Multiscale sampling model for motion integration

Lena Sherbakov

Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA, USA ✉

Arash Yazdanbakhsh

Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA, USA
Program in Cognitive and Neural Systems, Boston University, Boston, MA, USA 🏠✉

Biologically plausible strategies for visual scene integration across spatial and temporal domains continues to be a challenging topic. The fundamental question we address is whether classical problems in motion integration, such as the aperture problem, can be solved in a model that samples the visual scene at multiple spatial and temporal scales in parallel. We hypothesize that fast interareal connections that allow feedback of information between cortical layers are the key processes that disambiguate motion direction. We developed a neural model showing how the aperture problem can be solved using different spatial sampling scales between LGN, V1 layer 4, V1 layer 6, and area MT. Our results suggest that multiscale sampling, rather than feedback explicitly, is the key process that gives rise to end-stopped cells in V1 and enables area MT to solve the aperture problem without the need for calculating intersecting constraints or crafting intricate patterns of spatiotemporal receptive fields. Furthermore, the model explains why end-stopped cells no longer emerge in the absence of V1 layer 6 activity (Bolz & Gilbert, 1986), why V1 layer 4 cells are significantly more end-stopped than V1 layer 6 cells (Pack, Livingstone, Duffy, & Born, 2003), and how it is possible to have a solution to the aperture problem in area MT with no solution in V1 in the presence of driving feedback. In summary, while much research in the field focuses on how a laminar architecture can give rise to complicated spatiotemporal receptive fields to solve problems in the motion domain, we show that one can reframe motion integration as an emergent property of multiscale sampling achieved concurrently within lamina and across multiple visual areas.

Introduction

Visual scene integration is a well-studied topic, yet there is still little consensus about the necessary and sufficient network that affords the function observed. Historically, the classical view of visual processing is a local to global approach whereby earlier visual areas serve as edge and orientation detectors that pass on information to higher-order areas that perform more complex processing to complete the 3-D representation of the visual scene (Marr, 1982). However, recent research has shown that V1 contains highly multiplexed information about brightness, orientation, spatial frequency, and other stimulus properties (Ts'o & Gilbert, 1988; Rossi & Paradiso, 1999; Friedman, Zhou, & von der Heydt, 2003). Countering the view that early visual areas only process local information, a cell's response to border ownership was shown to be largely independent of spatial extent and is represented at a single neuron level (Craft, Schuetze, Niebur, & von der Heydt, 2007). Zhou, Friedman, and von der Heydt (2000) showed that as early as V1 18% of the cells responded to border ownership. Moreover, different sized receptive fields in different visual areas suggest that some stimulus properties may be sampled in higher-order areas in parallel with processing at lower areas through fast interareal connections (Bullier, 2001; Girard, Hupeì, & Bullier, 2001). Together, these new pieces of evidence suggest that much of the processing that was previously suggested to occur intra-areally within the same layer of the visual cortex may instead be computed by fast, parallel, bidirectional, interareal and interlaminar connections at different spatial resolutions.

In this paper, we explore whether a classic problem in visual motion integration—the aperture problem—can be solved with a simple model that samples the visual scene at different spatial and temporal scales in parallel. To frame what is meant by aperture problem, we note that a neuron's receptive field acts as a viewing aperture and only detects components of motion visible to its field of view (often not the same as the true global

motion). Due to the difference in size between stimulus and receptive fields (the latter being smaller), the true motion of a line viewed from this aperture is only unambiguous at line endings (assuming no significant texture is present); the rest of the cells only have view access to the perpendicular component of motion—this is commonly understood as the aperture problem (Stumpf, 1911; Wallach, 1935; Horn & Schunck, 1981). The neural trace of the aperture problem is therefore considered to be cells that only respond to the *component* direction of motion while the stimulus itself may move in a different global direction that cannot be perceived from the local aperture (known as the *pattern* motion). Surprisingly, 200 ms after the onset of the moving stimulus, area MT already "solves" the aperture problem (responds to the pattern of motion) while V1 still largely responds to components of motion (Pack & Born, 2001; Pack et al., 2003).

Historically, three broad classes of solutions have been proposed to explain how the aperture problem is solved: (a) intersection of constraints, (b) vector averaging of motion direction, and (c) feature tracking. The intersection of constraints method uses the normal components of velocity and predicts the perceived direction of motion from where those velocity-space lines intersect (Adelson & Movshon, 1982). In the vector averaging approach, the ambiguous line segments are perceived to move in a direction consistent with the average of the orthogonal components of the lines (Yo & Wilson, 1992). The unambiguous line ends are summed with varying weights together with ambiguous line segments to simulate the perception of global motion. In these models, cells are typically divided into two classes (terminator units, which can see the line ends, and contour units, which cannot). The aperture problem is solved by setting the weight applied to the terminator units to be larger than that applied to the contour units (Lorenceau, Shiffrar, Wells, & Castet, 1992). Lastly, the feature tracking approach propagates unambiguous signals (line ends or intersections of line ends) inward to fill in missing information of ambiguous retinal locations; this is done via a recurrent or feedback neural network (Chey, Grossberg, & Mingolla, 1997). The intersection of constraints and feature tracking methods yield the same outcome (i.e., equivalent at the computation level) and are different only at the algorithm and implementation levels. The above models approach motion integration in two stages: (a) compute local motion in early visual areas and (b) integrate local signals in higher visual areas to obtain the global percept of motion. Analogously, in feature tracking models, short-range filters and competitive interactions are assumed to take place in V1, and longer-range filters in MT establish the larger-scale representation of the moving object (Chey et al., 1997).

Our approach differs from the above three in several ways: (a) We de-emphasize intra-areal processing as the central mechanism that propagates the relevant information to solve the aperture problem, (b) fast interareal and interlaminar connections between V1 and MT feed back information onto V1, and (c) the computation done in our model areas V1 and MT is essentially identical with the only difference being spatial sampling scales. Henceforth, we use the terms "spatial sampling scale" and "multiscale sampling" to mean the integration of information from neural populations with heterogeneous receptive field sizes wherein some populations have receptive fields as much as an order of magnitude larger than other populations. This type of heterogeneity is well documented in biology (Bolz & Gilbert, 1986; Albright & Desimone, 1987), but its usefulness is underexplored in modeling work.

More recently, other models have suggested that multiscale sampling and feedback are the critical components to quickly and successfully solve the aperture problem in area MT (Bayerl & Neumann, 2004). However, it remains to be explained why V1 activity does not look more like MT activity (why it persists to show a component-like direction of motion). In the present work, we propose that this inconsistency with physiology can be addressed with a neural model that distinguishes what feedback layer 6 versus layer 4 of V1 receive from area MT. Moreover, while Bayerl and Neumann propose a modulatory, top-down connection from MT to V1, our model suggests that the interareal connections must, in fact, be driving with the only gating or modulatory connections coming intra-areally from V1 layer 6 to V1 layer 4.

