controls the width of the tuning curve across speeds and $\rho_s^{4,6}$ define the speeds that yield the maximal sensitivity in the MT population. As indicated in Table 1, we created units tuned to five speeds ($s = 1, 2, \dots, \hat{s}, \hat{s} = 5$ and defined the speeds that elicit the maximal filter response according to speed distribution quintiles present in the first frame of each input ($\hat{p}_1$).

For disparity, we evaluated the Gaussian tuning curve defined by Equation 18 at $q^{4,6}G(v_{h,f}; \delta_h^{4,6}, \sigma_h^{4,6})$, where $\sigma_h^{4,6}$ controls the width of the disparity tuning curves and $\delta_h^{4,6}$ define the depths that yield the maximal sensitivity in the MT population. We multiplied the filter by a disparity gain $q^{4,6}$ to simulate the tendency for stereo stimuli to garner greater activation in MT (Nadler et al., 2013). As indicated in Table 1, we created units tuned to five disparities ($h = 1, 2, \dots, \hat{h}, \hat{h} = 5$). We configured $\sigma_h^{4,6}$ using one of the following functions:

$$\lambda^{Inc}(h; \sigma_0, \sigma_{dh}) = \sigma_0 + \sigma_0 \sigma_{dh}(h - 1), \quad (19)$$

$$\lambda^{Const}(h; \sigma_0) = \sigma_0, \quad (20)$$

where $\lambda^{Inc}(h; \sigma_h^{4,6}, \sigma_{dh}^{4,6})$ represents the Increasing Variance model and $\lambda^{Const}(h; \sigma_h^{4,6}, \sigma_{dh}^{4,6})$ represents the default scheme for how disparity tuning varied with depth.

We computed the MT L4,6 activation $M_{x,y,d,s,h,f}^{4,6}$ by multiplicatively combining the input-tuning curve outputs that filter the input for direction, speed, and disparity:

$$M_{x,y,d,s,h,f}^{4,6} = [V(\Theta_{x,y,f}; \theta_d, b^{4,6})] [G(\rho_{x,y,f}; \rho_s^{4,6}, \sigma_s^{4,6})] [q^{4,6}G(v_{h,f}; \delta_h^{4,6}, \sigma_h^{4,6})]. \quad (21)$$

MT layer 2–3 (MT+). MT+ units perform a spatial integration of MT L4,6 activity ($M_{x,y,d,s,h,f}^{4,6}$) throughout the RF using a two-dimensional (2D) Gaussian filter:

$$G(x,y; x_0, y_0, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left(\left(\frac{x-x_0}{\sqrt{2}\sigma_x}\right)^2 + \left(\frac{y-y_0}{\sqrt{2}\sigma_y}\right)^2\right)}. \quad (22)$$

The following 2D convolution function specifies how this integration occurs:

$$C(w_{x,y}; \mu_x, \mu_y, \sigma_{x,y}, g) = g \sum_n \sum_m G(n, m; \mu_x, \mu_y, \sigma_{x,y}, \sigma_{x,y}) w_{x-n, y-m}. \quad (23)$$

We evaluate Equation 23 as $C_{x,y,d,s,h,t}^{4,6} = C(M_{x,y,d,s,h,t}^{4,6}; x,y, \sigma^{MT+,cent}, 1)$ to obtain the MT+ input, where $\lceil t \rceil$ indicates the ceiling operation taken with respect to time t to map continuous time to the discrete input signal at frame f .

We defined the MT+ activity ($M_{x,y,d,s,h,t}^{+}$) according to the leaky integrator equation:

$$\frac{dM_{x,y,d,s,h,t}^{+}}{dt} = -M_{x,y,d,s,h,t}^{+} + (1 - M_{x,y,d,s,h,t}^{+}) C_{x,y,d,s,h,t}^{+} \quad (24)$$

Among models that explain MSTd responses based on their feedforward input from MT, those that include a nonlinearity that compresses the MT

Table 1. Tuning curve parameter values used in joint motion parallax and disparity model

Parameter	Areas	Purpose	Description	Value	Unit
$\hat{d}$	All	Direction	Number of direction tunings	24	Ordinal
θ	All	Direction	Directions to which units are tuned	$[-180, \dots, 165]$	Degrees
$\hat{s}$	All	Speed	Number of speed tunings	5	Ordinal
$\hat{h}$	All except MSTd	Disparity	Number of disparity tunings	5	Ordinal
$\hat{x}$	All except MSTd	Spatial	Size of horizontal dimension	64	Pixels (°)
$\hat{y}$	All except MSTd	Spatial	Size of vertical dimension	64	Pixels (°)
$\hat{k}$	MSTd	Disparity	Number of MSTd disparities	3	Ordinal
q^{MSTd}	MSTd	Disparity	Disparity dependent gain	4.5	N/A
σ_h^{MSTd}	MSTd	Disparity	MT + /MSTd disparity tuning width	0.75	N/A
ϕ	MSTd	Disparity	MSTd disparities	1,3,5	Ordinal
b^{MSTd}	MSTd	Spatial	Extent of MSTd distance-dependent spatial integration from the preferred singularity position	5×10^{-3}	Pixels (°)
i	MSTd	Spatial	Number of evenly spaced horizontal singularities to which MSTd units are tuned	16	Ordinal
j	MSTd	Spatial	Number of evenly spaced vertical singularities to which MSTd units are tuned	16	Ordinal
b_d^v	MSTv	Direction	Surround direction tuning curve width	7	Degrees
g_d^v	MSTv	Direction	Surround direction tuning curve gain	0.26	N/A
q^{MSTv}	MSTv	Disparity	Surround disparity gain	1.5	N/A
$g^{v,cent}$	MSTv	Spatial	2D center spatial filter gain	4	N/A
$g^{v,surr}$	MSTv	Spatial	2D surround spatial filter gain	6	N/A
$r^{v,cent}$	MSTv	Spatial	Diameter of 2D center spatial filter	3	Pixels (°)
$r^{v,surr}$	MSTv	Spatial	Diameter of 2D surround spatial filter	7	Pixels (°)
$\sigma^{v,cent}$	MSTv	Spatial	Center spatial tuning extent	0.5	Pixels (°)
$\sigma^{v,surr}$	MSTv	Spatial	Surround spatial tuning extent	4	Pixels (°)
b_d^-	MT—	Direction	Surround direction tuning curve width	7	Degrees
g_d^-	MT—	Direction	Surround direction tuning curve gain	0.26	N/A
σ_h^-	MT—	Disparity	Surround disparity tuning curve width	1	Degrees
$g^{MT-,cent}$	MT—	Spatial	2D center spatial filter gain	1.5	N/A
$g^{MT-,surr}$	MT—	Spatial	2D surround spatial filter gain	0.5	N/A
$r^{MT-,cent}$	MT—	Spatial	Diameter of 2D center spatial filter	5	Pixels (°)
$r^{MT-,surr}$	MT—	Spatial	Diameter of 2D surround spatial filter	7	Pixels (°)
$\sigma^{MT-,cent}$	MT—	Spatial	Center spatial tuning extent	0.5	Pixels (°)
$\sigma^{MT-,surr}$	MT—	Spatial	Surround spatial tuning extent	4	Pixels (°)
$b^{A,6}$	MT L4,6	Direction	Direction tuning curve width	4	Degrees
$q^{A,6}$	MT L4,6	Disparity	Disparity dependent gain	45	N/A
$\delta^{A,6}$	MT L4,6	Disparity	Depths to which MT units are tuned	70, 85, 100, 115, 130	cm
$\sigma_{h_0}^{A,6}$	MT L4,6	Disparity	Disparity tuning curve width	8	Degrees
$\sigma_{dh}^{A,6}$	MT L4,6	Disparity	Change in disparity tuning curve width with depth	0.5	Degrees
b_d^{FB}	MT—, MSTv	Direction	Direction specificity of suppressive MSTd feedback	1	Degrees
b_s^{FB}	MT—, MSTv	Speed	Speed specificity of suppressive MSTd feedback	0.2	Pixels/s (°)
b_h^{FB}	MT—, MSTv	Disparity	Disparity specificity of suppressive MSTd feedback	4	Degrees
Γ^{FB}	MT—, MSTv	N/A	Activity threshold needed for MSTd units to send suppressive feedback	0.01	Mean spike rate
$r^{MT+,cent}$	MT+	Spatial	Diameter of feedforward 2D spatial filter	10	Pixels (°)
$\sigma^{MT+,cent}$	MT+	Spatial	2D spatial tuning extent	3	Pixels (°)

Where applicable, we parameterized the model in pixels for convenience for processing stimuli represented as a set of motion vectors in a sequence of digital images. Units of pixels (°) means that the given value in pixels may be readily converted to degrees of visual angle.

