Hierarchical processing of complex motion along the primate dorsal visual pathway

Patrick J. Mineault[a,1], Farhan A. Khawaja[a,1], Daniel A. Butts[b], and Christopher C. Pack[a,2]

[a]Montreal Neurological Institute, McGill University School of Medicine, Montreal, QC, Canada H3A 2B4; and [b]Department of Biology and Program in Neuroscience and Cognitive Science, University of Maryland, College Park, MD 20742

Edited by Robert Desimone, Massachusetts Institute of Technology, Cambridge, MA, and approved January 4, 2012 (received for review September 23, 2011)

Neurons in the medial superior temporal (MST) area of the primate visual cortex respond selectively to complex motion patterns defined by expansion, rotation, and deformation. Consequently they are often hypothesized to be involved in important behavioral functions, such as encoding the velocities of moving objects and surfaces relative to the observer. However, the computations underlying such selectivity are unknown. In this work we have developed a unique, naturalistic motion stimulus and used it to probe the complex selectivity of MST neurons. The resulting data were then used to estimate the properties of the feed-forward inputs to each neuron. This analysis yielded models that successfully accounted for much of the observed stimulus selectivity, provided that the inputs were combined via a nonlinear integration mechanism that approximates a multiplicative interaction among MST inputs. In simulations we found that this type of integration has the functional role of improving estimates of the 3D velocity of moving objects. As this computation is of general utility for detecting complex stimulus features, we suggest that it may represent a fundamental aspect of hierarchical sensory processing.

receptive field | optic flow

In the early stages of the primate visual system the receptive fields of neurons can be readily estimated from the responses to simple stimuli such as spots, bars, and gratings or even by hand mapping (1–3). However, for neurons farther along the visual pathways, the relationship between stimulus input and neuronal output is often far from obvious, particularly in areas that respond to complex stimuli such as faces, objects, or optic flow patterns (4–7). Uncovering this relationship is crucial for understanding the computations that underlie important behavioral functions such as object recognition and navigation.

One well-known example of complex cortical processing is the range of selectivities found in the medial superior temporal (MST) area of the primate visual cortex. Previous work has shown that MST neurons are highly selective for visual stimuli composed of combinations of motion patterns such as expansion, deformation, translation, and rotation (8–12). Although this selectivity has been documented many times over the last 25 y, very little is known about the computations by which it is derived. One prevalent hypothesis is that the selectivity of MST neurons is determined by specific strategies used by the brain to calculate one's direction of motion, or heading, through the world (13–15). In these models, heading is computed by combining the output of detectors tuned to specific motion patterns, and these patterns are reflected in the internal structure of an MST neuron's receptive field.

Although this hierarchical account of MST selectivity is appealingly simple, it has been difficult to confirm experimentally. Indeed previous studies have concluded that MST responses to complex stimuli often cannot be predicted, even qualitatively, from their responses to simple ones (7–9, 16). For example, a recent paper by Yu et al. (7) found that MST receptive field substructure failed to account for the response patterns of MST neurons to combinations of motions. This result led the authors to speculate that highly complex interactions must occur among MST inputs, perhaps involving specific wiring of dendritic compartments. Such findings call into question the simple hierarchical scheme that has been at the heart of most previous models.

In this work we have examined the hierarchical nature of MST processing, using a unique experimental stimulus and a rigorous computational framework. Specifically, we have developed a visual stimulus that efficiently and thoroughly explores the space of complex motion stimuli and used the resulting data to test MST models with different structures. We find that the most successful models take into account the specific properties of MST's most proximal source of afferent input, the middle temporal (MT) area (17–19). Furthermore, we find that such hierarchical models are capable of capturing all of the main features of MST stimulus selectivity, provided that a particular style of nonlinear integration is used to transform MT inputs into MST outputs. We show that this mechanism is consistent with the known properties of cortical neurons and that it can be expressed in a simple mathematical form. Finally, we demonstrate in simulations that this type of integration is useful for extracting the 3D velocity of objects relative to the observer, as it provides strong tuning for velocity with little dependence on other stimulus features. This work therefore provides quantitative validation of a number of existing notions about MST function, while supplying a crucial element (nonlinear integration) that has been previously missing.

Results

MST Neurons Are Tuned to Complex Optic Flow. We recorded from 61 neurons in area MST of two awake, fixating macaque monkeys. In most cases we first obtained an estimate of the neuron's selectivity for optic flow by measuring responses to the *tuning curve* stimuli depicted in Fig. 1*A*. For a given position in space, 24 tuning curve stimuli were presented, with 8 stimuli corresponding to translation (motion in a single direction), 8 corresponding to spirals (including expansion, contraction, rotation, and their intermediates), and 8 corresponding to deformation (expansion along one axis and contraction along the other). These tuning curve stimuli span the space of first-order optic flow patterns and have proved useful in characterizing optic flow selectivity in the dorsal visual stream (11, 20). These 24 tuning curve stimuli were presented at nine positions lying on a 3×3 rectangular grid that spanned most of the central 50° of the

Author contributions: P.J.M., F.A.K., D.A.B., and C.C.P. designed research; P.J.M. and F.A.K. performed research; P.J.M. analyzed data; and P.J.M., F.A.K., and C.C.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]P.J.M. and F.A.K. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: christopher.pack@mcgill.ca.

See Author Summary on page 5930 (volume 109, number 16).

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1115685109/-/DCSupplemental.

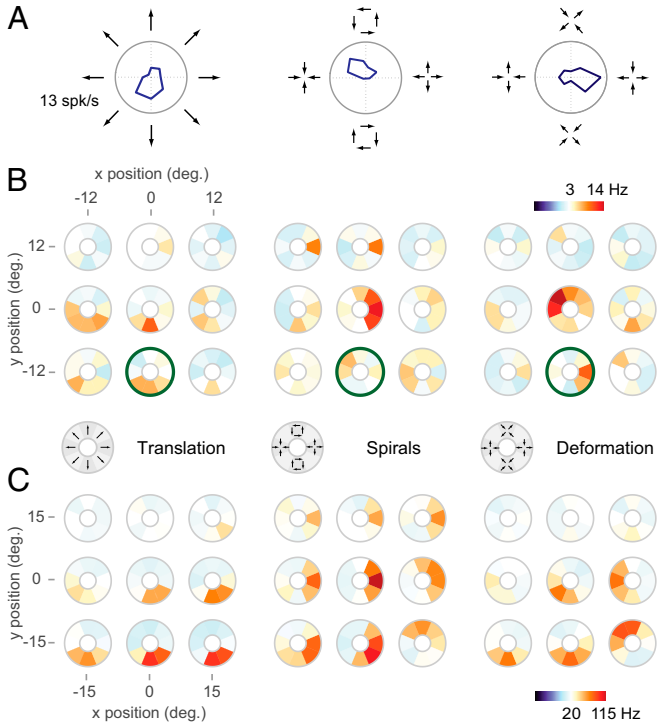


Fig. 1. Tuning of MST neurons for complex optic flow. (*A*) Tuning curves for a single MST neuron to visual motion composed of translation (*Left*), spirals (*Center*), and deformation (*Right*). Stimuli were presented at one position on a 3 × 3 grid centered on the fovea. (*B*) Tuning mosaics, in which large responses are represented by red colors, small responses by blue, and median responses by white. Each mosaic captures the tuning for one of the stimulus types shown in *A* at nine positions in the visual field. The mosaics highlighted in green correspond to the tuning curves shown in *A*. This cell consistently preferred downward translation (*Left*) and tuning for spirals (*Center*) and deformation (*Right*) varied across positions. (*C*) Tuning mosaics for a second example cell. This cell consistently preferred downward translation (*Left*) and expansion (*Center*) at most spatial positions.

visual field, allowing us to examine the positional invariance of the selectivity (21).

Fig. 1*A* shows the responses of an example MST cell to the 24 optic flow stimuli when they were displayed in the lower-middle part of the 3 × 3 grid. Here the cell preferred downward-translational motion (Fig. 1*A*, *Left*), contracting counterclockwise spirals (Fig. 1*A*, *Center*), and deformation with a horizontal divergent axis (Fig. 1*A*, *Right*). These responses are replotted in Fig. 1*B* as *tuning mosaics*, which are color-coded versions of the standard direction tuning curves. Each mosaic shows the response of a cell to 8 stimuli of a type at a given position in the receptive field, with red representing responses above baseline firing rate and blue responses below baseline. The most saturated red corresponds to maximal firing rate across all stimuli, whereas white corresponds to the median firing rate; tuning mosaics are not otherwise normalized. The mosaics outlined in green correspond to the tuning curves shown in Fig. 1*A*.

The translation mosaics (Fig. 1*B*, *Left*) indicate that this cell shows a preference for downward motion in the bottom and center portions of the screen. The spiral mosaics (Fig. 1*B*, *Center*) show that the cell's spiral tuning shifts from position to position, with the strongest preference being for expansion motion at the top and center of the visual field. A weaker response to contraction can be seen near the bottom of the visual field. The deformation mosaics (Fig. 1*B*, *Right*) show that tuning for deformation motion also varies from position to position. This cell therefore shows selectivity for a range of stimuli and a strong dependence of stimulus preference on spatial position.

In contrast, Fig. 1*C* shows a cell with tuning for expansion (Fig. 1*C*, *Center*) that is nearly invariant with spatial position. This second cell's translation tuning (Fig. 1*C*, *Left*) is similar to that of the cell in Fig. 1*B*, indicating that there is no obvious relationship between the tuning for translation and that for spirals. Thus, our results, in agreement with previous reports (8, 9), suggest that MST neurons exhibit complex tuning in a high-dimensional stimulus space. To explore this tuning in quantitative detail, we developed a stimulus that sampled the space of optic flow far more thoroughly than the tuning curve stimulus described above. Specifically, we used a *continuous optic flow* stimulus that consisted of continuously evolving, random combinations of translation, spirals, and deformation stimuli, each of which elicited robust responses from most MST neurons (Movie S1). This approach typically allowed us to measure responses to several thousand optic flow stimuli.

On the basis of the responses to this rich repertoire of stimuli, we sought to develop a quantitative account of the neuronal computations that lead to the variety and complexity of neuronal responses exemplified in Fig. 1. Our approach was to describe each neuron's responses using several mathematical models, all of which shared the same basic structure. In the first stage, the input stimulus is processed by a number of subunits, each of which is selective for motion in a part of the visual field. The output of these subunits is fed to the simulated MST neuron, which sums its inputs and translates the result into a predicted firing rate through an expansive static nonlinearity. Such linear–nonlinear cascade models have strong theoretical foundations that have been described elsewhere (22, 23).

For each MST neuron we optimized the choice of subunits to maximize the quality of the fit to the continuous optic flow data (*Methods*). We controlled the complexity of the model by cross-validation and evaluated its performance by predicting a neuron's response to the *tuning curve* stimuli, on which the model was not trained. As a check on the validity of our approach and its implementation, we verified that our methods converge to correct estimates of receptive fields in simulated data (*SI Appendix*, *SI Methods* and Fig. S1). As described in detail below, our approach allowed us to examine particular hypotheses about neuronal computation in MST.

Hierarchical Processing Partially Accounts for MST Responses. The simplest model that could in principle account for the data shown in Fig. 1 involves a computation in which MST neurons linearly compare the visual stimulus to an internal template, with the output reflecting the degree of match. This *linear model* is directly analogous to the linear spatiotemporal receptive field models that have been used in the luminance domain to study early visual areas (2, 24). Furthermore, it is mathematically tractable, and previous modeling work has shown promise in capturing the complex tuning properties seen in MST (25, 26). We found, however, that whereas such a model can capture some preference to translation, it is unable to capture the more complex selectivities of MST neurons (*SI Appendix, SI Methods* and Fig. S2).