There has also been an increased interest in statistical models that explain how and under what conditions the aperture problem is solved. Most of these models rely on a Bayesian framework in which the local motion is represented by likelihood functions of the line's position and velocity. Global motion is then inferred by introducing prior constraints and computing the posterior distribution (Perrinet & Masson, 2012). The prior constraints are a priori knowledge of cell or motion properties, such as preference for slow line speed (Montagnini, Mamassian, Perrinet, Castet, & Masson, 2007), smoothness of motion away from luminance discontinuities (Tlapale, Masson, & Kornprobst, 2010), or knowledge of line end versus line middle (Barthélemy, Perrinet, Castet, & Masson, 2008). Our approach differs in that while we do not doubt that the brain exploits various prediction strategies, we remain uncertain that these prior-like constraints are explicitly available to the neural system. Rather, like the solution to the aperture problem itself, the constraints of temporal continuity and slow line speed may arise as an emergent property of a multiscale sampling process. Moreover, most statistical models involve computing integrals over hidden vari-
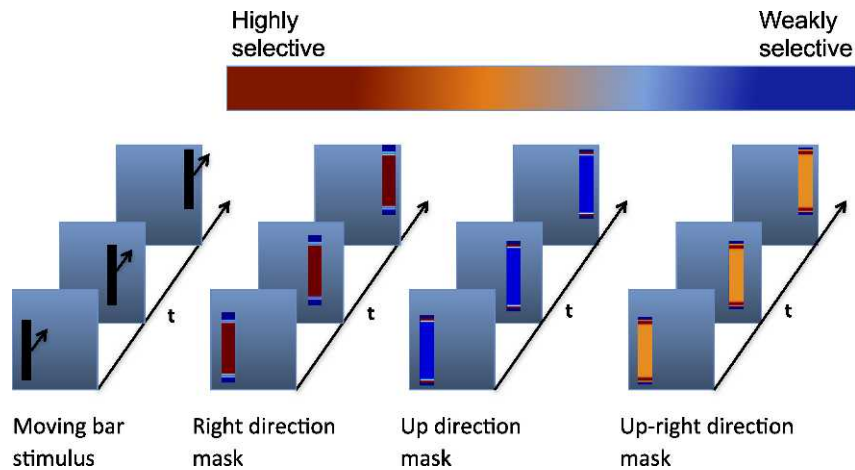
Figure 1. Selectivity mask representation. In all of the simulations, the bar moves in the up-right direction (leftmost figure). To simulate direction-selective cells, we introduce a mask that multiplies LGN's activity depending on the location of the receptive field. The rightward direction-selective mask is strongest in the center of the bar where only the horizontal direction of motion is registered by a small receptive field. At the bar ends where the true direction of motion is registered by cells with small receptive fields, the up-right direction selective mask is most active.

ables, which is not only time consuming on a single computer but also biologically questionable.

A last distinguishing feature of our model is the emergence of several observable cell properties that we did not explicitly set out to simulate. End-stopping, a phenomenon observed in area V1 and MT whereby cells develop suppressed responses to long but not short bar lengths (Pack & Born, 2001; Pack et al., 2003), emerges from our model area V1. We show that it is possible to solve problems in motion integration with this simple multiscale sampling approach in which fast interareal and interlaminar connections complement the relatively slow intralaminar communication. Moreover, the model is consistent with cell physiology and receptive field sizes.

Methods

In this work, we develop a computational model that simulates the response of three visual areas (LGN, V1 layers 4 and 6, and MT) to a vertically oriented bar moving at a 45° angle (Figure 1). For simplicity, the model only includes cell populations selective to three directions of motion (right, up, and up-right). We include only the detailed laminar structure we found necessary for the solution of the aperture problem to emerge. The model architecture is detailed in Figure 2.

The model

The model consists of LGN cells, V1 layer 6 neurons, V1 layer 4 interneurons, V1 layer 4 excitatory neurons,

and MT cells (Figure 2). Nondirection-selective LGN cells sample the moving bar with receptive fields whose excitatory regions are 1/25th the size of the bar. There is no within-LGN (intra-areal) connectivity; the LGN layer receives only feed-forward input from the moving bar.

To simulate direction-selective V1 neurons, we introduce the concept of a direction-selective mask that is applied to neurons of a given selectivity after they receive the LGN input. Model areas V1 layer 6, V1 layer 4 interneurons, and V1 layer 4 excitatory cells each have three motion direction-selective layers: rightward, upward, and right-up (45°). The rightward direction cells, for example, respond best to LGN input at the center of the moving bar where the only component of motion that is visible to the cell's receptive field is horizontal (for more detail, see Direction mask section).

Model LGN synapses onto three V1 populations: V1 layer 6 cells, V1 layer 4 interneurons, and V1 layer 4 excitatory cells. These synapses are not only well documented in physiology studies of area V1 (Van Essen, Anderson, & Felleman, 1992; Lamme, Super, & Spekreijse, 1998), but also serve as the backbone for many computational models of V1 (Grossberg & Williamson, 2001; Raizada & Grossberg, 2001). It should be noted that we were not seeking to complicate the model unnecessarily by adding laminar connections; rather, we derived this structure as the necessary and sufficient network to explain various aspects of the aperture problem. All V1 populations inherit the direction selectivity of the corresponding mask (therefore yielding nine V1 populations: V1 L6 rightward selective, V1 L4 interneuron rightward selective, V1 L4 excitatory rightward selective, and similarly for the
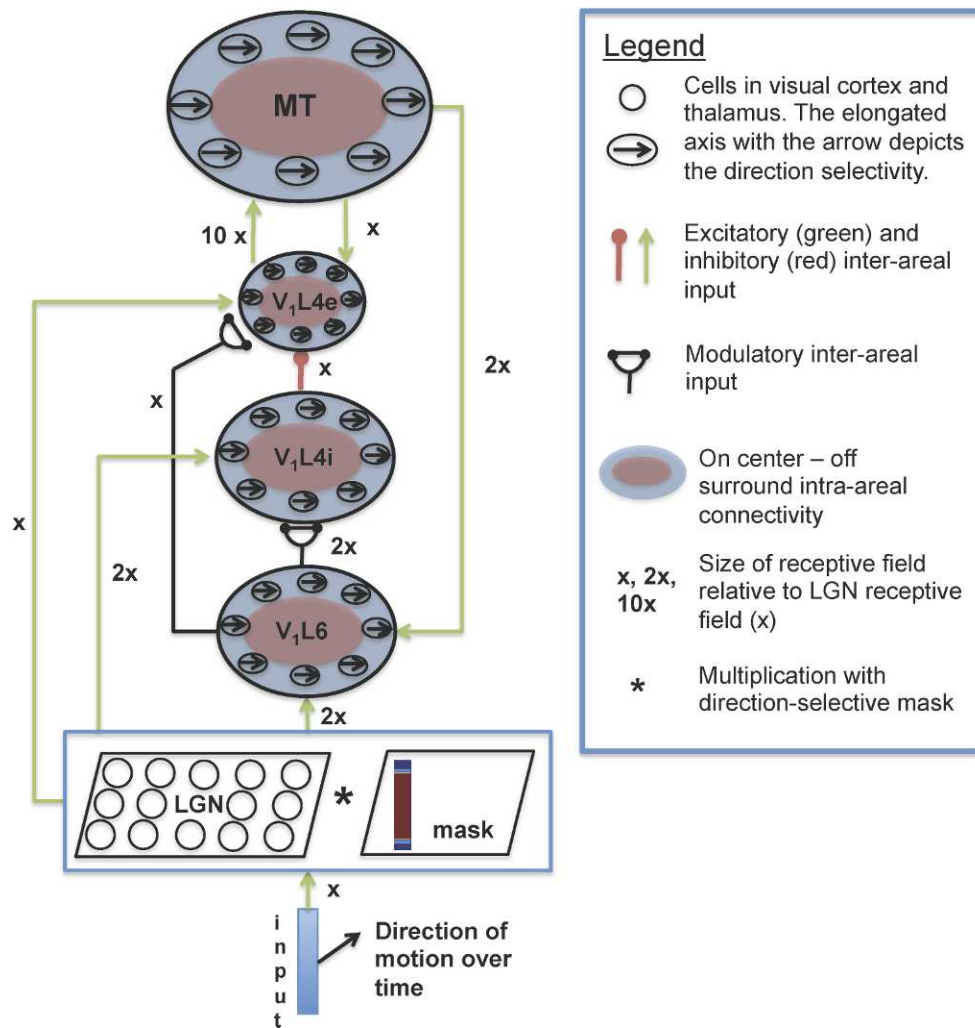
Figure 2. Model diagram. V1 layer 4 cells (both excitatory and inhibitory) and V1 layer 6 cells receive bottom-up input from LGN with different-sized sampling Gaussians as indicated by the size of the ovals and the x, 2x notation. This bottom-up activity is first passed through a direction-selective mask, which simulates the motion direction–selective cells of V1. MT receives input from V1 L4 and sends feedback to both V1 L6 and V1 L4, sampled with different-sized kernels. V1 L6 influences V1 L4 activity through inhibitory interneurons as well as through direct modulatory input. Green arrows indicate interareal excitatory connections, and red circles indicate interareal inhibition. Modulatory connections are in black. All feed-forward and feedback connections are driving (additive) and shunted by the cell's own activity with the exception of V1 layer 6, whose influence is always modulatory (multiplicative). A red oval with a blue oval surround symbolizes on-center-off-surround intra-areal connectivity. All receptive fields are Gaussian. While we do not show the diagrams for upward and right-upward selective cells, they are identical to this figure with the exception of the direction-selective mask applied at the beginning. No cross-orientation competition exists. V1L6 = V1 layer 6 cells that are rightward motion-direction selective, V1L4i = V1 layer 4 inhibitory interneurons that are rightward motion-direction selective, V1L4e = V1 layer 4 excitatory cells that are rightward motion-direction selective, and MT = area MT cells that are rightward motion-direction selective.