Table 2. Dynamical parameter values used in joint motion parallax and disparity model

Parameter	Area	Description	Value	Unit
κ^+	MT+	Synaptic depression recovery rate	10	Mean spikes/s
τ^+	MT+	Synaptic depression time constant	10	Milliseconds
α^{MSTd}	MSTd	Decay rate	0.1	Mean spikes/s
γ^{MSTd}	MSTd	Inflection point in recurrent feedback sigmoid transfer function	0.12	Mean spike rate
Γ^{MSTd}	MSTd	Activity threshold required to send recurrent feedback	0.28	Mean spike rate
β^-	MT—	Activity hyperpolarizing lower bound	0.4	Mean spike rate
β^{MSTv}	MSTv	Activity hyperpolarizing lower bound	0.3	Mean spike rate

signals perform best (Mineault et al., 2012). The compressive nonlinearity could be explained by synaptic depression, the tendency for the same inputs to lose their efficacy over time. We modeled MT+ synaptic depression ($Y_{x,y,d,s,h}^+$) as follows:

$$\frac{1}{\tau^+} \frac{dY_{x,y,d,s,h}^+}{dt} = 1 - Y_{x,y,d,s,h}^+ (1 + \kappa^+ M_{x,y,d,s,h}^+), \quad (25)$$

$$N_{x,y,d,s,h}^+ = M_{x,y,d,s,h}^+ Y_{x,y,d,s,h}^+, \quad (26)$$

where $N_{x,y,d,s,h}^+$ denotes the output signal to MSTd from MT+ units, τ^+ is the synaptic time constant, and κ^+ represents the rate at which the efficacy of the input signal $M_{x,y,d,s,h}^+$ declines over time (Grossberg, 1980).

MSTd

Visual self-motion pattern matching (MT+ – MSTd). The model includes MSTd units tuned to the following laminar ($\Lambda_{d,x,y}$) or radial ($\Pi_{i,j,x,y}$) patterns:

$$\Lambda_{d,x,y} = (\cos(\theta_d), \sin(\theta_d)), \quad (27)$$

$$\Pi_{i,j,x,y} = (x - i, y - j), \quad (28)$$

where θ_d refers to the directions of maximal MT tuning and (i, j) indicates the position of the singularity in the radial template. MSTd cells tuned to laminar patterns spatially average MT+ output signal tuned for a common direction and speed:

$$L_{d,s,h} = \frac{1}{\hat{x}\hat{y}} \sum_x \sum_y N_{x,y,d,s,h}^+ \quad (29)$$

where $(\hat{x}, \hat{y})$ refers to the spatial dimensions of the MT+ unit representation in the model. Considering that the input comes as a sequence of digital images, we assumed for convenience a 1-to-1 mapping between pixels in each image and the spatial distribution of MT units, meaning $(\hat{x}, \hat{y})$ refers to the number of pixels along each linear dimension of an input frame.

To match the radial pattern $\Pi_{i,j,x,y}$ with the responses generated each directional cell in a MT+ macrocolumn, we created direction templates that select MT+ signals tuned to a particular direction when they appear within the sector of the visual field that is ‘appropriate’ for the radial expansion pattern. For example, the rightward motion template pools the responses of MT+ cells tuned to rightward motion when their RFs coincide with the right side of the visual field. The following equations define these direction templates $T_{d,i,j,x,y}$ that integrate MT+ cells tuned to direction d , normalized by the number of pooled cells $(\hat{x}\hat{y})$ (Equation 32):

$$\chi_{i,j,x,y} = \arctan(2(y - j, x - i)), \quad (30)$$

$$\tilde{T}_{d,i,j,x,y} = \begin{cases} 1 & \frac{(2d - 1 - \hat{d})\pi}{\hat{d}} < \chi_{i,j,x,y} < \frac{(2d - \hat{d})\pi}{\hat{d}} \\ 0 & \text{otherwise} \end{cases}, \quad (31)$$

$$T_{d,i,j,x,y} = \frac{\tilde{T}_{d,i,j,x,y}}{\hat{x}\hat{y}}. \quad (32)$$

Equation 32 resolves the direction component of the radial pattern match input to MSTd ($R_{i,j,s,h}$), which compares the direction templates ($T_{d,x_0,y_0,x,y}$) and the output of MT+ ($N_{x,y,d,s,h}^+$). Each direction template integrates MT+ responses with greater weight near the preferred singularity position (i, j) and the weights decrease exponentially with distance (Layton and Fajen, 2016a,b).

$$R_{i,j,s,h} = \sum_x \sum_y \sum_d e^{-b^{MSTd}((x-i)^2 + (y-j)^2)} T_{d,i,j,x,y} N_{x,y,d,s,h}^+. \quad (33)$$

Speed tuning (MT+ – MSTd)

Model MSTd contains units tuned for two distinct speed profiles. Speed summing cells respond proportionally to the average retinal speed (Inaba et al., 2007; Perrone, 2012). We modeled this speed sensitivity with a weighted sum of MT+ speed signals:

$$U(w_s) = \frac{1}{\hat{s}} \sum_s s w_s, \quad (34)$$

where the sum is normalized by the total number of preferred speeds in MT+ ($\hat{s}$). In Equation 32, we express the summation as a function (U) because speed summing cells could be tuned to either laminar ($U(L_{d,s,h})$) or radial ($U(R_{i,j,s,h})$) patterns.

On the other hand, bandpass cells maintain their sensitivity to the same average speed as MT+ units:

$$B_s(w_s) = w_s. \quad (35)$$

Disparity tuning (MT+ – MSTd)

Because the disparity sensitivity in MSTd need not correspond to that in MT, we applied functions to map MT+ signals to MSTd. To transform MT+ disparities into corresponding disparities in MSTd, we applied a Gaussian weighting function $G(h; \phi_k, \sigma_h^{MSTd})$ defined by Equation 18, where h indexes MT+ disparities, ϕ_k defines the k th MSTd disparity, and σ_h^{MSTd} controls the degree of pooling between MT+ disparities to resolve those in MSTd. To give a concrete example, we simulated five MT+ dispar-

ities ($h = 1, \dots, 5$) and three MSTd disparities ($\phi_k = 1, 3, 5$), where 1 is the nearest disparity and 5 is the farthest. Therefore, disparities in MT map to respective disparities in MSTd, but the tuning is coarser in this example. This mapping is realized with the following convolution:

$$Q(w_h; \phi_k, \sigma_h^{MSTd}) = q^{MSTd} \sum_h G(h; \phi_k, \sigma_h^{MSTd}) w_h, \quad (36)$$

where q^{MSTd} is the disparity gain. As the next section indicates, model disparity tuning is independent of speed (bandpass, summing) and pattern (radial, laminar) tuning. That is, the type of disparity tuning defined by Equation 36 corresponds to non-direction-dependent disparity cells that comprise most of the disparity tuned cells in MSTd (Yang et al., 2011).

Network dynamics

The input signal to MSTd depends on the unit's preferred pattern ($L_{d,s,h}$ for laminar, $R_{i,j,s,h}$ for radial), speed profile [$U(\cdot)$ for summing, $B_s(\cdot)$ for bandpass], and disparity ($Q(\cdot; \phi_k, \sigma_h^{MSTd})$). We created the signal by composing each of these signals. For example, we computed the input to bandpass cells tuned to radial patterns as $I_{i,j,s,k}^{BP, Rad} = Q(B_s(R_{i,j,s,h}); \phi_k, \sigma_h^{MSTd})$ where k indexes disparities in MSTd (ϕ_k).

MSTd units compete with one another in recurrent networks (Layton et al., 2012; Layton and Fajen, 2016b, 2017) that implement divisive interactions like those in Equation 4, but in differential equation form. We define the activity of bandpass cells ($P_{i,j,s,k}^{BP, Rad}$) tuned to radial patterns using the following equation:

$$\begin{aligned} \frac{dP_{i,j,s,k}^{BP, Rad}}{dt} = & -\alpha^{MSTd} P_{i,j,s,k}^{BP, Rad} + (1 - P_{i,j,s,k}^{BP, Rad}) \\ & \times (I_{i,j,s,k}^{BP, Rad} + Z(P_{i,j,s,k}^{BP, Rad}, \gamma^{MSTd}, \Gamma^{MSTd})), \\ & - P_{i,j,s,k}^{BP, Rad} \left(\sum_{n \neq i} \sum_{m \neq j} \sum_{o \neq s} \sum_{p \neq k} Z(P_{n,m,o,p}^{BP, Rad}, \gamma^{MSTd}, \Gamma^{MSTd}) \right), \end{aligned} \quad (37)$$

where α^{MSTd} defines the passive decay rate of each unit, $I_{i,j,s,k}^{BP, Rad}$ refers to the MSTd input signal, and $Z(P_{i,j,s,k}^{BP, Rad}, \gamma^{MSTd}, \Gamma^{MSTd})$ is the on-center/off-surround recurrent feedback MSTd units send one another. The last term involving four summands means that each unit receives inhibition from units tuned for the same type of pattern (e.g., radial) and either have a different RF position (n, m), speed tuning (o), or disparity tuning (p). The recurrent feedback is defined by the sigmoid function:

$$Z(w; \gamma, \Gamma) = \frac{(w - \Gamma)^2}{\gamma^2 + (w - \Gamma)^2}, \quad (38)$$

where w is the activity of the unit sending the feedback signal, γ defines the inflection point of the sigmoid, and Γ represents the activity threshold that must be overcome to send the feedback signal. The sigmoid recurrent feedback function allows MSTd units to compete with one another to resolve winners without necessarily eliminating all the other weaker activity across the network (i.e., soft winner-take-all; Grossberg, 1973, 2013).