This result may be expected, as MST neurons have no direct access to the visual stimulus, instead receiving the bulk of their input from MT neurons, which are tuned for both direction and speed (17, 27). Thus, a more promising model involves a computation in which MST neurons linearly sum the output of appropriately tuned MT subunits. Indeed this idea is implicit in many existing MST models (9, 13, 14, 16, 28). We thus developed a *hierarchical model* in which the input stimulus is first transformed into the outputs of a population of MT-like subunits tuned for stimulus direction and speed (Fig. 2*A*). The mathematical form of these subunits was chosen to provide an accurate

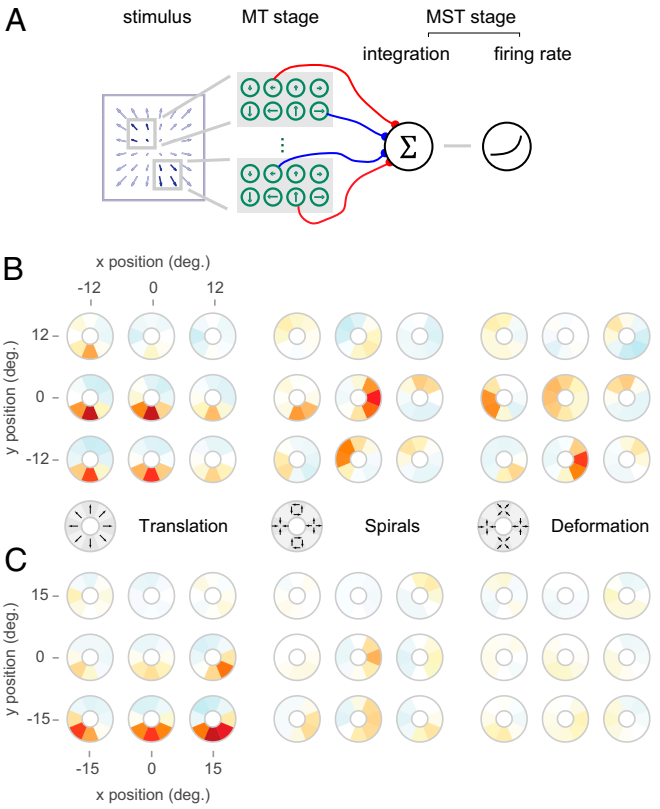


Fig. 2. Performance of the linear hierarchical model. (*A*) The stimulus was processed by groups of MT-like filters (only two groups shown for clarity), which could vary in preferred direction, spatial position, and speed. The outputs of these filters were weighted, summed, and nonlinearly transduced to a firing rate. (*B*) Predicted tuning mosaics for the same cell as in Fig. 1*B* under the hierarchical model. The hierarchical model correctly captures the optic flow tuning of this cell, including the preferences for spiral motion (*Center*). (*C*) Same as in *B* but for the example cell shown in Fig. 1*C*. The hierarchical model fails to capture this cell's tuning to complex optic flow (spirals and deformations).

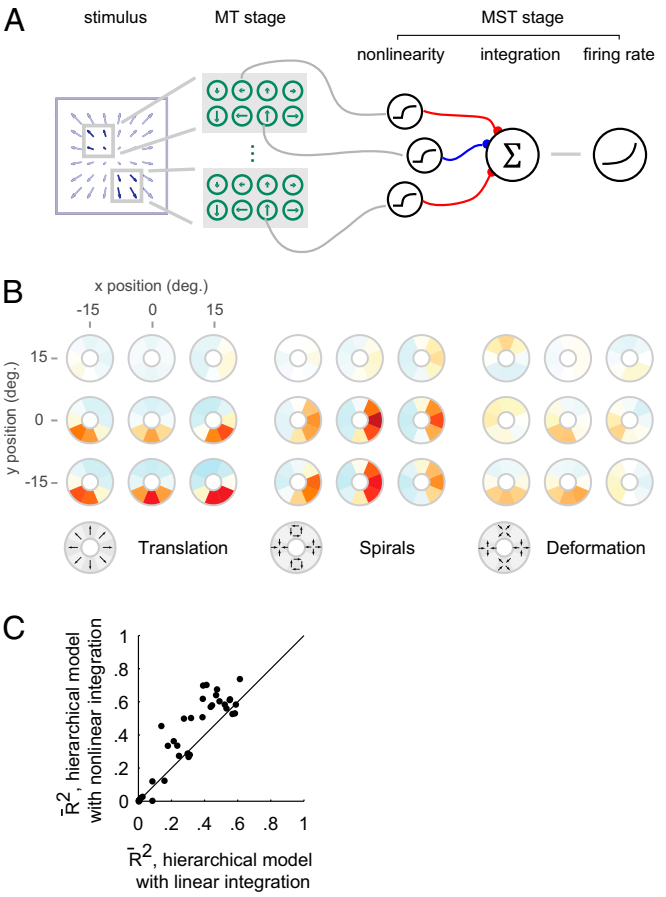


Fig. 3. Performance of the hierarchical model with nonlinear integration. (*A*) The stimulus was processed by groups of MT-like filters. The output of these filters was passed through a nonlinearity and then weighted, summed, and transduced to a firing rate. For each MST cell, the nonlinearity could vary from compressive to expansive and was identical across all subunits. (*B*) Predicted tuning mosaics for the same cell as in Fig. 1*C* under the nonlinear integration model. This model accurately captures the tuning and relative response levels of this cell to translation and spirals. (*C*) Quality of tuning curve predictions for the hierarchical model with and without nonlinear integration. The nonlinear integration model improves performance in 75% of the tested cells.

and parsimonious account of the responses of real MT cells. Specifically, MT subunits had receptive fields that were smaller than those found in MST and responses that were tuned for direction and speed, with bandwidths matching those found in real MT cells (*SI Appendix*, *SI Methods*).

Fig. 2*B* shows the predicted tuning curves under this hierarchical model for the example cell shown in Fig. 1*B*. In this case the model captures the tuning, including the general preference for downward translation (Fig. 2*B*, *Left*) and the variety of selectivities for spiral and deformation motion (Fig. 2*B*, *Center* and *Right*). The quality of the prediction can be assessed using $\bar{R}^2$, the proportion of explainable variance accounted for (29) (*Methods*). For this example cell, $\bar{R}^2 = 0.55$, which compares favorably with results reported previously in other areas (30–33). Across the MST population, however, the model fared considerably worse, with median $\bar{R}^2 = 0.31$. Indeed we found some cells with tuning characteristics that could not be explained even qualitatively with this model structure, and the neuron originally shown in Fig. 1*C* is an example of this category. Fig. 2*C* shows that, whereas the hierarchical model successfully captures this cell's tuning for translation (Fig. 2*C*, *Left*), it consistently underestimates the responses to spiral stimuli (Fig. 2*C*, *Center*). This pattern of errors in the hierarchical model was common across our population of cells, being present in 58% of the cells (21/36, stimulus class comparisons, $P < 0.001$) (*Methods*). Thus, we conclude that, although a hierarchical model can account for some MST tuning properties, there is strong evidence that such a model responds too strongly to translation and too weakly to complex optic flow.

Nonlinear Integration Is Necessary to Explain MST Stimulus Selectivity. Stated in more general terms, the stimulus selectivity of the hierarchical MST model is too similar to that of its inputs, and there appears to be no spatial arrangement of inputs that can bring this model into closer agreement with the data. This result suggests that MST selectivity requires a nonlinear operation that transforms the output of one area before summation by the next (2, 18, 34); indeed such a mechanism has been proposed in other contexts throughout the primate visual system (3, 5, 30, 35). We therefore examined the consequences of adding a nonlinearity (Fig. 3*A*) that shaped the output of each MT subunit. In particular, we added a flexible, static nonlinearity, represented by a single free parameter β, that could be either compressive ($\beta < 1$) or expansive ($\beta > 1$) (*Methods*). For each MST cell the nonlinearity was constrained to be identical across subunits.

Remarkably, this minimal change to the hierarchical model structure yielded far better fits to the data (Fig. 3*B*) for the ex-

pansion-selective cell originally presented in Fig. 1*C*. In particular, the *nonlinear integration model* showed enhanced responses to optic flow stimuli such as expansion and rotation, while maintaining strong tuning for translation, with an overall increase in the goodness-of-fit from an $\bar{R}^2$ = 0.41–0.70. This improved fit to the data was not a trivial consequence of the additional free parameter, as the model was evaluated with a validation procedure (defined in *Methods*) that was robust to the overall model complexity. Fig. 3*C* shows that predictions improved for the majority (75%) of MST cells from which we recorded, with the median goodness-of-fit improving from 0.31 to 0.50. These improvements are also reflected in the cross-validated goodness-of-fit measured with the continuous optic flow stimulus, shown in *SI Appendix*, Fig. S3*B*. Similar results were obtained if we allowed each subunit to have its own nonlinearity (Table 1, unrestricted nonlinear model) (*SI Appendix*, *SI Methods*), suggesting that the shared nonlinearity is sufficient.

In principle there are two ways in which the introduction of nonlinear integration could improve the fit of the model to the data. The first would be to increase the overall *level* of responses to spiral and deformation stimuli relative to translation stimuli, while preserving the shape of tuning curves within stimulus categories. This modulation would compensate for the above-mentioned tendency of the hierarchical model to underestimate firing rates for spiral and deformation stimuli. The second would be to improve the ability of the model to match the *shape*s of the tuning curves, apart from overall response levels for individual stimulus classes. To untangle these two factors, we performed an additional analysis after first normalizing the responses within each stimulus class (translation, spirals, and deformation). *SI Appendix*, Fig. S3*C* shows that the nonlinear integration model still improves the quality of predictions in 78% (28/36) of the cells (stimulus class comparisons; *Methods*). This result indicates that the nonlinear integration model captures aspects of the MST responses that cannot be related simply to stimulus-specific level modulation. Rather the nonlinear integration mechanism is necessary for producing the stimulus selectivity seen in MST responses to optic flow.

We also verified that the success of the model was not influenced by errors in the centering of the stimuli, as stimulus position profoundly affects MST stimulus selectivity (20). We estimated receptive field centers from the tuning curve stimuli and compared the quality of model fits for recordings in which the stimuli were well centered (within 7° of the centers) and those in which the centering was worse (12.4° on average). The addition of the nonlinearity improved the model fits for both groups of neurons (15/19 in the first group, 12/17 for the second group), indicating that our conclusions about nonlinear integration are robust to stimulus centering. Indeed the results were noticeably better when the stimulus was well centered (median $\bar{R}^2$ = 0.56 for well-centered cases and 0.36 when the centering was worse), which indicates that the model captures the bulk of the selectivity in the center of the receptive field.

Substructure of MST Receptive Fields. The success of the nonlinear modeling approach allowed us to examine the types of subunit arrangements that were recovered for each neuron. Fig. 4*A* shows the subunits that contribute most critically to the highly nonlinear neuron shown in Fig. 3*B* (*SI Appendix*, *SI Methods*). Each circle in Fig. 4*A* corresponds to the position and size of a single MT subunit's receptive field; the direction of each arrow indicates the preferred direction of the subunit; the opacity of the color indicates the weighting; and the color denotes the sign of the contribution, with red being excitatory and blue being inhibitory. The results of this analysis show that this MST neuron's response is largely explained by the selectivity of subunits tuned to downward-left motion in the bottom left portion of the visual field and downward-right motion in the bottom right. This result is consistent with this cell's tuning for both expansion and downward motion.

For some MST cells the subunit nonlinearity was less critical, and an example of this type of receptive field is illustrated in Fig. 4*B* (same cell as in Fig. 1*B*). Here the cell's receptive field is summarized by a single downward-tuned, centrally located subunit. This cell's nonlinearity had an exponent of 0.6, closer to unity than most neurons in the MST sample (see below for details); the quality of the prediction went from 0.55 to 0.62 with the additional nonlinearity, a comparatively small change. Thus, this MST cell's response properties were similar to those found in MT.

The receptive fields of three more MST neurons are shown in Fig. 4 *C*–*E*. Like the cell originally shown in Fig. 1*C*, these three cells are selective for expansion at multiple positions in the visual field. However, despite the similarity in the tuning, the most critical subunits of these neurons revealed a variety of receptive field substructures. In particular, the position and relative motion directions of the subunits varied substantially from cell to cell, suggesting that these MST cells are not detectors of expansion per se. Rather, the selectivity of these cells appears to be captured by nonlinear combinations of a small number of excitatory and inhibitory inputs,. Estimated time filters and additional examples of receptive fields are shown in *SI Appendix*, Fig. S6, and the tuning mosaics and predictions for more MST cells are shown in *SI Appendix*, Fig. S7.