upward-direction selective and up-right direction-selective cells). While both layers 6 and 4 of V1 receive LGN input, the receptive field sizes are distinct. Our model V1 layer 6 has twice the receptive field size of V1 layer 4, which has similar receptive field sizes to LGN (see Figure 2). These kernel sizes were chosen to be consistent with known physiology (Bolz & Gilbert, 1986).

Model V1 layer 6 cells have modulatory inputs to V1 layer 4 interneurons as well as V1 layer 4 excitatory cells. The idea that V1 layer 6 serves as a "gate" through which bottom-up and top-down activity is regulated has been proposed previously (Bolz & Gilbert, 1986). We find these modulatory connections necessary in explaining why end-stopped cells no longer emerge in V1 layer 4 in the absence of V1 layer 6 activity (Bolz & Gilbert, 1986). V1 layer 4 interneurons have inhibitory, driving synapses onto V1 layer 4 excitatory cells (Figure 2).

Model area MT is similarly split into three populations that inherit their motion-direction selectivity from V1: rightward-selective MT cells, upward-selective MT cells, and right-up selective MT cells. MT only receives input from V1 layer 4 cells of the same direction selectivity; no cross-orientation interactions are modeled in either area V1 or MT. MT receptive field sizes are simulated as roughly 10 times that of V1 layer 4 receptive field sizes (Albright & Desimone, 1987). There exists evidence for synapses directly from V1 layer 6 onto MT (Maunsell & Van Essen, 1983); however, we did not find this connection to be fundamental to the model and therefore did not include it.

The feedback connections in our model consist of MT onto V1 layer 6 and MT onto V1 layer 4 excitatory cells (Sillito, Cudeiro, & Jones, 2006). The receptive field with which layers 4 and 6 sample MT cells are the same as the bottom-up receptive fields of these neural populations.

All excitatory and inhibitory inputs to the model are driving (additive) and shunted (modulated by the cell's own activity) with the exception of V1 layer 6 synapses, which are modulatory (see Appendix).

All visual areas (with the exception of LGN) are modeled with distance-dependent shunting with on-center-off-surround intra-areal connections:

$$\frac{dx_{ij}}{dt} = -Ax_{ij} + (B - x_{ij})\left(I(t)*C\right)_{ij} \\ - (D + x_{ij})\left(I(t)*E\right)_{ij} + (x*F)ij$$

where x_{ij} is the model cell at location (i, j), A is the membrane potential decay rate, B stands for the depolarization threshold, $I(t)$ is the driving input to the cell at time t, C is a kernel for distance-dependent excitation, D is a surrogate for the hyperpolarization threshold, E is a kernel for distance-dependent inhibition, and F is a kernel for on-center-off-surround intra-areal interactions. The * operation denotes a convolution with the respective kernel. The parameters $B = 90$ and $D = 60$ are kept constant for all simulated brain regions. The decay rate, A, and the kernel sizes C, E, and F are varied as described in the section *Parameter selection*.

LGN is similar to other model areas with the simplification that it does not have any intra-areal interactions. For a detailed summary of the equations, see Appendix.

Direction mask

To address how our model neurons detect direction of motion, we introduce the direction-selective mask abstraction. The direction mask functions as a rudimentary Reichardt detector or any other mechanism that extracts "first-order" motion. We do not address how this direction mask emerges in a biological system; rather, the goal of this paper is to focus on multiscale sampling of the motion stimulus.

Motion direction selectivity is achieved in area V1 by introducing a direction mask over LGN cells that modulate the sampled activity based on which spatial region the V1 cells can perceive (Figure 1). For example, at the center of the bar where the V1 cells only see rightward direction of motion, the rightward mask has the strongest activity compared to the upward and right-up masks. Conversely, at the bar ends where the cells have access to the true direction of motion, the right-up mask is significantly more active than either the rightward or upward direction masks. The direction-selective masks move together with the bar over time to simulate which direction-selective V1 cells are receiving input from LGN and with what strength. The size of the direction-selective masks is no larger than the LGN receptive field size, which does not come close to "seeing" the full moving bar. Therefore, the direction mask concept alone cannot solve the aperture problem.

The stimulus

The stimulus we use is a vertically oriented bar 100 units in length and 1 unit in width, moving at a 45° angle relative to the horizontal (Figure 1). Given that the excitatory portion of an LGN receptive field covers roughly 0.2° of visual angle (Zhou et al., 2000), the moving bar roughly covers 4° of visual angle in length and 0.2° of visual angle in width. The bar moves at every time step to the upper right corner, and the time step refers to the Δt taken by the coupled ODE solver. The total simulation time is 30 time steps, simulating 300 ms.

Analysis of simulations

To determine whether the aperture problem was present in our simulation, we defined the solution to the aperture problem to be the case when, at some time t, the vector average of the preferred direction of motion pointed toward the pattern motion (45° from the horizontal) as opposed to the component direction of motion. The expected vector average component direction of motion was 2° from the horizontal for area V1 layer 4, 4° for V1 layer 6, and 18° for area MT. The expected component direction of motion is not uniquely 0° from the horizontal because cells that could see the bar ends and therefore the correct direction of motion (45°) are averaged with cells that can only see the middle of the bar (0° from the horizontal). We assessed whether the solution to the aperture problem

was achieved in V1 layer 6, V1 layer 4, and area MT separately. The preferred direction was assessed at different time points throughout the simulation.

Additionally, we investigated cell dynamics for model areas V1 L6, V1 L4, and MT by breaking down the analysis by cells whose receptive fields could see the bar ends versus those that could not. The presence of end-stopped cells was defined as suppressed activity after 20 ms of simulation time for long bars (cells that could not see the bar ends) without any changes in the activity for short bars (cells whose receptive fields could see the bar ends).

Parameter selection

To find the appropriate parameter range for our model, we attempted to match our LGN, V1, and MT cells to known latencies, peak response profiles, and spike distributions from available data in the macaque visual system. For LGN dynamics, our target cell was tuned to have a latency of roughly 20 ms (Schmolesky et al., 1998), a peak response at 50 ms, and complete response decay by 300 ms (Maunsell et al., 1999). The V1 cells were targeted to have a latency of 50 ms (Schmolesky et al., 1998), peak response at 80 ms, and response decay by 150 ms (Xing, Yeh, Burns, & Shapley, 2012). Model MT neurons were targeted to have a 70 ms latency (Schmolesky et al., 1998), a peak response at 100 ms, and a vanishing response by 200 ms (Raiguel, Xiao, Marcar, & Orban, 1999). Extended sustained responses over 1 s known to exist to a lesser or greater extent in each cell population were not considered. Due to feedback in the model, these target dynamics were not strictly enforced but rather served as guides and sanity checks for the model. The exact decay rates and other model parameters can be found in the Appendix.

To enforce the notion of different sized receptive fields in LGN, V1 layer 6, V1 layer 4, and MT, we used two-dimensional Gaussians to simulate the amount of excitatory and inhibitory influence of neighboring cells both within (intra-) and between (inter-) lamina and visual areas. We up-sampled or down-sampled the excitatory and inhibitory Gaussians by the same amount, which was determined by the relative receptive field size of the given visual area to the LGN receptive field size.