The following equation describes the dynamics of speed summing cells tuned to radial patterns, which is analogous to Equation 37:

$$\begin{aligned} \frac{dP_{i,j,k}^{SS, Rad}}{dt} = & -\alpha^{MSTd} P_{i,j,k}^{SS, Rad} + (1 - P_{i,j,k}^{SS, Rad}) \\ & \times (I_{i,j,k}^{SS, Rad} + Z(P_{i,j,k}^{SS, Rad}, \gamma^{MSTd}, \Gamma^{MSTd})) \\ & - P_{i,j,k}^{SS, Rad} \left(\sum_{n \neq i} \sum_{m \neq j} \sum_{p \neq k} Z(P_{n,m,p}^{SS, Rad}, \gamma^{MSTd}, \Gamma^{MSTd}) \right). \end{aligned} \quad (39)$$

The following two equations adapt Equations 35 and 37 to describe the dynamics of bandpass and speed summing cells tuned to laminar patterns, respectively. Note that the only differences are the addition of the direction subscript (d) and absence of spatial subscripts (i, j).

$$\frac{dP_{d,s,k}^{BP,Lam}}{dt} = -\alpha^{MSTd} P_{d,s,k}^{BP,Lam} + (1 - P_{d,s,k}^{BP,Lam}) \times (I_{d,s,k}^{BP,Lam} + Z(P_{d,s,k}^{BP,Lam}, \gamma^{MSTd}, \Gamma^{MSTd})) - P_{d,s,k}^{BP,Lam} \left(\sum_{n \neq d} \sum_{o \neq s} \sum_{p \neq k} Z(P_{n,o,p}^{BP,Lam}, \gamma^{MSTd}, \Gamma^{MSTd}) \right), \quad (40)$$

$$\frac{dP_{d,k}^{SS,Lam}}{dt} = -\alpha^{MSTd} P_{d,k}^{SS,Lam} + (1 - P_{d,k}^{SS,Lam}) \times (I_{d,k}^{SS,Lam} + Z(P_{d,k}^{SS,Lam}, \gamma^{MSTd}, \Gamma^{MSTd})) - P_{d,k}^{SS,Lam} \left(\sum_{n \neq d} \sum_{p \neq k} Z(P_{n,p}^{SS,Lam}, \gamma^{MSTd}, \Gamma^{MSTd}) \right). \quad (41)$$

Feedback (MSTd – MT–, MSTd – MSTv)

MSTd units send feedback to suppress cells in MT– and MSTv tuned to directions, speeds, and disparities that are locally consistent with the global pattern to which the active MSTd cell prefers. In our implementation of the model, we assume that only the maximally active cells tuned to each pattern and each disparity send feedback. If MSTd neurons are tuned to either near, fixational, and far disparities, this means that a total of 12 units send feedback: three bandpass radial units, three speed summing radial units, three laminar radial units, and three laminar speed summing units. The feedback described here implements the mechanism introduced above (see Self-motion and object motion interaction; Figs. 3, 4a).

The MSTd direction feedback weighting obeys the following Gaussian function:

$$W(w; \mu, \sigma) = e^{-\left(\frac{w - \mu}{\sigma}\right)^2}, \quad (42)$$

evaluated at $W(\theta_d; \chi_{i^*, j^*, x, y}, b_d^{FB})$ for MSTd cells tuned radial patterns and $W(\theta_d; \theta_d, b_d^{FB})$, where (i^*, j^*) refers to the singularity position of the maximally active radial MSTd unit, d^* refers to the direction preference of the maximally active laminar MSTd unit, θ_d represents the tuning directions in MT, and b_d^{FB} controls the extent to which the feedback signal affects MT–/MSTv units with similar direction preferences. The maximum weight in Equation 42 is 1 because only the most active MSTd unit tuned to each pattern type may send feedback.

We compute the feedback from bandpass ($K_{x,y,d,s,h}^{BP}$) and speed summing ($K_{x,y,d,s,h}^{SS}$) MSTd units, either tuned to radial ($K_{x,y,d,s,h}^{BP,Rad}$ and $K_{x,y,d,s,h}^{SS,Rad}$) or laminar ($K_{x,y,d,s,h}^{BP,Lam}$ and $K_{x,y,d,s,h}^{SS,Lam}$) motion patterns, as follows:

$$K_{x,y,d,s,h}^{BP,Rad} = [W(\theta_d; \chi_{i^*, j^*, x, y}, b_d^{FB})][W(s; s^*, b_s^{FB})][W(h; k^*, b_h^{FB})], \quad (43)$$

$$K_{x,y,d,s,h}^{SS,Rad} = W(\theta_d; \chi_{i^*, j^*, x, y}, b_d^{FB})W(h; k^*, b_h^{FB}), \quad (44)$$

$$K_{x,y,d,s,h}^{BP,Lam} = [W(\theta_d; \theta_d, b_d^{FB})][W(s; s^*, b_s^{FB})][W(h; k^*, b_h^{FB})], \quad (45)$$

$$K_{x,y,d,s,h}^{SS,Lam} = W(\theta_d; \theta_d, b_d^{FB})W(h; k^*, b_h^{FB}), \quad (46)$$

where s^* indicates the preferred speed of the most active MSTd unit and k^* indicates that unit's preferred disparity. Note from Equations 43–46 that speed-summing cells send equivalent feedback, regardless of the speed tuning of the recipient cell, whereas bandpass cells concentrate the feedback weight around recipient cells tuned to similar preferred speeds. Consistent with the feedforward direction, recipient cells receive feedback from MSTd cells tuned to corresponding disparities (e.g., near-to-near, far-to-far).

The feedback signal that reaches recipient cells in MT–/MSTv is the sum of Equations 43–46:

$$K_{x,y,d,s,h}^{MSTd} = \sum_n \omega_n H[\omega_n - \Gamma^{FB}] \vartheta_n, \quad (47)$$

where $\omega = \{P_{i,j,s,k}^{BP,Rad}, P_{i,j,s,k}^{SS,Rad}, P_{d,s,k}^{BP,Lam}, P_{d,s,k}^{SS,Lam}\}$, $\vartheta = \{K_{x,y,d,s,h}^{BP,Rad}, K_{x,y,d,s,h}^{SS,Rad}, K_{x,y,d,s,h}^{BP,Lam}, K_{x,y,d,s,h}^{SS,Lam}\}$, $H[\cdot]$ represents the Heaviside step function, and Γ^{FB} indicates the activity threshold that each type of MSTd unit must exceed to send feedback.

MT layer 2–3 (MT–)

MT– units perform an on-center/off-surround integration of MT L4,6 signals. We obtain the center input by evaluating the 2D convolution function $C(\cdot)$ (Eq. 23): $C_{x,y,d,s,h,t}^- = C(M_{x,y,d,s,h,t}^{4,6}, \gamma^{MT-cent}, g^{MT-cent})$, which mimics the MT+ center input (Eq. 24).

To resolve the spatial component of the inhibitory surround, we evaluate Equation 23 at $S_{x,y,d,s,h,t}^{MT-1} = C(M_{x,y,d,s,h,t}^{4,6}, \gamma^{MT-surr}, g^{MT-surr})$, where $\sigma^{MT-surr}$ controls the spatial extent of the integration with $\sigma^{MT-surr} > \sigma^{MT-cent}$, and $g^{MT-surr}$ is the filter gain. We pass the output $S_{x,y,d,s,h,t}^{MT-1}$ of this convolution through a von Mises filter that defines the directional selectivity of the surround:

$$S_{x,y,d,s,h,t}^{MT-2} = g_d^- V(S_{x,y,d,s,h,t}^{MT-1}; \theta_d, b_d^-), \quad (48)$$

where b_d^- controls the direction selectivity of the surround tuning, analogously to σ_{sd} (see Surround suppression from local motion signals), and g_d^- is the filter gain. The function $D(\cdot; \theta_d, \sigma_d^-)$ refers to the following 1D Gaussian filter convolution:

$$D(w_d; \mu, \sigma) = \sum_n G(n; \mu, \sigma) w_{d-n}. \quad (49)$$

We evaluate the disparity component of the feedforward surround inhibitory signal by performing a convolution with the disparity filters:

$$S_{x,y,d,s,h,t}^{MT-3} = D(S_{x,y,d,s,h,t}^{MT-2}; h, \sigma_h^-), \quad (50)$$

where h indexes MT disparities and σ_h^- determines the disparity selectivity of the surround.

MT– dynamics obey the following on-center/off-surround network:

$$\frac{dM_{x,y,d,s,h}^-}{dt} = -M_{x,y,d,s,h}^- + (1 - M_{x,y,d,s,h}^-) C_{x,y,d,s,h,t}^- - (M_{x,y,d,s,h}^- + \beta^-) (K_{x,y,d,s,h}^{MSTd} + S_{x,y,d,s,h,t}^{MT-3}), \quad (51)$$

where β^- specifies the hyperpolarizing lower bound of each unit's activation when suppressed and $K_{x,y,d,s,h}^{MSTd}$ refers to feedback from MSTd.