As can be seen in Fig. 4, another prominent feature of MST receptive fields is the spatial overlap of the subunits. Although differences in direction and speed preference tended to increase with spatial distance between subunits (*SI Appendix*, Fig. S8 *D*

Table 1. Summary of quality of fits of all models considered

Model	Median LL/s, continuous stimulus (median % difference relative to nonlinear MT model)	Median R^2, tuning curve stimuli (median % difference relative to nonlinear MT model)
Linear	0.38 (–54)	0.19 (–48)
MT	1.11 (–9)	0.31 (–21)
Nonlinear MT	1.23	0.50
Nonlinear MT (unrestricted)	1.23 (2)	0.45 (–6)
Divisive surround	1.20 (4)	0.48 (–1)
Asymmetric surround	1.12 (–2)	0.34 (–15)
Nonlinear asymmetric surround	1.22 (5)	0.45 (0)
Subtractive surround	1.09 (–3)	0.36 (–15)
Nonlinear subtractive surround	1.22 (4)	0.47 (–1)

Goodness-of-fit for continuous stimulus is defined as cross-validated log-likelihood accounted for per second of data. Quoted percentage values are the median ratio of goodness-of-fit for target model divided by goodness-of-fit for nonlinear MT model. Note that the ratio of medians is not necessarily equal to the median of individual ratios. LL, log-likelihood.

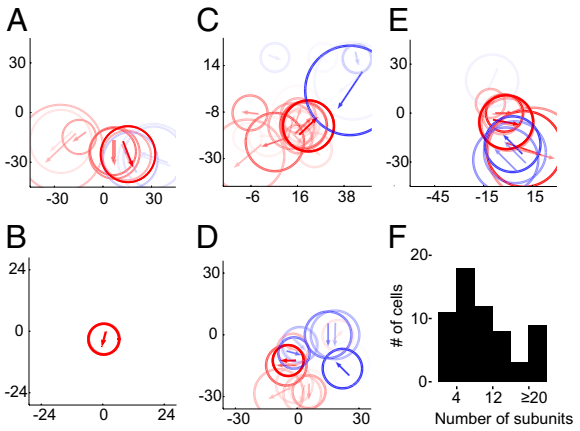


Fig. 4. Diversity of receptive field substructures in MST. (*A*) Receptive field substructure for the example cell shown in Fig. 1*C*. This visualization was produced by constructing a compact representation of the subunits in the nonlinear integration model. Red represents excitatory input, blue inhibitory input, opacity the magnitude of the weight of the subunit, and the direction of the arrow the preferred direction of the subunit. This cell's tuning for downward motion and expansion is explained by downward-left–tuned subunits in the lower left portion of the visual field and downward-right–tuned subunits in the lower right. (*B*) Substructure of example cell shown in Fig. 1*B*. This cell's receptive field was composed of a single, downward-left–tuned subunit. (*C*–*E*) The most critical subunits for three expansion-tuned cells. Whereas these cells and the one presented in Fig. 5*A* can all be described as expansion tuned, they show a diversity of receptive field arrangements. (*F*) Histogram of number of subunits found by the visualization procedure.

and *E*), there was also substantial variation on spatial scales smaller than a single subunit (e.g., Fig. 4*C*). This variation may be important for estimating optic flow quantities such as motion parallax, in which multiple motion vectors occur at nearby spatial locations. More generally, the complex selectivity observed here is likely to be useful in natural contexts, in which motion patterns are determined in part by the structure of the surrounding environment and hence are not constrained to resemble the canonical flow fields typically used experimentally. Overall these results parallel the finding that selectivity for analogous stimuli (e.g., non-Cartesian gratings) in the ventral stream of the visual cortex is related to selectivity for combinations of orientations or other features (4, 30, 36).

As suggested by Fig. 4, the number of subunits recovered by the model differed from cell to cell. This variability is summarized in Fig. 4*F*, which shows that the number of subunits contributing significantly to individual MST neurons ranged from 2 to 45, with a median value of 9 (*SI Appendix*, *SI Methods*). Most of these subunits were excitatory, with a median proportion of excitatory subunits of 81% across our population of cells. The remaining inhibitory subunits can be interpreted either as removal of excitation from tonically active MT cells or as indirect MT influences via MST interneurons, as interareal projections are almost exclusively excitatory. These conclusions are of course contingent upon the assumptions underlying our modeling approach. However, for the most part these assumptions are quite conservative, and, as we show in the next section, relaxing them does not change the main results.

Importance of Compressive Nonlinearities Across the MST Population. Although Fig. 4 shows that the receptive field substructure varied substantially from cell to cell, we found that the shape of the nonlinearity recovered by the model was highly consistent across neurons. This result is illustrated in Fig. 5*A*, which plots the distribution of the parameter β for all of the cells in our MST population. The distribution is heavily skewed toward values <1, as shown earlier in individual examples, suggesting that a *compressive* input nonlinearity is an important property of MST neurons.

Influence of Surround Suppression. Given the importance of the compressive nonlinearity in accounting for the MST data, we next sought to relate it to potential physiological mechanisms. One important candidate mechanism is *surround suppression* at the level of MT (37–40). Surround suppression attenuates the responses of MT neurons to pure translation, and so it might account for the above-mentioned observation that the compressive nonlinearity decreases the relative influence of translation on MST responses (Fig. 3*B*). We therefore extended the model output for each MT subunit to include divisive modulation (41) by a suppressive field that could vary in terms of its spatial extent, its tuning to motion, and its strength. We defined these quantities as free parameters and allowed the model to specify which characteristics best fit the data (Fig. 5*B*) (*SI Appendix*, *SI Methods*).

The results of these simulations indicate that in most cases the optimal surround was well tuned for motion direction and, surprisingly, that it covered a spatial extent similar to that of each subunit's excitatory receptive field (*SI Appendix*, Fig. S4*A*). In other words the suppressive influence recovered by the model was typically identical to the excitatory influence, so that stimuli that activated a subunit also limited its output. This type of suppressive mechanism is mathematically indistinguishable from a pure compressive nonlinearity. Indeed the full center-surround model yielded little or no improvement in the quality of the fits relative to the simple nonlinear integration model (*SI Appendix*, Fig. S4*B*, and Table 1). Similar results were obtained if we used spatially asymmetric surrounds (40), symmetric surrounds that interacted with the centers via subtraction (34, 38) rather than division, and surrounds that had their own output nonlinearities (*SI Appendix*, *SI Methods*). Although these models generally performed better than the linear integration model, none consistently outperformed the one-parameter nonlinear integration model. These results are summarized in Table 1.

Of course these results do not contradict the important role for MT surrounds in motion processing (38, 42), but they do suggest that the contribution of these surrounds to MST optic flow selectivity might be fairly subtle; we return to this issue in the *Discussion*.

Computational Properties of Nonlinear Motion Integration. Intuitively the compressive nonlinearity has a straightforward in-

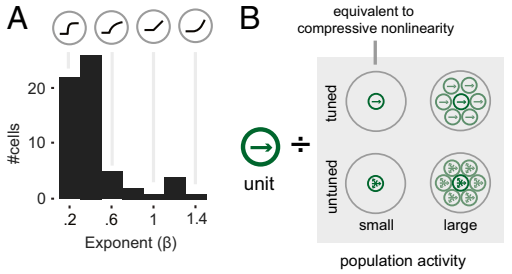


Fig. 5. Analysis of optimal subunit nonlinearity across the MST population. (*A*) In the nonlinear integration model, subunit outputs were processed by a nonlinearity of the form $f(x) = \max(0, x)^{\beta}$. β values <1 correspond to a compressive nonlinearity, whereas values >1 indicate an expansive nonlinearity. Most MST cells required a compressive nonlinearity at the level of each subunit. (*B*) In the divisive surround model, the output of the center of subunits is divided by the output of a pool of subunits differing in tuning bandwidth and spatial extent. A strongly tuned divisive surround with small spatial extent is equivalent to a static compressive nonlinearity.

terpretation: As the input to an individual subunit increases, the output saturates quickly, and as a consequence the MST cell responds best to stimuli that drive many different subunits, even if each subunit is activated weakly. This mechanism thus favors stimuli, such as complex motion, that activate many subunits.

This operation is similar to multiplicative subunit interactions described in other contexts (43–46). That is, the compressive nonlinearity is similar to a logarithm (*SI Appendix*, Fig. S3*A*), and thus the combination of compressive input nonlinearities and expansive output nonlinearity approximates multiplication through the identity $a \cdot b = \exp(\log a + \log b)$. Indeed, we verified in additional simulations that explicit multiplicative interactions between subunits outperformed models of similar complexity in 79% of the MST cells (*SI Appendix*, *SI Methods* and Fig. S5).

To quantitatively examine the functional utility of this mechanism we used optimal linear decoding to measure the ability of area MST to represent stimulus information, with and without the nonlinear integration mechanism in place. Specifically, we used our model MST cells to estimate the responses to various stimuli and then trained a simple decoding algorithm to extract various quantities from the population response (Fig. 6*A*). This method provides insight into the type of information that would be available to a brain region that had access to the output of the MST population (47, 48).

In our simulations the model MST population responded to a series of discrete objects moving in various directions and speeds, in various positions in the visual field (Fig. 6*B*). The goal of the decoder was to recover the different components of each object's velocity, independently of its position in visual space. Although we have not explored more complex situations involving different visual environments and observer motion, the position-invariant readout of 3D object velocity is necessary for common behavioral situations, such as vergence eye movement control (49, 50) and estimation of time to contact (51).

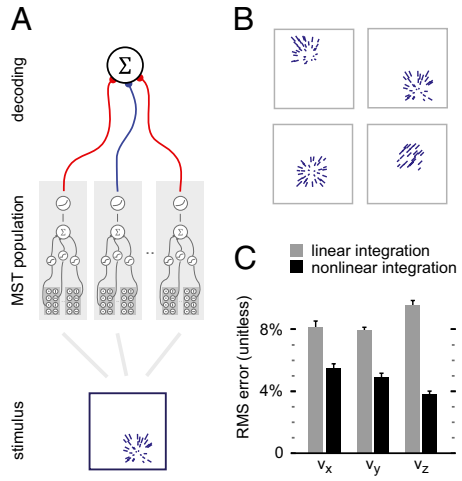


Fig. 6. Role of nonlinear integration revealed by population decoding. (*A*) In a decoding simulation, stimuli were processed by a population of MST model cells estimated from the recorded data. The goal of the linear decoder (*Top*) was to deduce physical parameters of the stimulus on the basis of the output of the MST population. (*B*) Example stimuli used in the object-decoding simulation corresponding to motion of an object in three dimensions. (*C*) Performance of the decoder based on input from the hierarchical model population with (black bars) and without (gray bars) nonlinear integration. Results are quantified as the mean error relative to the range tested; smaller values indicate better performance. Error bars indicate 1 SD from the mean, determined through a resampling procedure (*Methods*). The sensitivity of the nonlinear integration mechanisms to combinations of inputs facilitates the decoding of object velocity on the basis of the output of the MST population.

The results of this simulation (Fig. 6*C*) show that the model with nonlinear integration of MT inputs (Fig. 6*C*, black bars) outperforms the linear hierarchical model (Fig. 6*C*, gray bars) in reconstructing velocity in all three dimensions. The difference is especially large (a 60% drop in reconstruction error) in the case of the *z*-component of the velocity, which is defined by expansion optic flow. As mentioned above, the nonlinear integration approximates a multiplicative operation that renders the model less sensitive to the individual components of expansion stimuli, which are ambiguous with respect to the speed of motion in depth. This result suggests that the nonlinear aspects of MST motion encoding are useful for functions that rely heavily on measurement of motion in depth and for which retinal position is relatively unimportant (*Discussion*).

Discussion

Hierarchical Encoding of Visual Stimuli. In this work we have found that neurons in area MST can be effectively characterized by a hierarchical model that takes into account the properties of neurons in MT. An important result from this work is that cells with similar stimulus selectivity, as assessed by relatively low-dimensional tuning curve stimuli, can have subunit structures that differ significantly (Fig. 4). Although we cannot say that the subunits recovered by our model correspond exactly to the anatomical inputs received by each MST neuron, they do represent an optimal estimate under a conservative set of assumptions about MT responses. The formidable challenges associated with a direct characterization of the feed-forward inputs to the extrastriate cortex (52) suggest that a model-based approach is particularly valuable.

In addition to a plausible subunit representation, the model requires a nonlinear integration mechanism, which for most neurons is compressive (Fig. 5). Functionally, the compressive nonlinearity appears to be useful primarily for implementing a multiplicative operation similar to that seen in other visual cortical areas (5) and in sensory processing in other species (43–45). A similar approach has recently been proposed to account for the pattern and speed selectivity of MT neurons (53) and for shape selectivity in V4 (30). Indeed a similar idea was suggested as a qualitative account of optic flow tuning in MST (9). To the extent that the tuning properties found in different brain regions share the same nonlinear integration mechanism, one might expect to find that they share similar temporal dynamics (54, 55) and contrast dependencies (56); these predictions will be tested in future work.