All excitatory Gaussian kernels had a standard deviation = 0.15 and peak = 18, representing the spatial spread and amplitude of the outgoing signals passed from one visual area to another. The inhibitory Gaussians contributing to the off-surround had a standard deviation = 1.2 and peak = 0.5. These parameters were chosen for consistency with other models that use the shunting equation to represent the membrane potential of cell populations (Grossberg & Todorovi, 1988). We note that our choice for using shunting feedback for cell dynamics was driven by its inherent gain control property and ability to solve the noise-saturation dilemma (Grossberg, 1973). It remains to be proven that the model described in this paper can work and stabilize without shunting dynamics in the cell's membrane equation.

The LGN receptive field was used as the baseline receptive field, which was then up-sampled to simulate the receptive fields of V1 and MT. The excitatory portion of the LGN Gaussian had a radius of 2 units (cells), and the inhibitory portion had a radius of 5 units. V1 layer 6 was modeled as having twice the receptive field of LGN (excitatory radius = 4 units, inhibitory radius = 10 units). Our model V1 layer 4 had the same receptive field size as LGN, consistent with data that suggests layer 4 has smaller receptive fields than layer 6 of V1 (Bolz & Gilbert, 1986). We modeled area MT as having a receptive field that is 10 times that of LGN and V1 layer 4 (excitatory radius = 20 units, inhibitory radius = 50 units)—a modeling decision that is also rooted in physiology (Albright & Desimone, 1987). All feedback projections from MT to different lamina of V1 are sampled with the same size Gaussian as the feed-forward projections for that visual area.

The intra-areal sampling was simulated by a difference of Gaussians (excitatory-inhibitory), whose excitatory and inhibitory regions were down-sampled by two, relative to the cell's interareal sampling kernel (for example, MT's intra-areal sampling kernel had an excitatory radius of 10 units and an inhibitory radius of 25 units). This relatively smaller receptive field was meant to simulate slower intra-areal communication when compared to its interareal counterpart.

All simulations were performed in MATLAB 2009b. All equations and stimuli were modeled in 2-D in their differential equation form (see Appendix).

Results

Our simulation results show that the aperture problem can be solved in area MT with this relatively simple multiscale sampling model (Figure 4). The initial response of MT to the moving bar is largely in the component direction of motion (vector average preferred direction = 23.6° while the expected preferred direction, if the cell were listening to the components of motion, is 18°). However, after 60 ms, MT switches to responding entirely to the pattern motion (vector average preferred direction = 42.6° relative to the expected 45° if the cell were listening to the pattern motion).
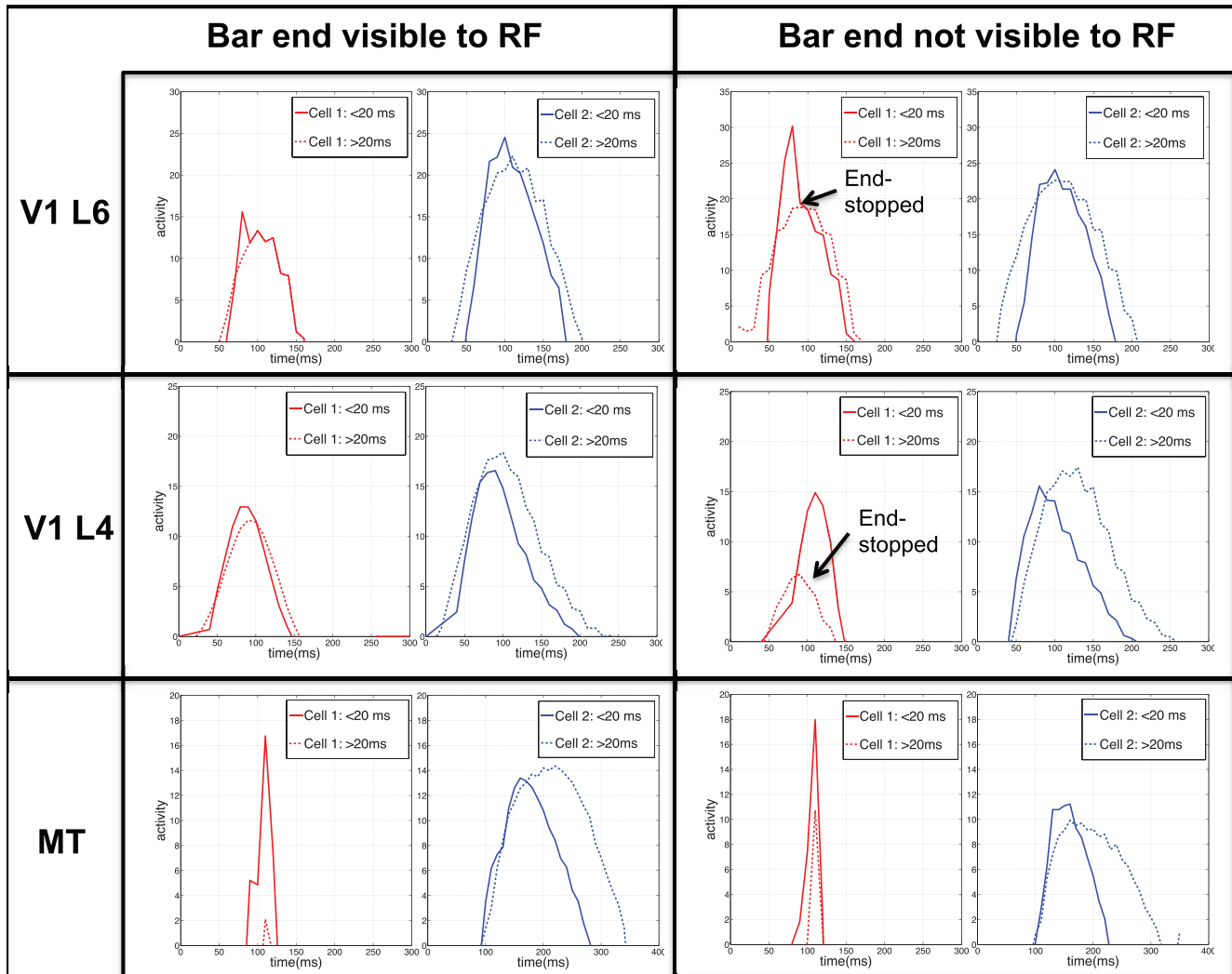
Figure 3. Model dynamics. Cell responses of two representative cells (in red and blue, respectively) in model areas V1 L6, V1 L4, and MT. The solid lines represent response of the cells early in the simulation (before 20 ms), and the dotted lines represent the response later in the simulation (after 20 ms). The first column shows the dynamics of the cells whose receptive field falls within the bar ends; the second column shows the dynamics of cells whose receptive fields only have access to the middle of the bar. The figure highlights the development of end-stopped cells largely in area V1 L4 and, to a lesser extent, in V1 L6. Unlike V1, certain cells in MT show suppression of response to both short bars (where the line end is visible) and longer bars (where the line end is not visible), implicating that an entire subset of direction-selective cells (in this case the rightward-direction cells) are being suppressed. The activity units are the threshold-scaled membrane potentials of the cells (see Appendix).

V1 layer 6 responds mostly to component motion throughout the simulation (vector average = 21° early in the simulation and 25° later in the simulation). V1 layer 4, however, begins to shift more strongly toward pattern motion as the simulation progresses (vector average = 22° early in the simulation and 33° later in the simulation). This phenomenon of V1 neurons being caught between component and pattern motion has been documented in end-stopped cells (most of which are coincidentally found in layer 4 of V1) (Pack et al., 2003).