MSTv

MSTv units perform an on-center/off-surround integration of signals from MT–, much in the same way MT– integrates feedforward signals from MT L4,6. Aside from larger RFs, the main difference is that MSTv cells respond proportionally to the average speed, like speed summing cells in MSTd. To model this, we perform a spatial integration of rectified MT– signals, weighted by their speed in both center and surround:

$$C_{x,y,d,h}^v = C(U(\lceil M_{x,y,d,s,h}^- \rceil^+); x, y, \sigma^{v,cent}, g^{v,cent}), \quad (52)$$

$$S_{x,y,d,h}^{v,1} = C(U(\lceil M_{x,y,d,s,h}^- \rceil^+); x, y, \sigma^{v,surr}, g^{v,surr}), \quad (53)$$

where $U(\cdot)$ is the weighted sum defined by Equation 34. We filtered the surround output $S_{x,y,d,h}^{v,1}$ by the von Mises direction filter V to establish direction selectivity:

$$S_{x,y,d,h}^{v,2} = g_d^v V(S_{x,y,d,h}^{v,1}; \theta_d, b_d^v). \quad (54)$$

We configured the surrounds as untuned for disparity; motion at any disparity inhibits MSTv units when it appears in the surround:

$$S_{x,y,d,h}^{v,3} = q^{MSTv} \sum_{p \neq h} S_{x,y,d,p}^{v,2} \quad (55)$$

where, q^{MSTv} denotes the surround disparity gain.

The following equation describes the network dynamics of MSTv units ($P_{x,y,d,h}^{v}$):

$$\frac{dP_{x,y,d,h}^v}{dt} = -P_{x,y,d,h}^v + (1 - P_{x,y,d,h}^v)C_{x,y,d,h}^v - (P_{x,y,d,h}^v + \beta^{MSTv})(K_{x,y,d,s,h}^{MSTd} + S_{x,y,d,h}^{v,3}). \quad (56)$$

Results

Our approach in this study is to simulate the model (or its individual mechanisms) to inform decisions about model development, to better understand how specific mechanisms behave and how they contribute to self-motion and/or object motion perception, and to compare model behavior to that of human observers. We organized the forthcoming six sets of analyses as follows. In the first two sections, we identify and simulate plausible mechanisms for perceptual stability. The focus is on direction of motion rather than speed and/or disparity to better understand each mechanism's properties and to inform subsequent model development. We then embed the mechanisms into a larger model system to investigate their interactions when dynamically processing disparity, speed, and direction (motion parallax) signals (Fig. 2*a*). In the fourth and fifth sections, we focus on the influence that depth tuning may have on object motion perception during self-motion and the detection of moving objects, respectively.

Mechanisms for establishing perceptual stability: divisive or subtractive interactions

Area MSTd is a reasonable candidate for mediating the estimation and suppression of the visual pattern arising from the observer's self-motion to establish perceptual stability given the RFs in many cells that span much of the visual field (Duffy and Wurtz, 1991), causal link to heading perception (Gu et al., 2012), and sensitivity to optic flow patterns generated by self-motion (also eye/head movements; Duffy, 1998; Liu and Angelaki, 2009). Importantly, with their direction and disparity tuning (Yang et al., 2011), MSTd neurons are capable of taking depth into account when integrating the self-motion component from retinal motion signals.

MSTd is reciprocally connected with MT, an area in which virtually all neurons are motion-tuned and unlike MSTd, have small enough RFs to respond to object motion. This makes MT a plausible brain region where the suppression of the motion pattern that arises from the observer's self-motion could begin to take place. In addition to neurophysiological studies that show contextual modulation in MT neurons based on motion surrounding the classical RF (Allman et al., 1985; Tanaka et al., 1986), preliminary evidence directly shows suppression in responses to small moving objects and direction modulation when RFs are surrounded by the type of radial motion experienced by a moving observer (Peltier et al., 2016).

For MT and MSTd to be involved in stabilizing perception during self-motion, given known properties of neurons contained therein, MSTd neurons could send a feedback signal that reflects the motion pattern most consistent with the observer's self-motion. We consider the possibility that this feedback signal exerts a suppressive influence on MT neurons with RFs distributed across the visual field. We illustrate how this feedback mechanism works in Figure 4*a*, top, left; a radial pattern is generated by an observer moving straight forward along a central heading in the presence of an object, moving upward on the retina, but up-and-to-the-left relative to the stationary world (as in Fig. 1*a*). MSTd neurons tuned to the pattern of radial expansion send feedback to MT neurons that integrate the local direction throughout the field (purple arrows, top, left). The purple Gauss-

ian profiles at the bottom show these feedforward local motion signals in two specific regions of the visual field. The orange Gaussian curves show the motion signals that correspond to the preferred direction of the MSTd neuron sending the feedback to MT locally within the MT unit's RF. The purple (retinal) and orange (feedback) curves may have a high degree of overlap, such as when the motion arises from observer movement relative to the stationary world (bottom, right), or much less overlap, such as when the motion arises from an independently moving object (bottom, left). To establish perceptual stability, feedback signals (orange curves) from MSTd could suppress responses to the different local directions distributed across the visual field. This suppression would be most effective when the orange feedback distributions match (i.e., overlap maximally) with the purple sensory input distributions (bottom, right).

In principle, there are several ways in which suppression of the self-motion component from MT responses could be achieved. As a starting point, we first simulated interactions between MT and MSTd signals using a simple divisive model (Eq. 4). According to this model, each unit within an interacting neural population receives driving excitatory input consistent with its tuning properties and inhibition normalizes the activity across the population. Figure 4*b* shows changes in the direction signaled by the MT population corresponding to a moving object as a function of different degrees of overlap between the purple ("Retinal motion") and orange ("Self-motion") distributions in Figure 4*a*. Focusing on the case wherein the broadness (SD σ) of both the feedforward and feedback direction distributions is 25° (red curve), Figure 4*b* shows that the object direction is unaffected when the distributions perfectly overlap (0° mismatch) or hardly overlap (180° mismatch). When feedback does overlap with the feedforward distribution originating from the object (Fig. 4*a*, bottom, left), interactions modulate the object direction represented across the population, shifting it toward the world-relative direction (Fig. 4*b*, positive shifts). Simulating the same model with different σ values (Fig. 4*b*, other colored curves) reveals the same qualitative behavior, indicating the pattern of results does not depend on a specific parameter value.

Following the suggestion from human studies that the brain "subtracts" out the global pattern (Warren and Rushton, 2007), we also simulated a model with subtractive (Eq. 7) rather than divisive interactions. The subtractive model differs in that inhibition subtracts from (rather than scales) the firing rate of the neuron (Mejias et al., 2014). This decreases the response of a neuron through a direct shift, without necessarily balancing activity across the population. The two models demonstrate qualitatively similar effects (Fig. 4*b,c*). They differ most substantially in their predictions about the broadness of the local MT population direction distributions, which corresponds to the variance of the purple and orange Gaussians depicted in Figure 4*a*, bottom. For comparably strong feedforward and feedback signals (e.g., Gaussian peaks in of Fig. 4*a*, bottom, have approximately the same height), the divisive model exerts a modulatory, sharpening effect (Fig. 4*d*, inset plot). This means that the model MT direction response after the purple and orange Gaussians interact would be narrower than any one of the original signals, which can be seen by inspecting the variance plotted in Figure 4*d*. For example, in Figure 4*d*, the red curve corresponds to a Gaussian object motion signal with $\sigma = 25^\circ$. The curve plateaus at the dashed horizontal line at $y = 25^\circ$ when there is little suppression and at $y = 5^\circ$ when suppression is maximal. Suppression is greatest when the feedforward and feedback signals match (0° mismatch; Fig. 4*a*, bottom, right), which has the effect of squar-

ing (i.e., narrowing) the direction distribution. Hence, the SD of the red curve ($\sigma = 25^\circ$) shrinks to, but does not drop $<5^\circ$. On the other hand, the subtractive model can completely suppress, or at least greatly diminish, the output signal when the feedforward and feedback signals overlap (e.g., when the moving object moves along a similar direction relative to the background self-motion pattern). This can be seen by inspecting how the SD of the output direction signal shrinks to 0° when the interacting signals match in Figure 4e (also see inset plot).

The property of having the gain and broadness of the output direction signal not drop below a minimum value in the divisive interaction model may offer computational benefits for maintaining acute sensitivity to moving objects while establishing perceptual stability, even when objects may move along similar trajectories to the background motion. The complete suppression of the object motion signal in the subtractive model may be problematic because it implies that in certain circumstances the object would not be perceived. On the other hand, the subtractive model does enable more precise signaling of the object motion direction (smaller minimum SD; Fig. 4e). For these reasons, we adopt a hybrid approach: a mostly divisive inhibitory interaction model in subsequent simulations with a weaker subtractive component.