In a complementary analysis, we tested the hypothesis that the compressive effect could be a result of center-surround interactions (37–40). We tested a wide variety of interaction types (Table 1), with the result that no mechanism consistently outperformed the simple nonlinear model. Moreover, the surrounds recovered by the model were typically the same size as the centers, suggesting that a spatially extended surround is not necessary to account for MST optic flow selectivity. A likely functional rationale for these surrounds is in performing motion segmentation and shape from motion (42).

Regardless of its precise functional interpretation, the compressive nonlinear operation could plausibly be implemented through inhibitory interactions among MT neurons with similar receptive field positions and stimulus selectivities; a similar "self-normalization" operation at the level of V1 has been posited to be of primary importance in explaining selectivity in MT cells (18, 34, 57). An alternate explanation is synaptic depression at the level of the MT–MST synapse (58). Both mechanisms are equivalent to a compressive static nonlinearity for slowly varying inputs (59). However, self-normalization would have visible effects on the tuning of MT cells, including bandwidth broadening. Given the current knowledge of MT, synaptic depression appears somewhat more plausible and would reconcile our use of a compressive nonlinearity with previous work showing that ex-

pansive output nonlinearities are sufficient for modeling the MT output (18). On the other hand, our results are unlikely to arise from contrast normalization or untuned surround suppression at the level of MT (*SI Appendix*, Fig. S4*A*).

An alternative explanation for the compressive effect is a form of normalization among MST neurons. A number of different nonlinear tuning operations can be performed through the interplay of feed-forward excitation and divisive normalization (60), including multiplicative input interactions. Although it is reasonable to assume that normalization shapes MST responses given its important role in areas V1 and MT (18, 34, 41, 61), the nature of the normalization pool in MST is unexplored, and as a result it would be difficult to incorporate into our model.

Previous MST models include those that are linear in the velocity domain (25, 26) and those that derive their selectivity primarily from the spatial arrangement of MT-like inputs (11, 13, 14, 28), as well as other more informal proposals (7, 9, 16). Each of these models is capable of reproducing certain qualitative aspects of the MST data, but to date there has been no statistical comparison of different model classes. Most recently, Yu et al. (7) attempted to estimate MST receptive field substructure by stimulating each cell with a small set of 52 canonical optic flow patterns. These authors concluded that the failure of the resulting receptive field models to account for tuning to complex optic flow stimuli implied that MST stimulus selectivity might result from an unknown mechanism that is sensitive to specific pairwise interactions within MST receptive fields.

Although this idea is of course possible, there are two main methodological shortcomings in the Yu et al. (7) work. First, the use of a small stimulus set permitted very limited inference power; our results suggest that thousands of different stimuli are necessary to estimate MST receptive field substructure. Second, the model-fitting approach implemented by the authors involved a comparable number of data points and free parameters and hence would be unlikely to generalize to novel stimuli even with a sufficiently rich training dataset. We therefore suggest that the previously reported lack of correspondence between receptive substructure and stimulus selectivity is not due to any intrinsic feature of MST, but rather to the stimulus and modeling methods used in that study.

Decoding of MST Population Activity. Functionally, MST neurons are likely to be involved in navigation (14, 62, 63). Indeed, many previous MST models have assumed that MST receptive fields are arranged to compute heading angle during self-motion (13, 14). However, our nonlinear integration model suggests that the properties of MST neurons reflect a more general mechanism that allows MST to participate both in heading and in 3D velocity estimation. Indeed, in naturalistic scenes, heading and object velocity often cannot be estimated separately (64).

In addition to heading, MST is likely involved in controlling tracking eye movements that maintain fixation on moving objects (50). Such eye movements require accurate estimates of motion direction, and our simulation results (Fig. 6*C*) suggest that the estimation of 3D object velocity relies critically on the computational properties we have identified in MST. Specifically, whereas frontoparallel motion can be recovered with reasonable accuracy by the MT population, accurate calculation of the velocity of motion in depth requires the nonlinear integration mechanism of the kind used by MST neurons. Consistent with this idea, previous work has shown that MST is important for estimating object velocity (49) and lesions of MST impair vergence movements (50).

Our simulations (Fig. 6*C*) show that a position-independent estimate of 3D velocity can be readily extracted from the output of the MST population and that nonlinear integration improves such estimates substantially. Thus, our findings indicate that nonlinear integration allows MST to form a distributed representation of 3D objects that supports a wide range of behaviors through a simple decoding mechanism (47, 48).

Methods

Electrophysiological Recordings. Two rhesus macaque monkeys took part in the experiments. Both underwent a sterile surgical procedure to implant a titanium headpost and a plastic recording cylinder. Following recovery the monkeys were seated in a custom primate chair (Crist Instruments) and trained to fixate on a small red spot on a computer monitor in return for a liquid reward. Eye position was monitored at 200 Hz with an infrared camera (SR Research) and required to be within 2° of the fixation point for the reward to be dispensed. All aspects of the experiments were approved by the Animal Care Committee of the Montreal Neurological Institute and were conducted in compliance with regulations established by the Canadian Council on Animal Care.

We recorded from well-isolated single neurons in the MST area. Single waveforms were sorted on-line and then resorted off-line, using spike-sorting software (Plexon). MST was identified on the basis of anatomical magnetic resonance imaging (MRI) scans and its position relative to MT (just past MT during a posterior approach to the superior temporal sulcus). Most of the neurons from which we recorded had large receptive fields that extended into the ipsilateral visual field and that responded to expansion and rotation stimuli in addition to translation. These tuning properties suggest that most of our recordings were from the dorsal, rather than the ventral, portion of MST, but this has not been verified histologically.

Procedure and Visual Stimuli. Upon encountering a well-isolated MST neuron, we performed a preliminary receptive field mapping with flashed bars and dot fields. For any neuron that was visually responsive, we characterized its responses in terms of *tuning curves* for three optic flow types: translation, expansion/rotation (spirals), and deformation (eight measurements per optic flow type; see *SI Appendix, SI Methods* for equations). Random-dot stimuli were presented in a 24° or a 30° aperture at nine different spatial positions on a 3 × 3 grid with adjacent center positions 12° or 15° apart. The grid was placed over the approximate center of the receptive field as determined by preliminary hand mapping.

To explore the space of optic flow stimuli more thoroughly, we also developed a novel *continuous optic flow* stimulus consisting of dots moving according to a continuously evolving velocity field generated by random combinations of six optic flow dimensions (see *SI Appendix, SI Methods* for equations). Dots moving according to this velocity field were presented in a circular aperture 24° or 30° wide, which moved slowly around the screen (Movie S1). The stimulus was presented for 6–10 min.

In all cases, dots were 0.1° in diameter at a contrast of 100% against a dark background. The screen subtended 104° × 65° of visual angle at a distance of 32 cm. The stimuli were presented at a resolution of 1,920 × 1,200 and refreshed at frame rates of 60 or 75 Hz. During continuous stimulus presentation, the animal was rewarded after maintaining fixation for 1 s.

Models. To understand the computations underlying MST optic flow selectivity, we fitted the continuous optic flow data from each cell to models with various types of subunits. In all cases we first binned the spike trains at 50 ms resolution and excluded time periods during which more than half of the stimulus was off the screen or the animal's gaze deviated >1.5° from the fixation point, as well as from 100 ms before loss of fixation to 250 ms following recovery of fixation. This method yielded a series of firing rates, which we describe as a response vector **y**. For the model, we assumed that this response was generated by a Poisson process with rate **r**, computed deterministically from the stimulus. The log-likelihood of the model $L(\mathbf{y}, \mathbf{r})$ is then given up to an additive constant (22) by

$$L(\mathbf{y}, \mathbf{r}) = \log p(\mathbf{y}|\mathbf{r}) = \sum_t y_t \log(r_t) - r_t. \quad [1]$$

We assumed that the firing rate was given by the rectified output of the receptive field acting on the stimulus, $r_t = g(\eta_t)$. g must be nonnegative for r to be meaningful; additional constraints on the derivatives of g are required to yield a model that is straightforward to optimize (22, 65). We thus chose $g \equiv \exp$.

The spatiotemporal receptive field acted on the stimulus to yield a response η_t:

$$\eta_t = c + \sum_\tau F(\rho(t,x,y), \theta(t,x,y)) w(t-\tau). \quad [2]$$

Here, $F(\rho, \theta)$ is a nonlinear spatial filter that acts on the optic flow stimulus, which is described by the local motion speed $\rho(t, x, y)$ and direction $\theta(t, x, y)$. c is a constant offset. We sampled the stimulus at a spatial resolution of 24 × 24 samples, generally covering from 48° to 60° of visual angle. The temporal filter $w(\tau)$ was assumed to last five time steps, spanning from −50 ms to −250 ms. This formulation embodies an assumption of separable, linear temporal

processing, which is supported by earlier studies of the temporal behavior of MST neurons (66).*

The nonlinear spatial filter $F(\rho, \theta)$ was assumed to be given by the sum of M nonlinear subunits $f(\rho, \theta, \mathbf{p}^m)$, where $\mathbf{p}^m$ denotes the parameters of the mth subunit:

$$F(\rho,\theta) = \sum_{m=1}^{M} f(\rho,\theta,\mathbf{p}^m). \quad [3]$$

We examined the compatibility of the data with several different models, each of which was defined by the structure of its subunits.

Hierarchical model. This model embodies the assumption that MST responses are approximately linear in terms of their feed-forward input from area MT, which provides one of the strongest projections to MST (27). The tuning of the modeled subunits is determined by three components. Subunits were assumed to have log-Gaussian speed tuning with preferred speed p_ρ:

$$R\Big(\rho(x,y),p_\rho\Big) = \exp\Big(-\Big(\log(\rho(x,y)+1)-p_\rho\Big)^2/2\sigma_\rho^2\Big) - \exp\Big(-\Big(\log(\rho(x,y)+1)+p_\rho\Big)^2/2\sigma_\rho^2\Big). \quad [4]$$

Note that a second log-Gaussian is subtracted from the first to constrain the response to be zero when there is no motion. Although MT cells tuned to low speeds have robust responses to static stimuli (67), we did not model such responses, as our stimulus poorly sampled slow speeds. We set the speed tuning width to $\sigma_\rho = 1$, similar to the mode of the distribution of speed tuning widths reported in MT (68).

The direction tuning of the subunits was given by a Von Mises function with preferred direction p_θ:

$$D(\theta(x,y),p_\theta) = \exp(\sigma_\theta \cos(\theta(x,y)-p_\theta)) - 1. \quad [5]$$

The value 1 is subtracted from the result so that by convention a stimulus moving in a direction orthogonal to the preferred direction elicits no response, and a stimulus moving in the nonpreferred direction elicits a negative response; a similar convention was used in previous models of MST (14, 15). The bandwidth parameter was chosen to be $\sigma_\theta = 2.5$, corresponding to a full-width at half-maximum bandwidth of 86°, similar to the mean value of 83° measured with moving random dots reported in ref. 19. Finally, subunits had a Gaussian spatial profile

$$G\Big(x,y,p_x,p_y,p_\sigma\Big) = \exp\Big(-\Big((x-p_x)^2+\Big(y-p_y\Big)^2\Big)\Big/2p_\sigma^2\Big). \quad [6]$$

The direction, speed, and spatial response of the subunits were combined to form the response of the subunit:

$$f(\rho,\theta,\mathbf{p}) = p_g h\Big(\sum_{xy} R\Big(\rho(x,y),p_\rho\Big)\cdot D(\theta(x,y),p_\theta)\cdot G\Big(x,y,p_x,p_y,p_\sigma\Big)\Big). \quad [7]$$

Here p_g denotes the gain of the subunit, and the function $h(x) = \max(x, 0)$ returns the positive part of the response (half-wave rectification).

Hierarchical model with nonlinear integration. This model provides each subunit with a nonlinearity that exhibits either compressive or expansive behavior depending on a free parameter (expansive when $\beta > 1$, compressive when $\beta < 1$). Subunits take the same form as Eq. **7**, but with the nonlinearity replaced by $h(x) = \max(x, 0)^\beta$. This model reduces to the previous model when $\beta = 1$. Importantly, β is shared across all subunits for a given model fit. In practice, we fitted the model for seven different values of β ranging from 0.2 to 1.4 and selected the optimal β for a cell on the basis of the cross-validated likelihood.