When we analyzed the dynamics of our model cells, we discovered that a strong end-stopping phenomenon emerged in our V1 layer 4 cells (and to a lesser extent in our V1 layer 6 cells) (Figure 3). After 20 ms, most of our V1 layer 4 cells had a significantly suppressed response to the middle of the bar relative to their initial response (peak response dropped by 65%) while the end of the bar response remained unchanged. Pack et al. (2003) also found that cells were most strongly end-stopped in layer 4 compared to layer 6. Interestingly, this phenomenon was not modeled explicitly but rather falls out of the multiscale sampling approach for reasons we describe in the Discussion section. Our model MT cells also had a suppressed response in the middle of the bar after 20 ms; however, these cells had a
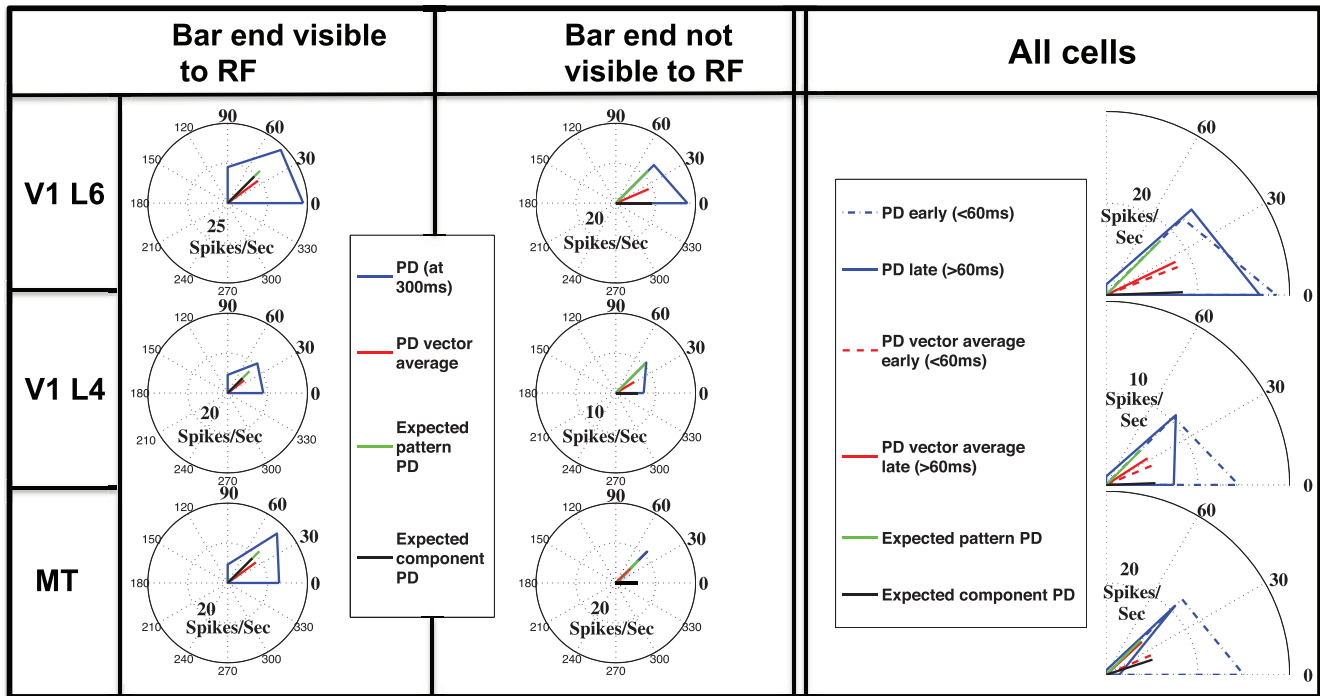
Figure 4. Preferred direction (PD) of cells whose receptive fields see the bar ends (leftmost column) and those that only see the middle of the bar (middle column) for areas V1 L6 (first row), V1 L4 (middle row), and MT (third row). The short red line indicates the vector average of the PD. The short black line indicates PD if the cells were only responding to the component direction of motion, and the green line corresponds to the expected PD if the cell was responding to the pattern direction of motion. To get a global view of direction coding in our model visual areas, the last column shows the average PD for the cells that see the line end and those that don't, together, in areas V1 L6 (first row), V1 L4 (second row), and MT (third row). The dotted blue lines indicate the PD early in the simulation (<60 ms), and the solid blue lines show the PD of the cells after 60 ms. Simulation area V1 L6 responds most to the component direction of motion and changes the least throughout the simulation. Area V1 L4 first responds to the component direction of motion but shifts closer toward the pattern direction of motion later in the simulation, such that the vector average of the PD is between the two extremes. Area MT responds to the component direction of motion at the beginning; however, after 60 ms, MT responds entirely to the pattern. While the expected pattern motion is the same for all cells (45°), the component motion is different based on the size of the receptive field of the model area. The expected component direction of motion is not uniquely 0° from the horizontal because cells that can see the bar ends and therefore whose component motion is the correct direction of motion (45°) are averaged with cells that can only see the middle of the bar (0° from the horizontal). The expected PD for component motion is 2° from the horizontal for V1 L4, 4° from the horizontal for V1 L6, and 18° from the horizontal for MT.

suppressed response at the bar ends and therefore cannot be called end-stopped. For a detailed breakdown of cell dynamics and preferred directions at the bar ends and middle of the bar, see Figures 3 and 4, respectively.

In an attempt to understand why end-stopped cells were emerging from our model, we cut all feedback connections from area MT to V1 (Figure 5A). We found that end-stopping can still develop in the absence of feedback only if layer 6 V1 cells have a larger receptive field than layer 4 V1 neurons (see the Discussion section for further explanation). Moreover, the aperture problem could still be solved by MT (albeit more slowly) without the feedback as long as end-stopped cells emerge from the dynamics. However, with only a single spatial sampling scale, we were never able

to produce end-stopped cells or the solution to the aperture problem in MT.

Lastly, when we deactivated V1 layer 6 in our model (Figure 5B), we no longer saw end-stopped cells developing in layer 4. This model phenomenon is consistent with physiology (Bolz & Gilbert, 1986) and reinforces the decision of modulatory or gating connections from V1 layer 6 to V1 layer 4 interneurons and excitatory cells.