Local center-surround influences

Feedback may not be solely responsible for modulating the object motion direction toward a world-relative reference frame. Local center-surround mechanisms within MT could also contribute. Indeed, a large, distinct subpopulation of MT cells (MT $-$) possess suppressive surrounds tuned to direction, speed, and disparity that matches the center region of the RF (Born and Bradley, 2005). They tend to be quiescent due to strong surround suppression when the same signal stimulates both center and surround; only when the direction, speed, and/or disparity in the surround mismatch the center does the cell fire vigorously. For example, one particular cell may respond best when slow speeds occupy the center and fast speeds occupy the surround. In general, the suppression falls off as the discrepancy from the preferred combination in any one of these dimensions grows (Xiao et al., 1997a,b, 1998; Born and Bradley, 2005). MT surrounds likely have a tuned inhibitory component derived from feedforward or intra-areal signals (Cui et al., 2013), distinct from feedback because the stimuli used to measure surround properties often only locally stimulate the MT RF and do not involve the global motion patterns that resemble those experienced during self-motion.

We performed simulations to test the extent to which local non-feedback-mediated MT surround suppression (e.g., inhibition from feedforward connections) could influence direction signals among MT cells with matching center-surround direction tuning. We focused on a scenario wherein the observer moves straight forward along a central heading toward a distant wall in the presence of a small object that moves rightward in the periphery of the visual field (Fig. 5a). We used a simple divisive MT model (Eq. 13) to quantify the influence that center-surround interactions have on object motion signals (rightward optical direction), as the local surround direction varied in the radial pattern along an iso-eccentricity circle (Fig. 5a). In Figure 5b, we varied the broadness of the center and surround RF direction tuning curves (e.g., broadness of the Gaussians in Fig. 4a, bottom). Focusing a unit that has equally broad center and surround direction tuning (orange curve; $\sigma_c/\sigma_s = 1$), responses were weakest when the same (0°) or very different (135°) motion di-

rections stimulated the center and surround. MT responses peaked somewhere in between these two extremes, depending on the ratio of center-surround direction tuning bandwidth. For units with a much narrower center direction tuning curve than the surround (e.g., $\sigma_c/\sigma_s = 0.75$), the response peaked when there was a small difference between the motion directions stimulating the center and surround. The peak shifted toward larger directional differences when the surround was relatively narrower [e.g., ($\sigma_c/\sigma_s = 1.25$)]. In most cases, shifts are toward the world-relative direction (positive).

The shift increased proportionally to the gain of the inhibitory surround tuning curve (Fig. 5c). A higher gain in surround tuning curve would resemble the peak of the orange curve Figure 4, bottom, if it were higher than that of the purple curve. Figure 5d shows the basis of the shift in one particular simulation, when the output of the center RF Gaussian filter coincides with rightward motion of the moving object (0° ; top) and the surround filter output deviates (middle). The bottom shows the shifted response distribution (i.e., shifted to the left), predicted by the divisive interaction model. This suggests that at least for motion patterns that arise during linear self-motion, despite the limited extent of MT RFs, the local mechanisms (e.g., within MT) could contribute, together with feedback that signals the self-motion estimate (e.g., from MSTd), to establish perceptual stability.

Suppression of self-motion component from object motion signals and spatiotemporal dynamics of object motion signals

Figures 4 and 5 illustrate how surround suppression arising from feedback carrying the self-motion signal and local inhibition could modulate the direction of object motion signals. To appreciate the joint influence on perceptual stability during self-motion, we developed a model of MT and MST that accounts for direction, speed, and disparity tuning, more elaborate than the simple direction-focused simulations described up until now. The model includes “wide-field” MT units with summing surrounds (MT $+$; Tanaka et al., 1986), and units with tuned suppressive surrounds (MT $-$; Born and Bradley, 2005). Our model contains distinct MT $+$ /MSTd (Born and Tootell, 1992; Born, 2000; Yu et al., 2018) and MT $-$ /MSTv (Thier and Erickson, 1992; Tanaka et al., 1993) pathways: the former pathway specializes in motion pooling, self-motion estimation, and detecting global patterns, whereas the latter is suited for signaling the motion of objects that move differently than their background (see Materials and Methods).

The division of areas MT and MST into functionally distinct pathways could lie at the heart of how the visual system establishes perceptual stability during self-motion. Although the presence of cells in the MT $-$ /MSTv pathway that selectively respond to object motion has been well established (Tanaka et al., 1993), the reference frame of these responses has not yet been examined and could shed light onto the strategy by which the visual system disambiguates the retinal motion field: signals in an observer-relative (retinal) reference frame would indicate a mere segmentation of moving objects, whereas shifts toward a world-relative reference frame would suggest the recovery of parameters related to the object's 3D movement, independent of the observer. We therefore focused our investigation on the frame of reference of object motion signals as they propagate through MT and MST. Given how MSTd feedback and local inhibition may shape the resultant direction through surround suppression from lateral,

Figure 6. Temporal dynamics in joint motion parallax and disparity model. *a*, Schematic of the observer's movement through the 3D volume of dots, with the object moving up and to the left through the world, along with the retinal radial flow. The stimulus is a stereo version of the Full condition used by Warren and Rushton (2009). *b*, Temporal dynamics of the self-motion signal in MSTd in the Stereo Full condition. The distribution of activity across MSTd reflects the confidence in the heading estimate. For example, broad activity distributed across MSTd (colors other than dark pink) indicates uncertainty in the heading direction. *c*, The temporal dynamics of the object motion signal in MT– and MSTv in the Stereo Full condition. Positive values indicate shifts toward the world-relative object motion direction (CCW). *d*, Input signals that MSTv units integrate that have the object within their RFs and inhibitory distributions from MSTd feedback and surround inhibition are superimposed. The x-axis corresponds to MT–/MSTv direction tuning in units whose local RF is centered on the object. *e*, Direction distribution shift over successive snapshots in time as feedback and surround inhibition interact with motion signals in MSTv. The blue dashed vertical line marks the early MSTv activity peak position (retinal direction) and the black dashed vertical line marks the object world-relative direction. Same axes conventions as *d*.

feedforward, and feedback signals, we simulated how activity in the MT–MST system dynamically unfolds over time.

Figure 6 depicts the temporal evolution of MT–MST signals generated by the model for a scenario that involves forward, straight-ahead simulated self-motion through a 3D cloud of dots viewed in stereo in the presence of a small moving object. Although the object's retinal motion is upward (i.e., in an observer relative reference frame), the world-relative movement is up-and-to-the-left (Fig. 6*a*). Figure 6*b* shows the progression in MSTd cell population tuned to radial expansion (forward self-motion) and the same disparity as the moving object. Model MSTd units integrate evidence about the observer's heading direction over time and competition suppresses signals incompatible with the evolving estimate. This reduces the variance across the MSTd distribution as units estimate the observer's self-motion with greater confidence. The ~200 ms model MSTd peak latency is shorter than the 300–500 ms of optic flow duration required until the accuracy of human heading judgments stabilizes (Layton and Fajen, 2016c), which is consistent with the idea that MSTd neurons contribute to heading perception (Britten and van Wezel, 2002; Gu et al., 2012). The time course of MSTd neurons varies considerably, but model unit dynamics are compatible with physiological response latencies (e.g., 120–440 ms interquartile range; Lagae et al., 1994).

As MSTd units send feedback to MT– and MSTv to signal the observer's self-motion, the direction signaled by populations responding to the object motion begins to shift from retinal to world-relative reference frames (Fig. 6*c*). To understand why this

transformation occurs, we plotted in Figure 6*d* the distribution of feedforward direction signals among MT– units (blue) that integrate object motion (i.e., the object appears within their RF) alongside the distributions that represent the inhibition that each respective unit receives from MSTd feedback (red) and local surround interactions (green). Figure 6*e* shows how these two inhibitory signals change the shape of the snapshots of MSTv activity distributions over time, which in turn influences the direction signaled by the population (shift to the left). Although early activity is aligned with the object's retinal direction (blue dashed vertical line), it shifts toward the world-relative direction. Similar dynamics and shifts toward a world-relative reference frame arise in MT–, through the joint influence of MSTd feedback and local surround inhibition, analogous to the green and red curves in Figure 6*d*. The shift is greater in MSTv than MT– because we model area MSTv as one level higher than MT and the direction shifts compound from one area to the next through both feedforward and feedback connections. Direction modulation may occur shortly following the motion onset (e.g., in MSTv) because we do not explicitly model a time delay between excitatory or inhibitory signals.

To examine the relative contribution of feedback and non-feedback sources of surround suppression in the modulation of direction, we introduced a circular mask around the moving object to remove a region of retinal motion. This manipulation eliminates the influence of either MT non-feedback-mediated surround suppression (i.e., from local inhibitory signals) when the motion inside the circle is removed (Fig. 7*a*, Stereo Global) or

Figure 7. Simulations of Stereo Global and Local conditions. *a*, Temporal dynamics of object motion signals in the Stereo Global condition. Positive values indicate shifts toward the world-relative object motion direction (CCW). *b*, Temporal dynamics of object motion signals in the Stereo Local condition. Positive values indicate shifts toward the world-relative object motion direction (CCW).