Model Fitting. Estimating the models described above is challenging, as they contain many free parameters and must be fitted with rather noisy data. To constrain the parameters and to obtain fits that extrapolate well to novel data, the fitting procedure must limit the dimensionality of the model. This dimensionality reduction is typically done by including explicit assumptions about the parameters (23). A particularly powerful assumption is that a model is sparse, meaning that most of its parameters are zero (69). In a neurophysiological context, this corresponds to the assumption that only a modest number of subunits are driving a given cell, which is consistent with anatomical and correlation studies of early sensory areas (70, 71). Models fitted with assumptions of sparseness have proved increasingly useful in estimating the receptive field properties of high-level neurons (22, 31, 33). We thus used gradient boosting, a stepwise fitting procedure that introduces an assumption of sparseness (69). The number of free parameters was limited through fivefold cross-validation (23).

Validation and Accuracy Metrics. For those cells for which the continuous optic flow stimulus spanned the spatial range of the tuning curve stimulus (36/61), we predicted the responses to the tuning curve stimuli on the basis of the continuous optic flow fit. Note that the continuous optic flow stimulus samples a large, six-dimensional space of optic flow, of which the tuning curve stimuli comprised a small number of points. Thus, this approach is a rigorous test of the model's ability to extrapolate to novel stimuli.

For these simulations we ignored the temporal component of the responses, instead predicting the total spike count in response to a stimulus. We allowed the gain and baseline firing rate to be estimated from the data using standard techniques (65) rather than predicted from the continuous optic flow stimulus. Given a predicted response $\mathbf{r}$ and an observed response $\mathbf{y}$, the quality of the prediction may be assessed using the standard R^2 metric of variance accounted for:

$$R^2 = \frac{\mathrm{Var}(\mathbf{y}) - \mathrm{Var}(\mathbf{y}-\mathbf{r})}{\mathrm{Var}(\mathbf{y})}. \quad [8]$$

In practice the value $R^2 = 1$ cannot be attained, as $\mathrm{Var}(\mathbf{y} - \mathbf{r})$ for a perfect prediction is the variance of the noise, which is nonnegligible in physiological measurements. To recover a natural scale we thus used a corrected R^2 metric, also known as predictive power (29):

$$\bar{R}^2 = \frac{\mathrm{Var}(\mathbf{y}) - \mathrm{Var}(\mathbf{y}-\mathbf{r})}{\mathrm{Var}(\hat{\mathbf{y}})}. \quad [9]$$

Here $\mathrm{Var}(\hat{\mathbf{y}})$ is the variance of the unobserved noiseless signal $\hat{\mathbf{y}}$. The explainable signal variance $\mathrm{Var}(\hat{\mathbf{y}})$ is estimated from the pattern of disagreement between responses in different presentations of the same stimulus (equation 1 in ref. 29).

To determine whether the relative level of responses to different classes of optic flow (translation, spirals, deformation) was correctly accounted for by the different models, we also computed a *stimulus class* $\bar{R}^2$ that introduced a free gain per optic flow type. In the case of the hierarchical model, we found that the relative level of responses across stimulus types was misestimated for 70% of cells (25/36, $P < 0.001$, likelihood-ratio test), and in a majority of these cases (84%, 21/25) predicted responses were too weak for spiral stimuli relative to translation stimuli. We emphasize that the stimulus class metric is not an accurate reflection of the quality of the model predictions, but rather is an artifice that allowed us to isolate one mechanism underlying quality of fit.

Decoding Simulations. We compared the capacity of an optimal linear estimator to extract information relevant to behavior. From the 61 fits (1 per cell) under the hierarchical models, we generated 61 × 4 = 244 virtual cells through reflections across the x and y axes to compensate for inhomogeneous sampling of visual space. Because the cells were tested at different resolutions and at different screen positions, we scaled and repositioned the receptive fields to span the central 120° × 120° of the visual field. Stimuli were cropped to the central 90° × 90° of the visual field to avoid artifacts around receptive field edges.

An object 1/16th the size of the visual field was simulated as undergoing 3D motion in 1 of 17 directions (left, up, down, right, toward the observer, and intermediate directions; Fig. 6*B*). The object could be located in 1 of 25 positions lying inside the receptive field. The speed of the object was chosen on a log scale from 2 to 16 Hz; the physical speed of the object may be reconstructed in meters per second or degrees per second if the distance to the object is known.

We reconstructed the physical parameters of the stimulus, using an optimal linear estimator given the outputs of a population of MST cells (47). We picked 122 cells at random from the pool of 244 to yield a decoding population of a size comparable to that previously used in the literature (47). The variables to reconstruct were the signed log velocities in each direction, for example $\mathrm{sign}(\nu_x)\log(|\nu_x| + 1)$ for the velocity in the x direction. To do so we computed the weights $\mathbf{w}$ that minimized the squared error between the reconstruction $\mathbf{Xw}$ and the variable to decode $\mathbf{y}$. Here $\mathbf{X}$ is a matrix with one row for each stimulus and 123 columns (1 for each cell and an offset). The quality of the reconstruction was determined by the root mean square (RMS) error and was expressed as a percentage of the range of log velocity in the x direction (5.67 log Hz). Each decoding simulation was repeated for 50 different random choices of decoding population to yield a mean value and SD.

*Khawaja F, Butts D, Pack C (2007) Towards the characterization of single cell MT and MST neuronal function in the context of natural vision. *Soc Neurosci Abs*, 715.715/FF718.

ACKNOWLEDGMENTS. We thank Drs. Curtis Baker and Maurice Chacron for comments on an early version of the manuscript and Julie Coursol and Cathy

Hunt for technical assistance. This work was supported by Canadian Institutes of Health Research Grant MOP-115178 and a Le Ministère du Développement Économique, de l'Innovation et de l'Exportation du Québec grant (to C.C.P.). C.C.P. and D.A.B. were supported by Collaborative Research in Computational Neuroscience Grant IIS-0904430 from the National Science Foundation. F.A.K. was supported by Fonds de Recherche Santé Québec Fellowship 13159. P.J.M. was supported by Fonds de recherche du Québec-Nature et technologies Scholarship 149928.

1. Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154.
2. DeAngelis GC, Ohzawa I, Freeman RD (1993) Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *J Neurophysiol* 69:1118–1135.
3. Movshon JA, Thompson ID, Tolhurst DJ (1978) Receptive field organization of complex cells in the cat's striate cortex. *J Physiol* 283:79–99.
4. Pasupathy A, Connor CE (2001) Shape representation in area V4: Position-specific tuning for boundary conformation. *J Neurophysiol* 86:2505–2519.
5. Brincat SL, Connor CE (2004) Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci* 7:880–886.
6. Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12:1187–1196.
7. Yu CP, Page WK, Gaborski R, Duffy CJ (2010) Receptive field dynamics underlying MST neuronal optic flow selectivity. *J Neurophysiol* 103:2794–2807.
8. Tanaka K, et al. (1986) Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *J Neurosci* 6:134–144.
9. Duffy CJ, Wurtz RH (1991) Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *J Neurophysiol* 65:1346–1359.
10. Graziano MS, Andersen RA, Snowden RJ (1994) Tuning of MST neurons to spiral motions. *J Neurosci* 14:54–67.
11. Orban GA, et al. (1992) First-order analysis of optical flow in monkey brain. *Proc Natl Acad Sci USA* 89:2595–2599.
12. Raiguel S, et al. (1997) Size and shape of receptive fields in the medial superior temporal area (MST) of the macaque. *Neuroreport* 8:2803–2808.
13. Lappe M (2000) Computational mechanisms for optic flow analysis in primate cortex. *Int Rev Neurobiol* 44:235–268.
14. Perrone JA, Stone LS (1994) A model of self-motion estimation within primate extrastriate visual cortex. *Vision Res* 34:2917–2938.
15. Perrone JA, Stone LS (1998) Emulating the visual receptive-field properties of MST neurons with a template model of heading estimation. *J Neurosci* 18:5958–5975.
16. Tanaka K, Fukada Y, Saito HA (1989) Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *J Neurophysiol* 62:642–656.
17. Maunsell JH, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol* 49:1127–1147.
18. Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9:1421–1431.
19. Albright TD (1984) Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol* 52:1106–1130.
20. 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.
21. 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.
22. Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15:243–262.
23. Wu MCK, David SV, Gallant JL (2006) Complete functional characterization of sensory neurons by system identification. *Annu Rev Neurosci* 29:477–505.
24. Carandini M, et al. (2005) Do we know what the early visual system does? *J Neurosci* 25:10577–10597.
25. Zhang K, Sereno MI, Sereno ME (1993) Emergence of position-independent detectors of sense of rotation and dilation with Hebbian learning: An analysis. *Neural Comput* 5:597–612.
26. Poggio T, Verri A, Torre V (1991) *Green Theorems and Qualitative Properties of the Optical Flow*, AI Memo AIM-1289 (MIT Press, Cambridge, MA).
27. Boussaoud D, Ungerleider LG, Desimone R (1990) Pathways for motion analysis: Cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *J Comp Neurol* 296:462–495.
28. Grossberg S, Mingolla E, Pack CC (1999) A neural model of motion processing and visual navigation by cortical area MST. *Cereb Cortex* 9:878–895.
29. Sahani M, Linden JF (2003) How linear are auditory cortical responses? *Advances in Neural Information Processing Systems*, eds Becker S, Thrun S, Obermayer K (MIT Press, Cambridge, MA), pp 109–116.
30. Cadieu C, et al. (2007) A model of V4 shape selectivity and invariance. *J Neurophysiol* 98:1733–1750.
31. David SV, Gallant JL (2005) Predicting neuronal responses during natural vision. *Network* 16:239–260.
32. Mante V, Bonin V, Carandini M (2008) Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. *Neuron* 58:625–638.
33. Willmore BDB, Prenger RJ, Gallant JL (2010) Neural representation of natural images in visual area V2. *J Neurosci* 30:2102–2114.
34. Tsui JMG, Hunter JN, Born RT, Pack CC (2010) The role of V1 surround suppression in MT motion integration. *J Neurophysiol* 103:3123–3138.
35. Anzai A, Peng X, Van Essen DC (2007) Neurons in monkey visual area V2 encode combinations of orientations. *Nat Neurosci* 10:1313–1321.
36. Gallant JL, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259:100–103.
37. Allman J, Miezin F, McGuinness E (1985) Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception* 14:105–126.
38. Raiguel S, Van Hulle MM, Xiao DK, Marcar VL, Orban GA (1995) Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *Eur J Neurosci* 7:2064–2082.
39. Xiao DK, Raiguel S, Marcar V, Koenderink J, Orban GA (1995) Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. *Proc Natl Acad Sci USA* 92: 11303–11306.
40. Xiao DK, Raiguel S, Marcar V, Orban GA (1997) The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cereb Cortex* 7:662–677.
41. Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9: 181–197.
42. Gautama T, Van Hulle MM (2001) Function of center-surround antagonism for motion in visual area MT/V5: A modeling study. *Vision Res* 41:3917–3930.
43. Hatsopoulos N, Gabbiani F, Laurent G (1995) Elementary computation of object approach by wide-field visual neuron. *Science* 270:1000–1003.
44. Gabbiani F, Krapp HG, Koch C, Laurent G (2002) Multiplicative computation in a visual neuron sensitive to looming. *Nature* 420:320–324.
45. Peña JL, Konishi M (2001) Auditory spatial receptive fields created by multiplication. *Science* 292:249–252.
46. Peirce JW (2007) The potential importance of saturating and supersaturating contrast response functions in visual cortex. *J Vis*, 10.1167/7.6.13.
47. Ben Hamed S, Page W, Duffy C, Pouget A (2003) MSTd neuronal basis functions for the population encoding of heading direction. *J Neurophysiol* 90:549–558.
48. DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. *Trends Cogn Sci* 11: 333–341.
49. Takemura A, Inoue Y, Kawano K, Quaia C, Miles FA (2001) Single-unit activity in cortical area MST associated with disparity-vergence eye movements: Evidence for population coding. *J Neurophysiol* 85:2245–2266.
50. Takemura A, Murata Y, Kawano K, Miles FA (2007) Deficits in short-latency tracking eye movements after chemical lesions in monkey cortical areas MT and MST. *J Neurosci* 27:529–541.
51. Sun H, Frost BJ (1998) Computation of different optical variables of looming objects in pigeon nucleus rotundus neurons. *Nat Neurosci* 1:296–303.
52. Movshon JA, Newsome WT (1996) Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J Neurosci* 16:7733–7741.
53. Perrone JA, Krauzlis RJ (2008) Spatial integration by MT pattern neurons: A closer look at pattern-to-component effects and the role of speed tuning. *J Vis* 8(9):11–14.
54. Brincat SL, Connor CE (2006) Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron* 49:17–24.
55. Pack CC, Born RT (2001) Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature* 409:1040–1042.
56. Pack CC, Hunter JN, Born RT (2005) Contrast dependence of suppressive influences in cortical area MT of alert macaque. *J Neurophysiol* 93:1809–1815.
57. Nishimoto S, Gallant JL (2011) A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J Neurosci* 31:14551–14564.
58. Abbott LF, Varela JA, Sen K, Nelson SB (1997) Synaptic depression and cortical gain control. *Science* 275:220–224.
59. Chance FS, Nelson SB, Abbott LF (1998) Synaptic depression and the temporal response characteristics of V1 cells. *J Neurosci* 18:4785–4799.
60. Kouh M, Poggio T (2008) A canonical neural circuit for cortical nonlinear operations. *Neural Comput* 20:1427–1451.
61. Britten KH, Heuer HW (1999) Spatial summation in the receptive fields of MT neurons. *J Neurosci* 19:5074–5084.
62. Britten KH, van Wezel RJA (1998) Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nat Neurosci* 1:59–63.
63. Gu Y, Fetsch CR, Adeyemo B, Deangelis GC, Angelaki DE (2010) Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron* 66:596–609.
64. Zemel RS, Sejnowski TJ (1998) A model for encoding multiple object motions and self-motion in area MST of primate visual cortex. *J Neurosci* 18:531–547.
65. Wood SN (2006) Generalized linear models. *Generalized Additive Models: An Introduction with R* (Chapman & Hall/CRC Press, Boca Raton, FL), pp 59–120.
66. Paolini M, Distler C, Bremmer F, Lappe M, Hoffmann KP (2000) Responses to continuously changing optic flow in area MST. *J Neurophysiol* 84:730–743.
67. Palanca BJA, DeAngelis GC (2003) Macaque middle temporal neurons signal depth in the absence of motion. *J Neurosci* 23:7647–7658.
68. Nover H, Anderson CH, DeAngelis GC (2005) A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J Neurosci* 25:10049–10060.
69. Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: A statistical view of boosting. *Ann Stat* 28:337–374.
70. Anderson JC, Binzegger T, Martin KAC, Rockland KS (1998) The connection from cortical area V1 to V5: A light and electron microscopic study. *J Neurosci* 18:10525–10540.
71. Alonso JM, Usrey WM, Reid RC (2001) Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *J Neurosci* 21:4002–4015.