Discussion

Our simulations show that it is indeed possible to solve the aperture problem through multiscale sampling between different lamina and visual areas. Our
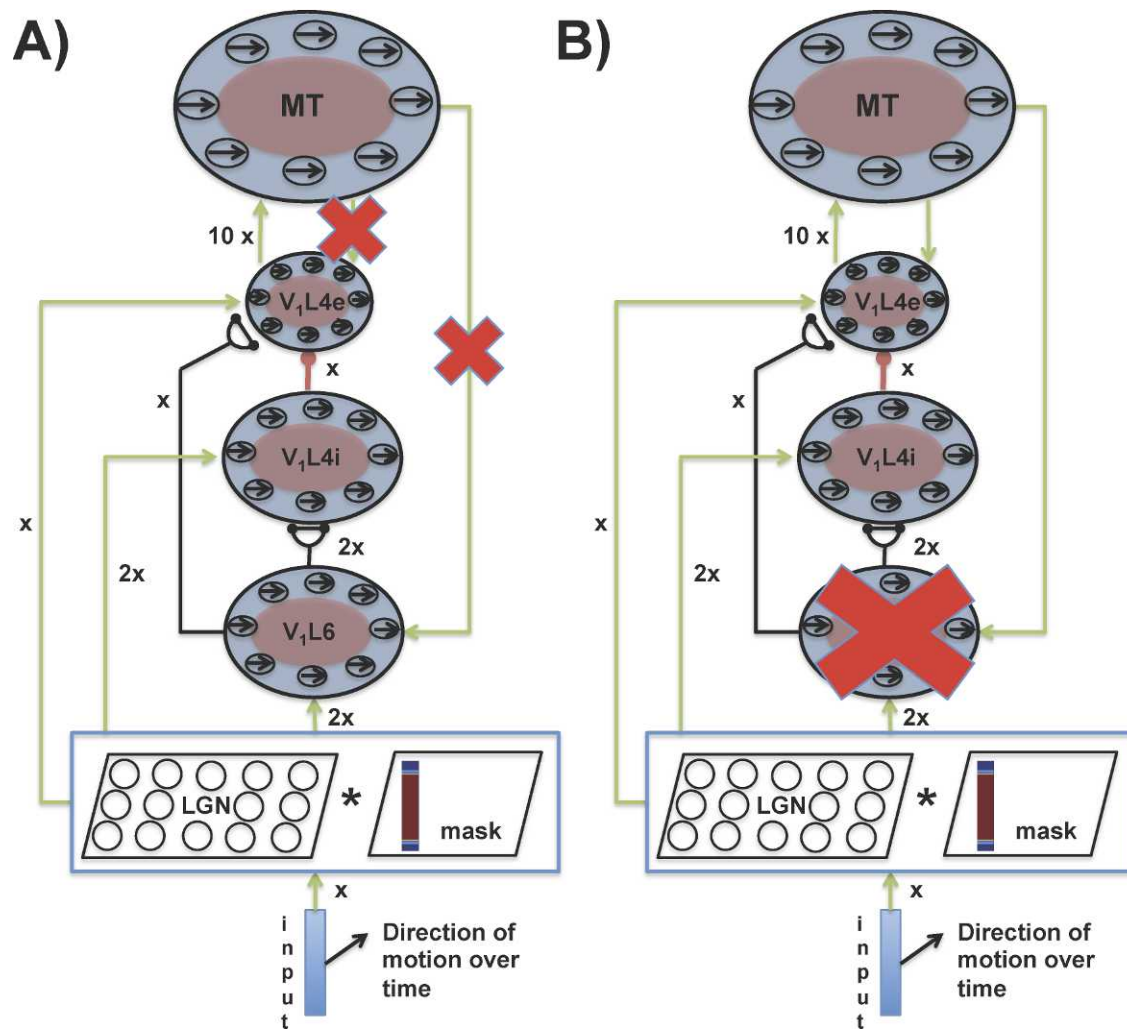
Figure 5. A) All feedback is disabled from the model. End-stopping in V1 and the solution to the aperture problem by MT still emerges due to the different receptive field sizes of layer 6 and layer 4 of V1. B) V1 layer 6 is disconnected from the model. End-stopped cells no longer emerge in layer 4 of V1 without layer 6 activity. Without end-stopped cells in layer 4, MT takes significantly longer to solve the aperture problem for a bar of the same length.

results are consistent with physiology, which shows that MT resolves the aperture problem while V1 continues to respond largely to the components of motion despite direct feedback from MT.

We believe that multiscale sampling (with or without feedback) is the key ingredient to the emergence of end-stopped cells in V1 layer 4, which, in turn, greatly facilitates the solution of the aperture problem in area MT. To give an intuitive explanation of why multiscale sampling works, consider a moving bar that elicits activity from LGN cells, which then synapse onto rightward direction-selective V1 cells. The activity in the rightward direction V1 cells is greatest in the middle of the bar where the receptive fields only perceive the horizontal component of motion. Now suppose these rightward-selective cells sample the LGN input at two different spatial scales and that the activity from the larger spatial scale is subtracted from the activity of the smaller spatial scale (this corresponds to V1 L4 cells receiving inhibition from V1 L4 interneurons, which receive their input from V1 L6 cells with larger receptive fields). The region that will be most suppressed because of this (smaller – larger receptive field) activity difference is precisely the middle of the bar. For this reason, we see that the strongest end-stopping occurs in our rightward-selective cells in V1 although some end-stopping can also be seen in right-up direction-selective cells.

While we find that feedback is not necessary for a successful solution to the aperture problem in area MT, it facilitates strong end-stopping in area V1 by providing a third spatial sampling scale. We hypothesize that the more spatial sampling scales the system is exposed to, the easier it becomes to suppress activity that does not agree between scales.

Furthermore, we show that it is possible to have a driving feedback model from area MT to V1 whereby a solution to the aperture problem in MT does not necessitate a solution in V1. This is explained by the facts that (a) MT synapses on different lamina in V1 (Sillito et al., 2006), where input to layer 6 eventually ends up inhibiting activity in layer 4 through inter-neurons, and (b) V1 is continually processing bottom-up input from LGN (which is riddled with the aperture problem) in addition to aperture problem–free feedback from area MT. Based on physiology and psychophysics (in monkeys and humans, respectively), we hypothesize that the visual system has evolved this way because there are evolutionary benefits to have one visual area (V1) preserve local motion information while another (MT) processes global motion. Local motion may be useful for more than just estimating global motion and solving the aperture problem. For example, a local motion signal from V1 can be used in other early vision tasks, such as optic flow estimation (Beauchemin & Barron, 1995) and image segmentation (Stoner & Albright, 1993). By maintaining both a local and a global registration of motion, the system remains flexible to different types of visual and cognitive tasks without binding itself to a given scale. We note that if the input from MT were modulatory (multiplicative) onto V1, V1 would solve the aperture problem as soon as MT does, which is inconsistent with physiology.

The fact that the aperture problem may be solved by fast interlaminar and interareal connections coupled with slower intralaminar and intra-areal sampling serves as a proof of concept for future work. There are several important limitations of this study. First, we note that the system of differential equations does not include temporal delay terms between different visual areas. We know that conduction latencies are different for area V1 and MT (Bullier, 2001); however, a systematic analysis of both feed-forward and feedback conduction delays is beyond the scope of this paper. Rather, we thought of the sampling kernel sizes as a surrogate measure for both spatial and temporal differences of receptive fields in V1 and MT. It remains to be seen how adding explicit temporal delays that match biological constraints impact the solution of the aperture problem in this model.

Another simplification of this work is the assumption that some filtering process enables certain cells to be sensitive to the rightward direction of motion while other cells are selective to the upward and up-right directions of motion. We introduced the concept of the direction mask as an abstraction for some process upstream that develops this selectivity biologically. A more complete model should show how motion-direction selectivity arises as an emergent property of the system and replaces the direction mask concept in this model. It is worth noting that V1 layer 6 has strong feedback connections onto LGN, which, when taken together with feedback from MT to V1 layer 6, may suggest a process by which direction-selectivity emerges (Sillito et al., 2006).

A third limitation to this work is the lack of any cross-orientation connectivity between the rightward, upward, and right-up motion-selective cells. There is evidence to suggest that this cross-orientation competition exists (Rose & Blakemore, 1974; Ferster & Miller, 2000); however, it remains to be seen what impact it would have on the aperture problem simulation. Lastly, the contribution of other visual areas as potential read-out layers for the aperture problem solution cannot be overlooked (for example, area MST has even larger receptive fields than MT). Many details of the model remain to be fleshed out; however, this work has served as a proof of concept that multiscale sampling with simple Gaussian inter-areal and intra-areal kernels is enough to solve the aperture problem. Future directions of research also include validating the model against psychophysical measurements of the aperture problem as a function of moving bar length, speed, duration of motion, and contrast (Lorenceau et al., 1992).

The contribution of this paper is to reframe motion integration as an emergent property of multiscale sampling rather than hierarchal processing of local-to-global information. Specifically, we investigated whether a simple model in which receptive fields of different spatial scales sampling a stimulus in parallel can solve the aperture problem. Our simulation results support the idea that fast, bidirectional, interlaminar and interareal sampling is the key concept that enables a network to solve the aperture problem without further need for cells of special function or receptive field shape.

Conclusion

In this paper, we presented a proof of concept that motion integration in a multiscale sampling model allows one to bypass the need for calculating intersection of constraints, propagation of signal from line ends, complicated spatiotemporal receptive fields, and other intricate methods. Moreover, the solution to the aperture problem, together with the development of end-stopped cells, pops out as an emergent property of the network. More work needs to be done to make this proof of concept biologically precise; however, we believe this multiscale sampling approach could be applied to many other classic problems in vision.