Figure 8. Simulations of monocular versions of Full, Global, and Local conditions from Warren and Rushton (2009).

feedback when motion when the motion outside the circular mask is removed (Fig. 7*b*, Stereo Local). For example, shifts in the Stereo Local condition (Fig. 7*b*) arise entirely due to local, non-feedback-mediated surround interactions in MT– and MSTv. Notice how the direction shift greatly diminishes when the mask removes the global surrounding pattern (Fig. 7*b*) compared with when it removes the local motion (Fig. 7*a*). Together, these mask manipulations show that MSTd feedback is responsible for a large proportion of the influence in the direction of object motion signals, consistent with the ordinal difference in gains of the inhibitory distributions shown in Figure 6*d* (red vs green).

The large contribution of the MSTd feedback mechanism suggests a reliance on the global motion pattern to suppress the self-motion component. There was no requirement for this to be the case; the suppression could instead heavily rely on the local surround suppression signals. To assess the plausibility that the MSTd feedback mechanism has a greater contribution than local center-surround interactions, we compared model-derived object motion direction estimates with human judgments about object trajectory reported by Warren and Rushton (2009). Their displays resemble those used in Figures 6 and 7, except the subjects viewed monocular displays. To facilitate comparison, we performed the simulations again, but this time without disparity sensitivity throughout the model. We obtained a model object-direction estimate by plotting the direction signaled by the MSTv population at the end of each stimulus presentation (Fig. 8). This analysis shows that model estimates of object direction closely correspond to human judgments, which suggests a dominant role of a feedback-based mechanism in MT surround suppression that may arise to disambiguate retinal motion during self-motion. Interestingly, when we ran the simulations using displays

with disparity (Fig. 6*c*), the shift was ~10% greater compared with the monocular version shown in Figure 8.

Influence of disparity tuning

In this section, we use the MT–MST model system to explore the potential role of disparity within neural mechanisms involved in perceptual stability. We focused on how surround suppression may influence the direction estimates of a moving object positioned at different depths. To constrain model disparity tuning, we selected MT disparity tuning curves for which sensitivity remains constant across depth (i.e., constant variance disparity tuning curves; Fig. 2*d*). For comparison, we considered an alternative based on the weak tendency for disparity sensitivity to become coarser with depth (DeAngelis and Uka, 2003), which we call the “increasing variance model”. Because we normalized each disparity tuning curve, the peak tuning gain decreases with depth (Fig. 9*a*), which implies increased uncertainty about depth at farther disparities.

We tested both disparity tuning variants by simulating the model with the stereo displays from Warren and Rushton (2007) to compare model-derived direction estimates with human object trajectory judgments. An object moved upward at a fixed speed through the world and maintained one of three depths from the observer. Because of motion parallax, the optical speed decreased the further it was positioned from the observer. The two main conditions are schematized in Figure 9, *b* and *c*: the Rotation (Fig. 9*b*) and Translation (Fig. 9*c*) conditions. The Rotation condition simulates pure lateral rotation about the point of observation, as may arise during certain horizontal eye movements. Rotation creates a global optical motion pattern that is independent of the depth of points in the environment (Fig. 9*b*, top). Hence, if the visual system stabilizes perception during rotational movement, the amount of suppression should be constant across depth. Because the object’s optical speed decreases with depth, the relative effect of this suppression increases as the object is moved farther away. As such, the shift in the object direction signal should grow with depth. Indeed, such an increasing trend is evident in the human object motion judgments from Warren and Rushton (2007; Fig. 9*d*, red solid curve).

Object direction estimates decoded from model MSTv units also capture this increasing trend (Fig. 9*d*, red dashed curve). Because rotation creates similar speed and direction patterns across depth due to the observer’s self-motion (Fig. 9*b*, top), the stimulus activates MSTd units regardless of their preferred disparity (i.e., at all depths). MSTd sends feedback to suppress MT–/MSTv units tuned to corresponding disparities (e.g., near–near, far–far), which leads to different amounts of suppression in

Figure 9. Effect of disparity tuning on object motion signals. **a**, Disparity tuning curves in the increasing variance model. **b**, The Warren and Rushton (2007) Full Volume Rotation condition. **c**, The Warren and Rushton (2007) Full Volume Translation condition. **d**, The Warren and Rushton (2007) Full Volume Translation and Rotation condition data and simulations. **e**, Simulations of Near, Middle, and Far depth conditions from Warren and Rushton (2007).

the MT object motion signal, depending on the object's depth in the scene. In the model, the closer the object's speed is to the background, the greater the directional modulation (Fig. 3*b*).

The other main condition, Translation (Fig. 9*c*), simulates rightward lateral translation of the observer, which results in motion parallax: a decreasing speed gradient in depth with faster speeds arising at nearby depths, all moving in the same direction (Fig. 9*c*, bottom). If the visual system stabilizes perception during translational movement, the amount of suppression should decrease with depth. Because the moving object is subject to the same motion parallax as the stationary elements of the scene, shifts in the object motion signal should be constant across depth. Although the human judgments from Warren and Rushton (2007) do not demonstrate this ideal flat curve, the slope is weaker than in the Rotation condition.

Model-derived object direction estimates also exhibited a weaker slope in the Translation condition than in the Rotation condition. This occurred because changes in object depth correlated with the speed preference of MSTd units suppressing the self-motion pattern in the moving object signal. Surround suppression therefore exerted a roughly constant effect on the shift across depth (i.e., the inhibition sent to MT-/MSTv tracks the main diagonal in Fig. 3*a*).

Surprisingly, we found that varying the disparity filter tuning properties only had a modest effect on the shift in the object direction signals garnered in both conditions (Fig. 9*d*, dotted

curves). We therefore focused on the constant variance disparity tuning curves for remaining simulations.

Warren and Rushton (2007) also measured human object motion judgments in the Rotation and Translation conditions in conditions where dots only occupy the nearest (Near), middle (Middle), and farthest (Far) third of the 3D volume (Warren and Rushton, 2007, their Experiment 3). The object moves at its designated depth, whether or not dots in the volume also occupy the same depth. We simulated these variable depth conditions because they recruit local- and feedback-based disparity dependent suppression in complex ways and to provide a more challenging test for the model. Starting with the Rotation condition, shifts in the object direction derived from the model increased with depth (Fig. 9*e*, dashed red curves), consistent with the human data from the Full condition and in the Near, Middle, and Far conditions (solid red curves). Again, MSTd mediated suppression exerted greater modulation for slower, more distant objects. In the Translation condition, the model-derived shift remained fairly constant in the Near and Middle conditions, but showed a clear positive slope in the Far condition, consistent with human judgments. This transition happens in the model because the slow speeds in the Rotation and Translation conditions stimulate far disparity MSTd units responsible for the suppression to roughly the same extent. Together, the Rotation and Translations simulations summarized in Figure 9 demonstrates that model MT-MST dynamics between direction, speed, and disparity signals

Figure 10. Detecting moving objects from model MSTv signals. **a**, Monocular and stereo object motion detection compared with the Royden and Connors (2010) data (blue curve). **b**, Simulation showing model MSTv units detecting a moving object based on disparity alone. The object possessed a speed and direction consistent with movement as part of the background scene (110 cm), but we independently manipulated the object depth (x-axis). The background 3D dot volume was positioned at 110 cm and further depths (shaded region). Curves show differences between MSTv unit responses to moving object and those to the nearest background dot at three different object eccentricities. Positive y values indicate how well model MSTv units detect the object solely based on disparity differences.

capture important characteristics of human object motion judgments during self-motion.

The role of joint motion parallax and disparity tuning in detecting moving objects

Suppressing the observer's estimated self-motion to disambiguate the retinal motion field using motion parallax and disparity also provides a means to resolve whether an object is moving or stationary. This is because mechanisms that suppress the observer's estimated world-relative motion silence responses to stationary objects, because they generate motion parallax consistent with world-fixed elements. For this solution to object detection to be feasible in general, the mechanisms should take both motion parallax and disparity into account. Otherwise, a close stationary object might be considered moving because of its fast optical speed when juxtaposed with distant background elements. We hypothesized that the joint motion parallax and disparity tuning in MT and MST neurons would allow the feedback and non-feedback mediated suppression mechanisms to detect the presence of moving objects. To test this, we simulated forward

self-motion in the presence of an object that initially moved along the same direction as the background pattern (Fig. 10a, inset). For different optical speeds, we determined the minimal angular deviation required for the object motion to exceed a fixed activity detection threshold in model MSTv.

Figure 10a summarizes the monocular moving object detection threshold angles decoded from model MSTv activity. The threshold defines the smallest angular deviation of a small object moving relative to the surrounding optic flow pattern to yield a suprathreshold object motion response (see Materials and Methods, Direction threshold object detection conditions). We compared the model derived thresholds with those from Royden and Connors (2010), a psychophysical study wherein subjects performed a similar monocular object detection task. Thresholding model MSTv object motion signals to detect the object's presence closely resembles the human data, which suggests that the model MSTv activity can account for human performance. We also considered a stereo version of the display to assess how disparity could influence moving object detection by a mobile observer. Figure 10 shows lower thresholds, which indicates that depth improves the model's ability to detect moving objects.