Supplementary information

List of supplementary figures

Supplementary Figure S1: Our methods can estimate veridical receptive fields.

(A) Receptive fields of simulated neurons (see Supplementary Methods for details). (B) Estimated receptive fields based on hierarchical model and subunit visualization procedure. Our methods are able to estimate veridical receptive fields within the limits imposed by noise. (C) Cross-validation example for cell in Figure S7, page 1. Top: evolution of quality of fit for each fold as a function of number of boosting iterations. The likelihood of the data always increases as more parameters are added into the model. Bottom: evolution of the validated goodness-of-fit. Thin lines: evolution of the quality of the predictions for each leave-aside fold as the number of boosting iterations increases. As more parameters are added, the model starts overfitting to noise, and beyond a certain point the predictions on the leave-aside fold become worse. The optimal number of iterations varies from fold to fold, as more or less noisy data is placed at random in the fit and validation folds. Thick line: the average of the validation scores is used to determine the number of boosting iterations (indicated by a gray arrow) to be used for the final model fit, which uses all data. The number of boosting iterations is not equal to number of degrees of freedom in the model, because of the use of a damping parameter $\alpha < 1$ (see Main Methods).

Supplementary Figure S2: Failure of linear model to account for MST responses.

(A) and (B): Predicted responses to tuning curve stimuli based on linear model for cells depicted in Figure 1B and 1C. The predicted responses to translation are approximately correct, but responses to spirals and deformation are not captured. (C) percentage difference in cross-validated log-likelihood between hierarchical and linear models. (D) $\bar{R}^2$ of tuning curve predictions compared between

hierarchical and linear models. Despite its higher dimensionality, the hierarchical model performs better on validation sets than the linear model.

Supplementary Figure S3: Analysis of relative goodness-of-fit of nonlinear integration model and linear integration model.

(A) Relative cross-validated log-likelihood between nonlinear and linear hierarchical models. Note the sizable number of cells showing improvements of 20% or more after the addition of a single parameter (β). (B) Stimulus class $\bar{R}^2$ of tuning curve predictions compared between linear and nonlinear hierarchical models. The stimulus class $\bar{R}^2$ metric measures fraction of variance accounted for assuming independent gains for different stimulus classes (translation, spirals, deformation). By construction, it is insensitive to how well each model predicts the relative gain of responses between different stimulus classes. The hierarchical model with nonlinear integration retains better predictive ability according to this metric, indicating that the better fits are not entirely due to better accounting of relative gain between stimulus classes.

Figure S4: Gain control model results.

(A) Histogram of optimal parameters for gain control pool. Left: optimal bandwidth was skewed towards small bandwidths (tuned gain control), although some cells preferred untuned gain control pools. Middle: optimal pool size was the same as that of the subunits themselves, ruling out strong center-surround effects. Right: optimal gain control strength was skewed towards strong gain control, which gives more compressive effects, consistent with results in Figure 6. Overall, these results are consistent with a strong, tuned and spatially limited gain control mechanism mathematically equivalent to a static compressive nonlinearity (B) Quality of predictions of gain control model with optimal unconstrained pool (full model) compared with that of gain control model with pool constrained to be tuned and of limited spatial extent (restricted model). The full model does not lead to predictions

appreciably better than the restricted model whose gain control pool has an effect equivalent to a static compressive nonlinearity.

Supplementary Figure S5: Multiplicative interaction model confirms existence of nonlinear integration mechanism.

(A) Cross-validated log-likelihood of linear hierarchical model versus a model with multiplicative pairwise interactions (see Supplementary Methods for definitions). Note that only the most responsive cells were used for these fits because the multiplicative interactions model has a high number of degrees of freedom. (B) Relative $\bar{R}^2$ of tuning curve predictions. A model with explicit multiplicative interactions provides a better description of the data, consistent with a prominent nonlinear integration mechanism.

Supplementary Figure S6 (two pages): Receptive field parameters for sample cells.

Page 1, top left: mean temporal filter and random subset of temporal filters found for the population of MST cells. In most cases, temporal filters are integrative and peak at 100 ms. Page 1, other positions and Page 2: Subunits of 13 sample cells, including rotation, translation, contraction, and deformation tuned cells. To the right of the receptive fields are pictured the two tuning curve stimuli eliciting the greatest response in the cell; the number underneath these diagrams is the measured firing rate in Hertz.

Supplementary Figure S7 (6 pages): Tuning mosaics and predictions for more sample cells

Tuning mosaics of 6 sample cells and predictions of the hierarchical models with linear and nonlinear integration.

Supplementary Figure S8: Analysis of overlap of subunits

(A) Difference in direction selectivity as a function of normalized distance between subunits for single example cell (same cell as Figure 4B). This normalized distance is defined as the distance between pairs

of subunits divided by the sum of their radii. A normalized distance > 1 indicates that the subunits are non-overlapping. Differences in direction selectivity build gradually as a function of normalized distance. (B) and (C) Difference in direction and speed selectivity as a function pairwise normalized distance across all cells. These plots were created repeating the analysis in (A) for all cells and plotting all subunit pairs at once. Blue trend line is the running median of pairwise differences in selectivity (neighboorhood size of 50). (D) and (E) Detailed view of selectivity difference as a function of pairwise normalized distance. Differences in selectivity build up gradually as a function of normalized distance.

List of supplementary movies

Supplementary Movie S1: A movie showing an example of the continuous optic flow stimulus. Note that due to limitations in the screen capture software, some aspects of the stimulus have been changed from the real stimulus (dot size and overall speed).

Supplementary Methods - Stimulus generation

Dots within the aperture of the random-dot patterns were assigned an instantaneous velocity in the horizontal and vertical directions (u,v) depending on their position (x,y) within their aperture (1):

$$(u,v) = v_0(\cos\theta_1, \sin\theta_1) \text{ (for translation)} \quad (1)$$

$$(u,v) = \omega_0\left(x\cos\theta_2 - y\sin\theta_2, x\sin\theta_2 + y\cos\theta_2\right) \text{ (for expansion/rotation)} \quad (2)$$

$$(u,v) = \omega_0\left(x\cos\theta_3 + y\sin\theta_3, x\sin\theta_3 - y\cos\theta_3\right) \text{ (for deformation)} \quad (3)$$

We set v_0 = 40 deg/s and $\omega_0 = 2Hz$, as these values elicited robust responses from most MST neurons (2, 3). The values of $\theta_1, \theta_2, \theta_3$ were sampled at 45 degree intervals, yielding a basic set of 24 stimuli. Stimuli were presented in 500 ms trials in pseudorandom order, with 5 repeats per stimulus.

The continuous optic flow stimulus was generated according to:

$$u(x, y, t) = s_1(t)\cos\theta + \left(s_3(t) + s_5(t)\right) x \cos\theta + \left(-s_3(t) + s_5(t)\right) y \sin\theta \quad (4)$$

$$v(x, y, t) = s_2(t)\sin\theta + \left(s_4(t) + s_6(t)\right) x \sin\theta + \left(s_4(t) - s_6(t)\right) y \cos\theta \quad (5)$$

where s_1 and s_2 correspond to the translation speeds in the x and y directions, s_3 and s_4 to the magnitude of expansion and contraction, and s_5 and s_6 to the components of deformation. These values were determined by low-pass filtering independent streams of Gaussian-distributed values, with a cutoff of 2Hz. The magnitude of each component was scaled so that the distribution of optic flow components (expansion, rotation, and deformation) had a standard deviation of 1 s^{-1}, while that of the translation components was 20 deg./second. The position of the aperture was determined by another pair of low-pass filtered Gaussian variables with a cutoff of 0.05-0.10 Hz and standard deviation of 10 – 15 deg.

Supplementary Methods - Model fitting and validation

Gradient boosting and cross-validation

The form of the hierarchical models, in which subunits are combined additively, makes them amenable to an estimation procedure known as gradient boosting (4), which is a stepwise fitting procedure that introduces an assumption of sparseness (5). Briefly, gradient boosting starts with an empty model (consisting entirely of a constant firing rate), and iteratively adds subunits whose output is most similar

to the current model residual (i.e., the difference between actual and predicted firing rates). Early subunits tend to fit to prominent effects while later tend to fit to noise. The process is deliberately slowed by setting a subunit's gain to a fraction α of its optimal value when it is added. This which makes the procedure less greedy and allows subsequently added subunits to account for subtler features of the data.

In order to limit the number of degrees of freedom in the model, we determine the optimal number of boosting iterations by 5-fold cross-validation (6). Here the data are split into 5 non-overlapping subsets of equal size, and a model is fit to the data in 4 of these subsets and used to predict the data in the leave-aside set. This process is repeated for the 5 different partitions of the data, and the prediction scores are averaged to form the cross-validated goodness-of-fit. The optimal number of iterations is then defined as the one that maximizes this cross-validated goodness-of-fit; an example run is shown in Figure S1C. Importantly, increasing the number of parameters beyond the optimal value decreases the quality of the predictions, as the extra parameters fit to noise in the training set. Thus cross-validated goodness-of-fit measures are largely insensitive to the number of model parameters.

Model fitting algorithm

The main optimization problem in boosting is to find, at each iteration, the parameters of the subunit whose output is most similar to the current model residual; similarity here is measured by the absolute value of the correlation between a unit's output and the current residual. A direct maximization of the correlation with respect to the parameters of MT-like units would have been challenging to perform rapidly (thousands of these optimizations have to be performed for a given model fit). Alternatively, choosing a unit out of a pool of precomputed units entails storing the output of a high-resolution filter bank of MT cells which would have stretched the memory capacity of the desktop computers performing the optimizations. We instead adopted a hybrid approach to find the optimal subunit for the

linear and nonlinear integration hierarchical models, first finding the approximate optimal parameters out of a low-resolution pool of pre-computed subunits (3 speeds, 8 directions of motion, 144 unit centers, one unit size) and refining the parameters through numerical gradient ascent. In all cases, the unit size p_σ was constrained to be ≥ 1.5 grid units (3 or 3.75 degrees). We fit the spatial and temporal structure of each model in an alternating fashion, according to the following algorithm (7):

1. Assume an initial temporal filter
2. Do 3 times:
    a) Boost model for 25 iterations, $\alpha = 0.5$
    b) refine temporal filter (see next section)
3. Boost model for up to 1000 iterations, $\alpha = 0.1$, with 5-fold cross-validation.