Acknowledgments

Corresponding author: Arash Yazdanbakhsh.
Email: yazdan@bu.edu.
Address: Center for Computational Neuroscience and Neural Technology, Program in Cognitive and Neural Systems, Boston University, Boston, MA, USA.

References

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

Albright, T. D., & Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Experimental Brain Research, 65*(3), 582–592.

Barthélemy, F. V., Perrinet, L. U., Castet, E., & Masson, G. S. (2008). Dynamics of distributed 1D and 2D motion representations for short-latency ocular following. *Vision Research, 48*(4), 501–522.

Bayerl, P., & Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Computation, 16*(10), 2041–2066.

Beauchemin, S. S., & Barron, J. L. (1995). The computation of optical flow. *ACM Computing Surveys (CSUR), 27*(3), 433–466.

Bolz, J., & Gilbert, C. D. (1986). Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature, 320*(6060), 362–365.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews (Review), 36,* 96–107.

Chey, J., Grossberg, S., & Mingolla, E. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America, 14,* 2570–2594.

Craft, E., Schuetze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology, 97,* 4310–4326.

Ferster, D., & Miller, K. D. (2000). Neural mechanisms of orientation selectivity in visual cortex. *Annual Review of Neuroscience, 23,* 441–471.

Friedman, H. S., Zhou, H., & von der Heydt, R. (2003). The coding of uniform colour figures in monkey visual cortex. *The Journal of Physiology, 548*(Pt. 2), 593–613.

Girard, P., Hupeì, J.-M., & Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *Journal of Neurophysiology, 85,* 1328–1331.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52,* 217–257.

Grossberg, S., & Todorovi, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics, 43,* 241–277.

Grossberg, S., & Williamson, J. R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cerebral cortex (New York, NY: 1991), 11*(1), 37–58.

Horn, B. K. R., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence, 17,* 185–204.

Lamme, V. A., Super, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology, 8*(4), 529–535.

Lorenceau, J., Shiffrar, M., Wells, N., & Castet, E. (1992). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research, 33*(9), 1207–1217.

Marr, D. (1982). *Vision.* San Francisco: W.H. Freeman.

Maunsell, J. H., Ghose, G. M., Assad, J. A., McAdams, C. J., Boudreau, C. E., & Noerager, B. D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Visual Neuroscience, 16*(1), 1–14.

Maunsell, J. H., & Van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of Neuroscience, 3*(12), 2563–2586.

Montagnini, A., Mamassian, P., Perrinet, L., Castet, E., & Masson, G. S. (2007). Bayesian modeling of dynamic motion integration. *Journal of Physiology-Paris, 101*(1), 64–77.

Pack, C. C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature, 409*(6823), 1040–1042.

Pack, C. C., Livingstone, M. S., Duffy, K. R., & Born, R. T. (2003). End-stopping and the aperture

problem: Two-dimensional motion signals in macaque V1. *Neuron, 39*(4), 671–680.

Perrinet, L. U., & Masson, G. S. (2012). Motion-based prediction is sufficient to solve the aperture problem. *Neural Computation*, 24(10), 2726–2750.

Raiguel, S. E., Xiao, D. K., Marcar, V. L., & Orban, G. A. (1999). Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. *Journal of Neurophysiology*, 82(4), 1944–1956.

Raizada, R. D., & Grossberg, S. (2001). Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition, 8*(3–5), 431–466.

Rose, D., & Blakemore, C. (1974). An analysis of orientation selectivity in the cat's visual cortex. *Experimental Brain Research*, 20(1), 1–17.

Rossi, A., & Paradiso, M. (1999). Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex. *Journal of Neuroscience, 19*(14), 6145–6156.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., et al. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology*, 79(6), 3272–3278.

Sillito, A. M., Cudeiro, J., & Jones, H. E. (2006). Always returning: Feedback and sensory processing in visual cortex and thalamus. *Trends in Neurosciences, 29*(6), 307–316.

Stoner, G. R., & Albright, T. D. (1993). Image segmentation cues in motion processing: Implications for modularity in vision. *Journal of Cognitive Neuroscience, 5*(2), 129–149.

Stumpf, P. (1911). Über die abhängigkeit der visuellen bewegungsrichtung und negativen nachbildes von den reizvorgangen auf der netzhaut [Translation: About visually perceived direct motion]. *Zeitschrift für Psychologie, 59,* 321–330.

Tlapale, É., Masson, G. S., & Kornprobst, P. (2010). Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. *Vision Research, 50*(17), 1676–1692.

Ts'o, D. Y., & Gilbert, C. D. (1988). The organization of chromatic and spatial interactions in the primate striate cortex. *Journal of Neuroscience, 8,* 1712–1727.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science, 255*(5043), 419–423.

Wallach, H. (1935). Über visuell wahrgenommene bewegungsrichtung. *Psychologische Forschung, 20,* 325–380.

Xing, D., Yeh, C. I., Burns, S., & Shapley, R. M. (2012). Laminar analysis of visually evoked activity in the primary visual cortex. *Proceedings of the National Academy of Sciences, USA, 109*(34), 13871–13876.

Yo, C., & Wilson, H. R. (1992). Moving 2-D patterns capture the perceived direction of both lower and higher spatial frequencies. *Vision Research, 32,* 1263–1270.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience, 20*(17), 6594–6611.

Appendix: Model equations

For all of the equations below, A represents the membrane potential decay rate, B stands for the depolarization threshold, and D is for the hyperpolarization threshold. The A parameter is specific to the cell population, and the B and D thresholds are 90 and 60, respectively, for all simulated cell populations. All excitatory, inhibitory, and intra-areal sampling kernels (C, E, and F, respectively) are 2-D Gaussian kernels. The absolute value of the peak amplitudes and standard deviations of the excitatory kernels is always 18 and 0.15, respectively. The inhibitory kernels always have a peak amplitude of 0.5 and a standard deviation of 1.2 (the ratio of E:I peak amplitude is always 36:1, and the ratio between the E:I standard deviation is maintained at 1:8). The size of the excitatory and inhibitory Gaussians (how many cells or units they span) varies by visual area. The scaling is accomplished by up- or down-sampling of the same baseline Gaussian kernels. The * operation denotes a convolution with the respective kernel. To simplify reading the equations, excitatory terms have been colored green, inhibitory terms red, modulatory terms blue, and intra-areal connections purple.

LGN

The population of LGN cells, LGN_{ij}, receives input from the moving bar, $I(t)$. The excitatory sampling kernel, C_{LGN}, has a peak amplitude of 18 and a standard deviation of 0.15 and spans a radius of 2 cells or units. The inhibitory sampling kernel, E_{LGN}, has a peak amplitude 0.5 and standard deviation 1.2 and spans a radius of 5 cells or units. The membrane potential decay rate, A_{LGN}, is 50 (the same as the speed of the moving bar in Hz).

$$\frac{dLGN_{ij}}{dt} = -A_{LGN} \cdot LGN_{ij} + (B - LGN_{ij})$$
$$\cdot \left(I(t) * C_{LGN} \right)_{ij} - (D + LGN_{ij})$$
$$\cdot \left(I(t) * E_{LGN} \right)_{ij} \tag{A1}$$

Because Equation A1 and the surrogate depolarization and hyperpolarization constants (B and D, respectively) describe the membrane potential (the actual spiking activity), we pass the LGN activity through a threshold linear signal function with a threshold of 30 before it can influence upstream activity. The cells meeting this threshold are also divisively normalized by the maximum activity in that population of cells (such that the most active cell that meets the threshold is scaled to one, and all other cells meeting the threshold are scaled in proportion with the maximally active cell).