We challenged the model's object detection capabilities by investigating whether MSTv units can detect a moving object on the basis of disparity alone. In the simulation depicted in Figure 10b, the object moved with a direction and speed that matched that of the stationary background (at 110 cm), but we independently manipulated depth. For different object depths in front of and within the 3D cloud

of stationary elements in the scene, we examined the activity difference between MSTv units responding to the moving object and to those responding to the nearby background. We interpreted any value greater than zero as an indication that object motion was detected. Figure 10b shows that model MSTv units generate consistently greater activation to the object across depth, even though its direction and speed matches the nearby background. The periodic fluctuations in depth arise due to the limited number of disparity tuning curves (Fig. 2d), and the sudden drop near 110 cm occurs due to suppression from the high degree of similarity to the background (Fig. 3). Together, these simulations show that the model leverages disparity to detect the presence of moving objects; even when disparity is the only differentiating factor. The model produces human-like object detection thresholds that improve when disparity information is available.

Discussion

Although neurophysiological and psychophysical studies have indicated that the visual system disambiguates self-motion from

object motion in the retinal motion field, the underlying computational mechanisms had previously not been examined. We developed two related mechanisms that use surround suppression (one that operates globally and the other locally) to segregate self-motion and object motion signals along separate neural pathways. Our analysis serves to improve our understanding of the mechanisms that underlie how the visual system establishes perceptual stability during self-motion.

Here is a summary of key predictions derived from model mechanisms and ideas for experiments to support or falsify them:

- Feedback suppresses MT−/MSTv neurons with direction, speed, and disparity tuning that is compatible with the preferred RF pattern of the most active MSTd neurons (Mechanism 1). This could be tested by characterizing the direction tuning curve of MT or MST neurons with antagonistic surrounds; both in the presence (“Global-Pattern” condition) and absence (“No-Global-Pattern” condition) of a global (e.g., radial) pattern that surrounds the RF. In the Global-Pattern condition, individual neurons should demonstrate suppression around the preferred direction established in the No-Pattern condition, particularly if it is similar to the nearby direction of the global pattern, whereas others should show enhancement at a non-preferred direction. These effects should modulate the direction decoded at the population level toward one that would be consistent with the world-relative direction (Fig. 6e).
- Local sources of inhibition in MT−/MSTv support the segregation of object motion and self-motion (Mechanism 2). To tease apart local and global (feedback-mediated) mechanisms, it would be valuable to test a condition wherein the global pattern does not directly stimulate the region adjacent to the RF center. Stimulating the “near surround” and “far surround” should differently modulate direction tuning.
- Surround suppression influences the reference frame within which object motion is signaled at the population-level (Figs. 6c,e, 7a): a transformation occurs in the model from the observer-relative frame of reference of retinal signals in earlier stages of the motion (e.g., MT−) to a reference frame more closely aligned with the stationary world, independent of the observer’s self-motion, in the later stages (e.g., MSTv). Characterizing neurons in the Global-Pattern and No-Global-Pattern conditions in areas further downstream from MT (e.g., MSTv) should reveal increased direction modulation toward the world-relative direction. Such a progression is not necessarily expected, given that the visual system could simply suppress the surrounding motion to enhance selectivity, while preserving the retinal reference frame of object motion signals.
- The visual system relies on a temporal solution to resolve the world-relative motion of moving objects and establish perceptual stability. In the Global-Pattern condition, neurons tuned to the retinal direction should peak at early poststimulus latencies and suppression should occur at later times, compared with the No-Global-Pattern condition. Neurons tuned to the world-relative direction should demonstrate enhancement at later poststimulus latencies.
- Neurons should only demonstrate directional modulation in the Global-Pattern condition if they are subject to surround suppression.

Areas involved in establishing perceptual stability

The global–local interactions and joint motion parallax/disparity tuning needed to disambiguate the retinal motion field during self-motion led us to consider the involvement of brain areas, MT and MST. In humans, this may correspond to the MT complex and area V6, respectively (Pitzalis et al., 2010). However, there is no reason to rule out the involvement of other areas at similar levels in the visual hierarchy. Our simulations demonstrate the importance of estimating the observer’s self-motion, which could implicate VIP, STPa, 7a, and other areas that exhibit the requisite full-field motion pattern sensitivity. These regions exhibit extensive interconnectivity (Maunsell and van Essen, 1983; Van Essen and Maunsell, 1983; Felleman and Van Essen, 1991) and the feedback-mediated suppression proposed in our model could influence object motion signals in areas other than MT, such as V4 (Li et al., 2013) and CIP (Rosenberg et al., 2013). We focused on MT and MSTd in particular because of the well established complementarity between local and global motion signals and studies have identified robust tuning to direction, speed, and disparity in both areas (Born and Bradley, 2005).

Comparison to existing models

The present model builds on the model of Layton and Fajen (2016b), which also included a divisive feedback mechanism from MSTd to MT. There are important advances worth highlighting. First, the present model makes specific predictions about the functional role and connectivity of different subpopulations of MT cells (MT+ and MT−), whereas the former model did not subdivide MT. Second, the former model focuses on direction alone, whereas here we consider how the visual system estimates and multiplexes depth, speed, and direction to disambiguate the optical motion signal. Third, despite the added overall complexity, the present model uses a simple Gaussian shape for feedback-mediated suppression (Fig. 6d, red curve), a more parsimonious account than its bimodal predecessor. This simplification is possible because the present model implements the simple rule: maximally suppress MT motion signals most similar in direction, speed, and disparity to the self-motion estimate in MSTd. This principle could extend more broadly to encompass additional sensorimotor signals (e.g., vestibular). Fourth, the present model demonstrates that feedforward on-center/off-surround interactions could account for local modulation of direction signals; the previous model used opponent lateral interactions. Finally, the former model implemented a simplistic notion of time, whereas the present model leverages extensive dynamic interactions between MT and MST.

The present model also advances the ViSTARS model of Browning et al. (2009b), which simulates the dynamics of visual areas that include MT and MST. Although ViSTARS estimates heading, it does not suppress the activity of neurons in “early” visual areas (e.g., MT) based on the activity pattern in MSTd. The present model does this to establish perceptual stability. Both models segregate the independent motion of objects, however, the mechanism in ViSTARS is local, whereas the present model also takes into account the global motion pattern induced by self-motion. Moreover, the present model recovers object motion relative to the stationary environment (world-relative), but this need not occur in ViSTARS.

A number of psychophysical studies have proposed a conceptual model whereby the visual system “subtracts” out the self-motion component from the object motion to disambiguate the retinal motion field (Warren and Rushton, 2007, 2009). Our analysis cautions against a literal interpretation of this model

when it comes to understanding underlying neural mechanisms. Although subtractive interactions may be capable of factoring out the influence of the observer's self-motion, other mechanisms that, for example, employ divisive interactions could offer a number of advantages, such as the potential to preserve direction information and signal bandwidth regardless of the object direction relative to the background (Fig. 2*d*).

Suppressing model MT motion signals consistent with the self-motion estimate could be viewed as canceling out a "prediction" or "expectation" produced by model MSTd, which connects our model with the principles of predictive coding (Friston, 2010). The theory posits that feedback carries predictions about how sensory signals should appear and feedforward signals transmit the error between the sensory array and prediction. From this perspective, disambiguating the retinal flow field (sum of observer's self-motion and object's independent movement) amounts to canceling out the observer's self-motion based on an estimate to recover the object's independent movement through the world (Fig. 1). Our model achieves this by matching feedforward MT optic flow signals with feedback signals derived from MSTd, carrying the predictions about the expected global motion pattern associated with the observer's self-motion. The model suppresses sensory signals that match the predicted motion parallax and disparity and signals that mismatch emerge as a prediction error, representing the world-relative object motion.

Model limitations

We either assumed a smooth depth gradient or constant depth in MSTd disparity and motion pattern tuning, but in reality, the spatial distribution of disparity is likely more diverse. In most cases here, this is not a problem because environments were simple 3D dot volumes. However, this could be a factor in the Variable Depth (Warren and Rushton, 2007) displays wherein the object potentially occupies a different depth than the rest of the scene. Recall that model MSTd cells send feedback to suppress MT signals that match at each spatial location the disparities corresponding to the observer's estimated scene-relative motion. Model MSTd cells would not strongly suppress the motion within the contours of a nearby object that moves in the condition wherein the dot volume is far away. Therefore, in cases where the world-relative observer motion arises at very different depths than the object, the present model may underestimate the directional modulation. It is noteworthy that this is a limitation of current implementation, rather than with the general computational principles or mechanisms.

Conclusion

We have developed computational principles and biologically plausible mechanisms that explain how the visual system could rely on depth through motion parallax and disparity to stabilize visual signals during self-motion. To that end, our account links surround suppression and joint motion parallax-disparity tuning properties to the coordinate system of object motion signals generated during self-motion. We foresee many exciting opportunities to leverage these computational principles in future studies to advance our understanding of the underlying neural mechanisms.