Each model took on average about one hour to fit on a recent desktop computer. For the nonlinear MT model, we fit the model using this procedure for $\beta = 0.2$ to $\beta = 1.4$ in steps of 0.2 and defined the optimal exponent as the one which yielded the model with the highest cross-validated likelihood.

As an additional validation of this estimation procedure we performed simulations in which the parameters were optimized with respect to simulated neurons with fixed collections of subunits (see Supplementary Methods and Supplementary Figure 1).

Fitting the temporal filter

Temporal processing is linear and separable with respect to spatial processing in our model. Thus, If the spatial parameters of the model are fixed, refining the temporal filter can be done by fitting a standard generalized linear model (GLM) with one parameter per time lag and an offset (7). We assumed that the temporal filter was smooth through the use of a Gaussian smoothness prior (8). We fit this penalized GLM using standard methods (9), using cross-validation to adjust the strength of the prior. Example resulting time filters are shown in Supplementary Figure S7.

Application of the model to simulated data

To verify the validity of our fitting procedures, we applied them to simulated neuronal responses, for which the subunits and nonlinearities were known. Testing was done on four spatial receptive field profiles (Supplementary Figure 1A), which were endowed with subunits corresponding to translation and expansion selectivity, with and without suppression in the anti-preferred direction. For each simulated neuron the temporal filters were characterized by five time points [.25,1,.25,-.1,0] (from shortest to longest lag). Responses were normalized to have a standard deviation of 10 Hz, then passed through an exponential output nonlinearity, then normalized again to have a mean firing rate of 10 Hz, and finally truncated to have a maximum peak rate of 140Hz. These values were chosen so that the simulated cells were driven relatively weakly compared to the observed distribution of responses in our sample cells. The cells' outputs were then transduced to a firing rate by a Poisson noise generator. We fit the responses of the neurons with the same continuous optic flow stimulus used in the main text with the hierarchical model, and ran our subunit visualization procedure on the fitted model neurons. The results are shown in Supplementary Figure 1B. The model and visualization procedures are able to recover the correct receptive fields within the limits imposed by noise.

Validation metrics

Given a predicted response $\mathbf{r}$ and an observed response $\mathbf{y}$, the quality of the prediction may be assessed using the standard R^2 metric of variance accounted for):

$$R^2 = \frac{\mathrm{Var}(\mathbf{y}) - \mathrm{Var}(\mathbf{y} - \mathbf{r})}{\mathrm{Var}(\mathbf{y})} \tag{6}$$

In practice the value $R^2 = 1$ cannot be attained, as $\mathrm{Var}(\mathbf{y} - \mathbf{r})$ for a perfect prediction is the variance of the noise, which is non-negligible in physiological measurements. To recover a natural scale we thus used a corrected R^2 metric, also known as predictive power (10):

$$\bar{R}^2 = \frac{\mathrm{Var}(\mathbf{y}) - \mathrm{Var}(\mathbf{y} - \mathbf{r})}{\mathrm{Var}(\hat{\mathbf{y}})} \quad (7)$$

Here $\mathrm{Var}(\hat{\mathbf{y}})$ is the variance of the unobserved noiseless signal $\hat{\mathbf{y}}$. The explainable signal variance $\mathrm{Var}(\hat{\mathbf{y}})$ is estimated from the pattern of disagreement between responses in different presentations of the same stimulus (equation 1 in 10).

To determine whether the relative level of responses to different classes of optic flow (translation, spirals, deformation) was correctly accounted for by the different models, we also computed a *stimulus class* $\bar{R}^2$ which introduced a free gain per optic flow type. As the model underlying this second prediction is a superset of the one-gain model, a likelihood ratio test was used to establish the significance of this second set of predictions over the first. The magnitudes of the gains were then compared to determine whether the model under- or overestimated the responses of one stimulus type over another. In the case of the hierarchical model with linear integration, we found that the relative level of responses across stimulus types was misestimated for 70% of cells (25/36, likelihood ratio test, $p<0.001$), and in a majority of these cases (84%, 21/25) predicted responses were too weak for spiral stimuli relative to translation stimuli. In addition, stimulus class predictions were compared for the linear and nonlinear hierarchical models to determine whether the increase in quality of fit was due to better accounting of gain. We emphasize that the stimulus class metric is not an accurate reflection of the quality of the model predictions, but rather is an artifice that allowed us to isolate one mechanism underlying quality of fit.

Visualization of subunits

For most cells the model recovered many subunits that were similar. Displaying all the subunits recovered by the model made it difficult to discern the structure of the receptive fields because many subunits overlap. Thus, we used a second procedure to select a subset of these subunits which could

account for 80% of the likelihood captured by the full model. We refit the weights of the subunits of each cell, imposing a Laplace prior on these weights (11, 12). This yields even sparser models than those achieved with boosting. We then adjusted the strength of the prior for each cell in order to obtain the sparsest model that accounted for 80% of the likelihood relative to the full model. We then plotted the resulting subunits (Figure 4, for example), modulating the opacity and color of the subunits in proportion to the weight of the recovered subunits.

Analysis of subunit overlap

To gain more insight into the degree to which subunits overlap in space and in tuning, we computed the pairwise normalized distance between subunits with positive weights discovered by the visualization procedure (see previous section), and correlated it with the pairwise difference in direction and speed tuning. The normalized distance was defined as the spatial distance between subunits divided by the sum of their radii. A normalized distance > 1 indicates that the subunits are non-overlapping. An example of this analysis is shown in Figure S8A for the direction tuning of the subunits of the cell originally shown in Figure 4B. In this case, we see that differences in direction selectivity build up gradually with normalized distance.

We repeated this analysis for every cell and compiled all the pairwise differences in Figures S8B and S8C. Again, we see that differences build up gradually with normalized distance; this is true for both direction and speed tuning. This is most clearly seen with the blue trend line, which computes the running median pairwise difference in selectivity.

Figure S8D and S8E replots the same data, zooming in on the x axis (normalized distances <1.5). It is clear that strongly overlapping subunits (normalized distance <.1) generally have very similar direction and speed tuning, indicating that they correspond to the same input. On the other hand, median pairwise differences in direction selectivity reach values > 45 degrees around a normalized distance of

.5, where the overlap is substantial. We conclude that MST receptive fields are densely tiled by subunits whose tuning varies rapidly as a function of position.

Supplementary methods - Alternative models

Linear model

The simplest model that we explored performs a linear match between the local velocity of the observed optic flow field and a preferred template. This model has linear speed tuning and cosine direction tuning, and so it is not tuned in the same sense as the hierarchical models explored in the main text. Rather it is most similar to a linear receptive field model in the luminance domain, as is often used to model LGN or V1 simple cells (13, 14). While this model, endowed with an exponential output nonlinearity and Poisson noise, can be fit directly through maximum likelihood methods (9), we used the same boosting methodology we applied to our other models, so that the results could be directly compared. In the boosting formulation, a model cell contains subunits whose activation is proportional to $u(x, y)$ and $v(x, y)$, the horizontal and vertical components of the velocity of the stimulus inside their receptive fields:

$$f(u(x,y), v(x,y), \mathbf{p}) = p_g \sum_{xy} \left(\cos p_\theta \, u(x,y) + \sin p_\theta \, v(x,y)\right) G(x, y, p_x, p_y, p_\sigma) \quad (1)$$

Here p_g denotes the gain of a unit, p_θ denotes its preferred direction of motion, and G denotes a Gaussian as in the main text (equation 11).

The resulting tuning curve predictions for the cells originally shown in Figure 1B and 1C are presented in Supplementary Figure 2A and 2B. The same patterns of approximately correct predictions for tuning to translation and inadequate predictions for complex optic flow were visible in most cells. Supplementary Figure 2C and 2D compare the cross-validated likelihood and prediction $\bar{R}^2$ for the linear and linear

hierarchical models. While the linear model is attractive because of its mathematical tractability, it fails to capture the more complex selectivity seen in MST responses.

Unrestricted nonlinear MT model

In the nonlinear MT model considered in the main text, all subunits shared a single power law nonlinearity for a given MST cell; the shape of this nonlinearity was determined by an exhaustive search. In the unrestricted nonlinear MT model, this constraint was relaxed such that each subunit had its own power-law nonlinearity selected out of a range from 0.2 to 1.4 in steps of 0.2.

Because of the added computational burden of this model, we used only precomputed MT subunits during fitting (3 speeds, 8 directions of motion, 144 unit centers, one unit size, 7 exponents for ~24000 precomputed subunits). Other aspects of the fitting procedure were similar to the models presented in the main text. The resulting quality of fits (Table 1) indicate that the added flexibility does not lead to enhanced fits.

Divisive center-surround model

In this model, the output of a MT subunit is divided by a weighted sum of the output of other units in its neighborhood. Calling $U(p_x, p_y, p_\theta, p_\rho)$ the raw output of an MT subunit with center at (p_x, p_y) and tuned to direction and speed (p_θ, p_ρ), a corresponding surround-suppressed subunit $N(p_x, p_y, p_\theta, p_\rho)$ is given by:

$$N(p_x, p_y, p_\theta, p_\rho) = \frac{U(p_x, p_y, p_\theta, p_\rho)}{1 + \alpha \sum_{\text{all } \Delta} S(\Delta x, \Delta y) T(\Delta\theta, \Delta\rho) U(p_x + \Delta x, p_y + \Delta y, p_\theta + \Delta\theta, p_\rho + \Delta\rho)} \quad (2)$$

Here $S(\Delta x, \Delta y)$ is the spatial weighting function, a 2D Gaussian centered at the origin with a width σ_s, while $T(\Delta\theta, \Delta\rho)$ is the tuning weighting function, given by:

$$T(\Delta\theta, \Delta\rho) = \exp(-(c\Delta\rho^2 + (1 - cos\,\Delta\theta)^2/4)/2\rho_t^2) \quad (3)$$

c was chosen so that the range of $c\Delta\rho^2$ was equal to 1. By varying α, σ_s and σ_t , we obtained several distinct center-surround models ranging in suppression strength (α), size of the spatial integration pool (σ_s), and tuning strength (σ_t).

We preset the integration size of the raw units to $p_\sigma = 1.8$ grid units. We manually picked 5 values for σ_s ranging from a small to a large surround, $\sigma_s = [0.5,1,2,3,4.5]$. We also manually picked σ_t= $[.13, .3, .42, .72, 2]$ corresponding to angular bandwidths (full-width at half max) of roughly [90,145,180,270, ∞] degrees. For each pair of parameters, we wished to find values of α corresponding to "weak" and "strong" suppression. For every parameter pair, we thus computed $N\left(p_x, p_y, p_\theta, p_\rho\right)$ for a typical stimulus sequence for a unit in the center of the screen over a large range of suppression strengths α. For small α, the output of $N\left(p_x, p_y, p_\theta, p_\rho\right)$ was highly correlated ($r \approx 1$) with the raw unit $U\left(p_x, p_y, p_\theta, p_\rho\right)$, while for large α this correlation reached an asymptote $r_{\min}$. We chose α values corresponding to $r = r_{\min} + [.25, .5, .7, .9, .95] \cdot (1 - r_{\min})$. We thus obtained 125 triplets of parameters (σ_s, σ_t, α) in addition to a reference triplet corresponding to $\alpha = 0$.

Because of the large number of model fits (126 per cell) involved, we used only precomputed MT subunits while fitting the divisive center-surround models (3 speeds, 8 directions of motion, 144 unit centers, one unit size). By examining the cross-validated likelihood of the fits to the continuous optic flow stimulus, we determined the optimal parameters of the surround for each cell (Figure S4A).

Symmetric and asymmetric subtractive surround models

In this model, we considered the possibility that the tuned, asymmetric surrounds of MT cells could contribute significantly to the optic flow selectivity of MST neurons. Rather than a single stereotyped Gaussian envelope, MT cells came in different varieties: no surround (as before), asymmetric one-sided surround, and bilaterally symmetric surround (15, 16). One-sided surrounds were created by the

difference of two Gaussians: the centre was a positive symmetric Gaussian, while the surround was created by a spatially offset, larger Gaussian with negative weight. The output of the center and surround were combined before the half-rectifying nonlinearity.

The surround was offset from the center by a distance 1.5 times the radius at half-height of the center Gaussian (15). The radius of the surround was 1.5 times the radius of the center (15). The surround could be located at 0, 90, 180 or 270 degrees with respect to the preferred direction of the MT cell. Bilateral surrounds were created similarly by the difference of a center Gaussian and two lateral Gaussians.