V1 layer 6 rightward-selective population

The population of V1 layer 6 right cells, $v_{ij}^{1L6\rightarrow}$, receives shunted input from both LGN and right-sensitive MT cells, LGN and $MT^{\rightarrow}$, respectively. The LGN input is filtered through a right-selective mask (see Direction mask section). The excitatory sampling kernel, C_{V1}^{excite}, is twice the size of the LGN sampling kernel (spans a radius of 4 units). The inhibitory sampling kernel, E_{V1}^{inh}, is likewise twice the size of the LGN inhibitory sampling kernel (radius = 10 units). The membrane potential decay rate, A_{v1}, is 400 (tuned to fit the latency, peak amplitude, and decay rate of typical V1 cells). The excitatory intra-areal kernel, F_{V1}^{excite}, has peak amplitude 18, standard deviation 0.15, and is half of the size of the interareal excitatory kernel (radius = 2 units). The inhibitory intra-areal kernel, F_{V1}^{inh}, has peak amplitude 0.5, standard deviation 1.2, and is also half of the interareal inhibitory kernel (radius = 5 units).

$$\frac{dv_{ij}^{1L6\rightarrow}}{dt} = -A_{v1} \cdot v_{ij}^{1L6\rightarrow} + (B - v_{ij}^{1L6\rightarrow})$$
$$\times \left[\left(f(LGN) * C_{V1}^{excite} \right)_{ij} \cdot mask_{ij}^{\rightarrow}(t) \right.$$
$$\left. + \left(f(MT^{\rightarrow}) * C_{V1}^{excite} \right)_{ij} \right] - (D + v_{ij}^{1L6\rightarrow})$$
$$\times \left[\left(f(LGN) * E_{V1}^{inh} \right)_{ij} \cdot mask_{ij}^{\rightarrow}(t) \right.$$
$$\left. + \left(f(MT^{\rightarrow}) * E_{V1}^{inh} \right)_{ij} \right]$$
$$+ \left(v_{ij}^{1L6\rightarrow} * (F_{V1}^{excite} - F_{V1}^{inh}) \right)_{ij} \tag{A2}$$

The signal function, f, is again threshold linear (threshold for LGN cells, LGN, is 30; threshold for MT cells, $MT_{ij}^{\rightarrow}$, is 35) and divisively scaled by the maximally active cell.

V1 layer 4 inhibitory interneurons rightward-selective population

The population of rightward-selective V1 layer 4 interneurons, $v_{ij}^{1L4i\rightarrow}$, receives bottom-up input from LGN, modulatory input from rightward-selective V1 layer 6 neurons, and intra-areal on-center-off-surround. The decay rate and sampling kernels are the same as for V1 layer 6 cells.

$$\frac{dv_{ij}^{1L4i\rightarrow}}{dt} = -A_{v1} \cdot v_{ij}^{1L4i\rightarrow} + (B - v_{ij}^{1L4i\rightarrow})$$
$$\times \left[\left(\left(f(LGN) \cdot f(v_{ij}^{1L6\rightarrow}) \right) * C_{V1}^{excite} \right)_{ij} \right.$$
$$\left. \cdot mask_{ij}^{\rightarrow}(t) \right] - (D + v_{ij}^{1L4i\rightarrow})$$
$$\times \left[\left(\left(f(LGN) \cdot f(v_{ij}^{1L6\rightarrow}) \right) \right. \right.$$
$$\left. \left. * E_{V1}^{inh} \right)_{ij} \cdot mask_{ij}^{\rightarrow}(t) \right]$$
$$+ \left(v_{ij}^{1L4i\rightarrow} * (F_{V1}^{excite} - F_{V1}^{inh}) \right)_{ij} \tag{A3}$$

The signal function, f, is threshold linear (threshold for LGN cells, LGN, is 30; threshold for V1 layer 6 cells, $v_{ij}^{1L6\rightarrow}$, is 35) and divisively scaled by the maximally active cell.

V1 layer 4 excitatory rightward-selective population

The population of rightward-selective V1 layer 4 cells, $v_{ij}^{1L4e\rightarrow}$, receives bottom-up input from LGN, inhibitory input from V1 layer 4 interneurons, top-down input from MT, modulatory input from V1 layer 6 cells, and intra-areal on-center-off-surround. The decay rate is the same as for V1 layer 6 and V1 layer 4 interneurons, but the sampling kernels are reduced to half the size of V1 layer 6 Gaussians (excitatory kernel spans a radius of 2 units, inhibitory kernel radius = 5 units). Note the inhibitory contribution of V1 layer 4 interneuron factors in the equation in the reverse order (the inhibitory convolution is added to the excitation while the excitatory convolution is added as inhibition).

$$\frac{dv_{ij}^{1L4e\rightarrow}}{dt} = -A_{v1}\dot{s}v_{ij}^{1L4e\rightarrow} + (B - V_{ij}^{1L4e\rightarrow})$$

$$\times \left[\left(f(LGN) * C_{V1}^{excite} \right)_{ij} \cdot mask_{ij}^{\rightarrow}(t) \right.$$

$$+ \left(f(MT_{ij}^{\rightarrow}) * C_{V1}^{excite} \right)_{ij}$$

$$\left. + \left(f(v_{ij}^{1L4i\rightarrow}) * E_{V1}^{inh} \right)_{ij} \right]$$

$$\cdot \left(f(v_{ij}^{1L6\rightarrow}) * C_{V1}^{excite} \right)_{ij} - (D + v_{ij}^{1L4e\rightarrow})$$

$$\times \left[\left(f(LGN) * E_{V1}^{inh} \right)_{ij} \cdot mask_{ij}^{\rightarrow}(t) \right.$$

$$\left. + \left(f(MT_{ij}^{\rightarrow}) * E_{V1}^{inh} \right)_{ij} + \left(f(v_{ij}^{1L4i\rightarrow}) * C_{V1}^{excite} \right)_{ij} \right]$$

$$\cdot \left(f(v_{ij}^{1L6\rightarrow}) * E_{V1}^{inh} \right)_{ij}$$

$$+ \left(v_{ij}^{1L4e\rightarrow} * (F_{V1}^{excite} - F_{V1}^{inh}) \right)_{ij} \qquad (A4)$$

The signal function, f, is threshold linear (threshold for LGN cells, LGN, is 30; threshold for V1 layer 6 cells, $v_{ij}^{1L6\rightarrow}$, is 35; threshold for V1 layer 4 interneurons, $v_{ij}^{1L4i\rightarrow}$, is 25; threshold for MT cells, $MT_{ij}^{\rightarrow}$, is 35) and divisively scaled by the maximally active cell.

MT rightward-selective population

The population of MT rightward-selective cells, $MT_{ij}^{\rightarrow}$, receives bottom-up input from rightward-selective V1 layer 4 excitatory cells and intra-areal on-center-off-surround. MT's decay rate, A_{MT}, was set to 800 to tune the cells to realistic latencies, peak response times, and decay profiles. The excitatory sampling kernel, C_{MT}^{excite}, is 10 times the size of the LGN sampling kernel (spans a radius of 20 units). The inhibitory sampling kernel, E_{MT}^{inh}, is likewise 10 times the size of the LGN inhibitory sampling kernel (radius = 50 units).

$$\frac{dMT_{ij}^{\rightarrow}}{dt} = -A_{MT} \cdot MT_{ij}^{\rightarrow}$$

$$+ (B - MT^{\rightarrow}) \left[\left(f(v^{1L4e\rightarrow}) * C_{MT}^{excite} \right)_{ij} \right]$$

$$- (D + MT^{\rightarrow}) \left[\left(f(v^{1L4e\rightarrow}) \right) * (E_{MT}^{inh})_{ij} \right]$$

$$+ \left(MT^{\rightarrow} * (F_{MT}^{excite} - F_{MT}^{inh}) \right)_{ij} \qquad (A5)$$

The signal function, f, is threshold linear (threshold for V1 layer 4 cells, $v_{ij}^{1L4i\rightarrow}$, is 10) and divisively scaled by the maximally active cell.

The same equations and parameters hold for upward and right-up direction-selective cells with the exception that $mask_{ij}^{\rightarrow}(t)$ is replaced by $mask_{ij}^{\uparrow}(t)$ and $mask_{ij}^{\nearrow}(t)$, respectively. In general, there are no cross-orientation interactions; only cells of the same orientation (within and across layers) contribute the cell's activity.