References

- Allman J, Miezin F, McGuinness E (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci* 8:407–430.
- Born RT (2000) Center-surround interactions in the middle temporal visual area of the owl monkey. *J Neurophysiol* 84:2658–2669.
- Born RT, Bradley DC (2005) Structure and function of visual area MT. *Annu Rev Neurosci* 28:157–189.
- Born RT, Tootell RB (1992) Segregation of global and local motion processing in primate middle temporal visual area. *Nature* 357:497.
- Bremmer F, Kubischik M, Pökel M, Lappe M, Hoffmann K (1999) Linear vestibular self-motion signals in monkey medial superior temporal area. *Ann N Y Acad Sci* 871:272–281.
- Britten KH, van Wezel RJA (2002) Area MST and heading perception in macaque monkeys. *Cereb Cortex* 12:692–701.
- Browning NA, Grossberg S, Mingolla E (2009a) A neural model of how the brain computes heading from optic flow in realistic scenes. *Cogn Psychol* 59:320–356.
- Browning NA, Grossberg S, Mingolla E (2009b) Cortical dynamics of navigation and steering in natural scenes: motion-based object segmentation, heading, and obstacle avoidance. *Neural Netw* 22:1383–1398.
- Cui Y, Liu LD, Khawaja FA, Pack CC, Butts DA (2013) Diverse suppressive influences in area MT and selectivity to complex motion features. *J Neurosci* 33:16715–16728.
- Cullen KE (2012) The vestibular system: multimodal integration and encoding of self-motion for motor control. *Trends Neurosci* 35:185–196.
- DeAngelis GC, Uka T (2003) Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J Neurophysiol* 89:1094–1111.
- Duffy CJ (1998) MST neurons respond to optic flow and translational movement. *J Neurophysiol* 80:1816–1827.
- Duffy CJ, Wurtz RH (1991) Sensitivity of MST neurons to optic flow stimuli: I. A continuum of response selectivity to large-field stimuli. *J Neurophysiol* 65:1329–1345.
- Duffy CJ, Wurtz RH (1995) Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *J Neurosci* 15:5192–5208.
- Duffy CJ, Wurtz RH (1997) Medial superior temporal area neurons respond to speed patterns in optic flow. *J Neurosci* 17:2839–2851.
- Dupin L, Wexler M (2013) Motion perception by a moving observer in a three-dimensional environment. *J Vis* 13(2):15 1–14.
- Fajen BR, Matthis JS (2013) Visual and non-visual contributions to the perception of object motion during self-motion. *PLoS One* 8:e55446.
- Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1:1–47.
- Friston K (2010) The free-energy principle: a unified brain theory. *Nat Rev Neurosci* 11:127–138.
- Gibson JJ (1950) The perception of the visual world. Houghton Mifflin.
- Grossberg S (1973) Contour enhancement, short term memory, and constancies in reverberating neural networks. *Stud Appl Math* 52:213–257.
- Grossberg S (1980) Intracellular mechanisms of adaptation and self-regulation in self-organizing networks: the role of chemical transducers. *Bull Math Biol* 42:365–396.
- Grossberg S (2013) Recurrent neural networks. *Scholarpedia* 8:1888.
- Gu Y, DeAngelis GC, Angelaki DE (2012) Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. *J Neurosci* 32:2299–2313.
- Inaba N, Shinomoto S, Yamane S, Takemura A, Kawano K (2007) MST neurons code for visual motion in space independent of pursuit eye movements. *J Neurophysiol* 97:3473–3483.
- Kim HR, Angelaki DE, DeAngelis GC (2016) The neural basis of depth perception from motion parallax. *Philos Trans R Soc Lond B Biol Sci* 371:20150256.
- Lagae L, Maes H, Raiguel S, Xiao DK, Orban GA (1994) Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. *J Neurophysiol* 71:1597–1626.
- Layton OW, Browning NA (2014) A unified model of heading and path perception in primate MSTd. *PLoS Comput Biol* 10:e1003476.
- Layton OW, Fajen BR (2016a) A neural model of MST and MT explains perceived object motion during self-motion. *J Neurosci* 36:8093–8102.
- Layton OW, Fajen BR (2016b) Competitive dynamics in MSTd: A mechanism for robust heading perception based on optic flow. *PLoS Comput Biol* 12:e1004942.
- Layton OW, Fajen BR (2016c) The temporal dynamics of heading perception in the presence of moving objects. *J Neurophysiol* 115:286–300.
- Layton OW, Fajen BR (2017) Possible role for recurrent interactions between expansion and contraction cells in MSTd during self-motion perception in dynamic environments. *J Vis* 17(5):5 1–21.

- Layton OW, Mingolla E, Browning NA (2012) A motion pooling model of visually guided navigation explains human behavior in the presence of independently moving objects. *J Vis* 12(1):20 1–19.
- Li P, Zhu S, Chen M, Han C, Xu H, Hu J, Fang Y, Lu HD (2013) A motion direction preference map in monkey V4. *Neuron* 78:376–388.
- Liu S, Angelaki DE (2009) Vestibular signals in macaque extrastriate visual cortex are functionally appropriate for heading perception. *J Neurosci* 29:8936–8945.
- Matsumiya K, Ando H (2009) World-centered perception of 3D object motion during visually guided self-motion. *J Vis* 9(1):15 1–13.
- Maunsell JH, van Essen DC (1983) The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J Neurosci* 3:2563–2586.
- Mejias JF, Payeur A, Selin E, Maler L, Longtin A (2014) Subtractive, divisive and non-monotonic gain control in feedforward nets linearized by noise and delays. *Front Comput Neurosci* 8:19.
- Mineault PJ, Khawaja FA, Butts DA, Pack CC (2012) Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proc Natl Acad Sci U S A* 109:E972–E980.
- Nadler JW, Angelaki DE, DeAngelis GC (2008) A neural representation of depth from motion parallax in macaque visual cortex. *Nature* 452:642–645.
- Nadler JW, Barbash D, Kim HR, Shimp S, Angelaki DE, DeAngelis GC (2013) Joint representation of depth from motion parallax and binocular disparity cues in macaque area MT. *J Neurosci* 33:14061–14074.
- Naka KI, Rushton WA (1966) S-potentials from colour units in the retina of fish (Cyprinidae). *J Physiol* 185:536–555.
- Peltier N, Angelaki DE, DeAngelis GC (2016) Optic flow parsing in macaque monkeys. *Soc Neurosci Abstr* 42:328.19.
- Perrone JA (2012) A neural-based code for computing image velocity from small sets of middle temporal (MT/V5) neuron inputs. *J Vis* 12(8):1 1–31.
- Pitzalis S, Sereno MI, Committeri G, Fattori P, Galati G, Patria F, Galletti C (2010) Human v6: the medial motion area. *Cereb Cortex* 20:411–424.
- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
- Rosenberg A, Cowan NJ, Angelaki DE (2013) The visual representation of 3D object orientation in parietal cortex. *J Neurosci* 33:19352–19361.
- Royden CS, Connors EM (2010) The detection of moving objects by moving observers. *Vision Res* 50:1014–1024.
- Rushton SK, Warren PA (2005) Moving observers, relative retinal motion and the detection of object movement. *Curr Biol* 15:R542–R543.
- Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9:1421–1431.
- Smolyanskaya A, Ruff DA, Born RT (2013) Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable. *J Neurophysiol* 110:2806–2816.
- Tanaka K, Hikosaka K, Saito H, Yukie M, Fukada Y, Iwai E (1986) Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *J Neurosci* 6:134–144.
- Tanaka K, Sugita Y, Moriya M, Saito H (1993) Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *J Neurophysiol* 69:128–142.
- Thier P, Erickson RG (1992) Vestibular input to visual-tracking neurons in area MST of awake rhesus monkeys. *Ann N Y Acad Sci* 656:960–963.
- Van Essen DC, Maunsell JH (1983) Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci* 6:370–375.
- Warren PA, Rushton SK (2007) Perception of object trajectory: parsing retinal motion into self and object movement components. *J Vis* 7(11):2 1–11.
- Warren PA, Rushton SK (2008) Evidence for flow-parsing in radial flow displays. *Vision Res* 48:655–663.
- Warren PA, Rushton SK (2009) Optic flow processing for the assessment of object movement during ego movement. *Curr Biol* 19:1555–1560.
- Warren WH, Hannon DJ (1988) Direction of self-motion is perceived from optical flow. *Nature* 336:162–163.
- Xiao DK, Marcar VL, Raiguel SE, Orban GA (1997a) Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. *Eur J Neurosci* 9:956–964.
- Xiao DK, Raiguel S, Marcar V, Orban GA (1997b) The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cereb Cortex* 7: 662–677.
- Xiao DK, Raiguel S, Marcar V, Orban GA (1998) Influence of stimulus speed upon the antagonistic surrounds of area MT/V5 neurons. *Neuroreport* 9:1321–1326.
- Yang Y, Liu S, Chowdhury SA, DeAngelis GC, Angelaki DE (2011) Binocular disparity tuning and visual-vestibular congruency of multisensory neurons in macaque parietal cortex. *J Neurosci* 31:17905–17916.
- Yu X, Hou H, Spillmann L, Gu Y (2018) Causal evidence of motion signals in macaque middle temporal area weighted-pooled for global heading perception. *Cereb Cortex* 28:612–624.