We considerd 3 different surround strengths (50%, 100%, 150%). At 100% surround strength, a full screen homogeneous stimulus yielded a net null response in surround-suppressed MT cells; the weight of the surround corresponding to 100% strength was scaled by .5 or 1.5 to yield the 50% and 150% surround strengths. Models were fit using the same method as the divisive surround model.

While this change improved the quality of fits on the continuous stimulus in a manner comparable to the addition of nonlinear integration, predictions on the tuning curve stimulus set were poorer (Table 1). This latter test is a more stringent test than the first, since it contains stimuli not found in the initial set.

We also considered a model with a subtractive *symmetric* surround. This surround was created by summing the output of 8 Gaussians surrounding the center; the parameters of these Gaussians were as in the asymmetric surround model. MST cells had access to both these surround-suppressed MT cells and cells with no surrounds. Other aspects of the model were identical to the asymmetric surround model. These yielded essentially identical fits to the subtractive *asymmetric* surrounds.

Finally, we considered the possibility that the combination of a subtractive surround and an output nonlinearity could act synergistically to explain MST selectivity. In this case we considered power-law

nonlinearities (exponents of .2, .4, .6, and 1.4) interacting with either symmetric or asymmetric surrounds of 3 different strengths. This more complex model yielded very similar fits to the more parsimonious single-parameter nonlinearity, regardless of whether the subtractive surround was symmetric or asymmetric (Table 1).

In summary we find that the addition of a subtractive surround (whether symmetric or asymmetric) can improve the performance of the model relative to a simple model comprised only of excitatory subunits. However, none of the subtractive surround models consistently outperformed the simple model with a simple output nonlinearity. Our conclusions are therefore that the single-parameter model provides a powerful and parsimonious account of MST selectivity.

Multiplicative interaction model

The results described in the main text suggest that a multiplicative interaction among inputs is important for explaining MST responses. To test this idea explicitly, we fit the 22 least noisy cells in our sample with a model that contained explicit pairwise multiplicative interactions. The functional subunits $f(\rho, \theta, \mathbf{p})$ of the models computed the sum of a pair of MT cells $M_1(\rho, \theta)$ and $M_2(\rho, \theta)$ and their multiplicative interaction:

$$f(\rho, \theta, \mathbf{p}) = p_a M_1(\rho, \theta) + p_b M(\rho, \theta) + p_c \sqrt{M_1(\rho, \theta), \cdot M_2(\rho, \theta)} \quad (4)$$

Here p_a, p_b and p_c are gains. The square root of the product of the subunits is taken to compress the dynamic range of the interaction term.

This multiplicative interaction model was compared against a baseline model similar to the hierarchical model with linear integration used in the main text. We fitted these models through boosting. The parameters of the multiplication model to be fitted in each boosting iteration were the 3 gains as well as the parameters of the MT cells M_1 and M_2. To make the problem tractable, the parameters of MT cells were restricted to discrete values along a grid (5 speeds, 8 directions of motion, 144 unit centers, one

unit size). As an exhaustive search over all interactions was impractical, we used a greedy algorithm to determine the parameters of each subunit:

1. Find the MT filter M_1 whose output is most similar to the current residual.
2. Project out the output of M_1 from the residual to obtain a second residual.
3. Find the MT filter M_2 such that $\sqrt{M_1(\rho,\theta),\cdot M_2(\rho,\theta)}$ is most similar to the second residual computed in step 2.
4. Fit all gains through least squares.

The subunits of the baseline linear model were also restricted to discrete values along a grid to facilitate comparisons.

We compared the cross-validated likelihood of these models on the continuous optic flow stimulus; results are shown in Supplementary Figure 5A. The model with multiplicative interactions performed better than the linear integration in 100% (22/22) of cases. For those cells for which the *continuous optic flow* stimulus spanned the spatial range of the *tuning curve* stimulus (14/22), we evaluated the quality of the model predictions, the results of which are presented in Supplementary Figure 5B. The multiplicative interaction model performed better than the linear integration model in 78% of cells (11/14, $p < .05$, binomial test). These results are consistent with the notion that multiplicative input interactions are an important property of MST cells.

Mathematical appendix

Linear scaling

The need for nonlinear integration in our MST model might seem inconsistent with the results of a recent study (17), which found that MST neurons, like those in MT (18), respond linearly as a function of

the coherence of optic flow stimuli. Here we demonstrate that linear scaling as a function of coherence may be achieved in a model which includes nonlinear interactions. Assume that an MST neuron's output is $f(a,b)$ in response to two MT inputs a and b (this generalizes to an arbitrary number of inputs). a and b are in turn assumed to be linear as a function of the coherence c (18). Without loss of generality, we let the firing rate at zero coherence for all cells be 0. Then $a(c) = a_0 c$, $b(c) = b_0 c$. An MST cell is linear as a function of coherence if and only if the following relationship holds for all values of $a_0, b_0, c \geq 0$:

$$f(ca_0, cb_0) = cf(a_0, b_0) \tag{5}$$

The family of "power law" integration rules (19):

$$f(a,b) = \left(a^\beta + b^\beta\right)^{\frac{1}{\beta}} \tag{6}$$

Satisfy this linearity condition for all β. For example, when $\beta \to \infty$, the integration rule becomes:

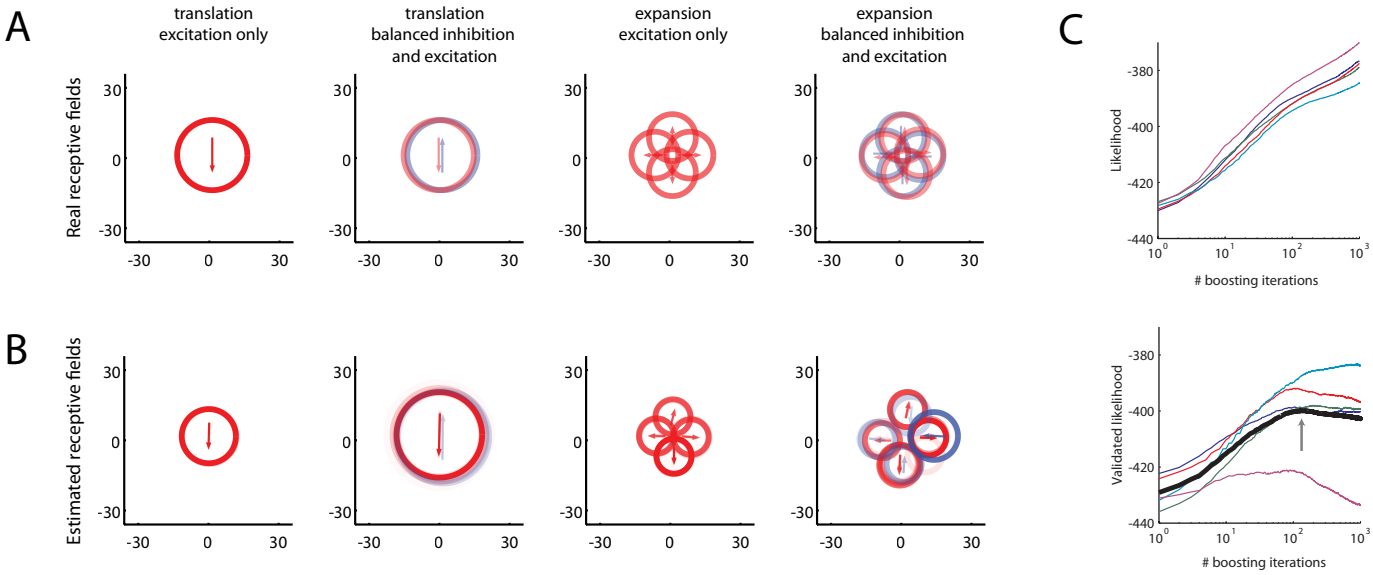
$$f(a,b) = \lim_{\beta\to\infty} \left(a^\beta + b^\beta\right)^{\frac{1}{\beta}} = \max(a,b) \tag{7}$$

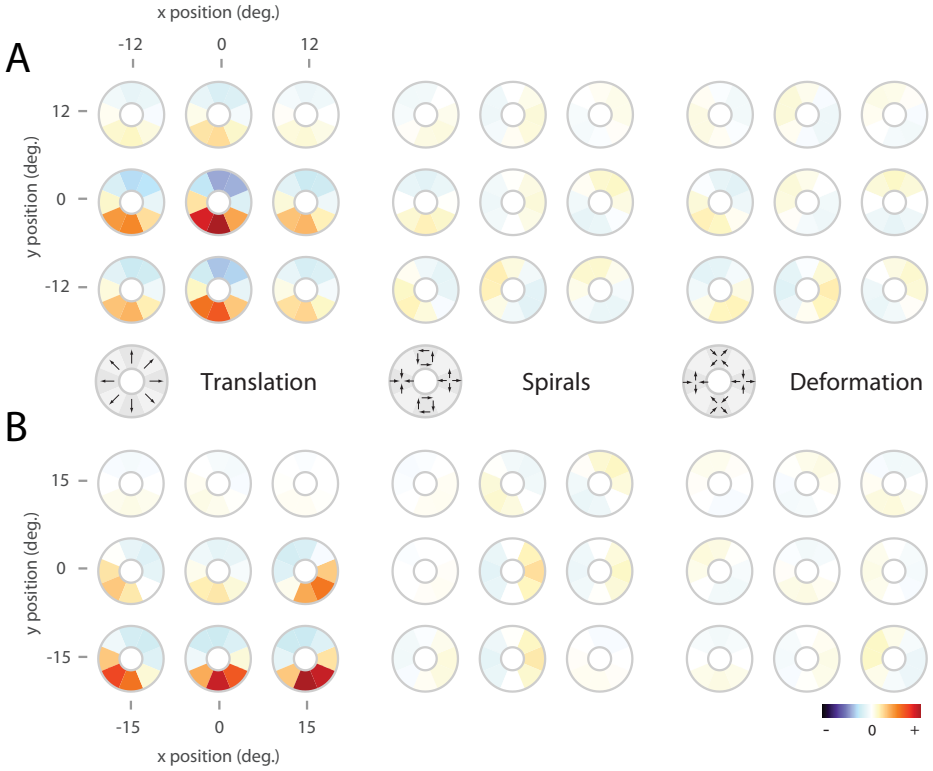
Clearly, $\max(ca, cb) = c \max(a,b)$, and a linear response to coherence is obtained. Hence nonlinear integration is not inconsistent with linear responses to coherence. In our framework, a linear response to coherence could be achieved by replacing the output linearity $\exp(x)$ with $(x^+)^{\frac{1}{\beta}}$; this would considerably complicate the fitting procedure, however.

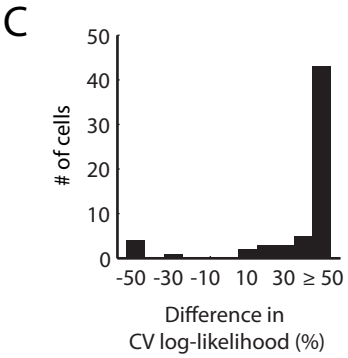
Supplementary references

1. Koenderink JJ (1986) Optic flow. *Vision Res.* 26(1):161-180.
2. 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(6):1329-1345.
3. Tanaka K & Saito H (1989) Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* 62(3):626-641.

4. Buhlmann P & Hothorn T (2007) Boosting algorithms: regularization, prediction and model fitting. *Stat. Sci.* 22(4):477-505.
5. Friedman J, Hastie T, & Tibshirani R (2000) Additive logistic regression: A statistical view of boosting. *Ann. Statist.* 28(2):337-374.
6. Bishop CM (2006) *Pattern Recognition and Machine Learning* (Springer).
7. Ahrens MB, Paninski L, & Sahani M (2008) Inferring input nonlinearities in neural encoding models. *Network Comp. Neural* 19(1):35-67.
8. Wu MCK, David SV, & Gallant JL (2006) Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.* 29(1):477-505.
9. Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. *Network Comp. Neural* 15(4):243-262.
10. Sahani M & Linden JF (2003) How linear are auditory cortical responses? *Advances in neural information processing systems*:125-132.
11. Mineault PJ, Barthelmé S, & Pack CC (2009) Improved classification images with sparse priors in a smooth basis. *J. Vis.* 9:10-17.
12. Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. *J. Roy. Stat. Soc. B. Met.* 58(1):267-288.
13. Carandini M*, et al.* (2005) Do we know what the early visual system does? *J. Neurosci.* 25(46):10577-10597.
14. Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. *Network* 12(2):199-213.
15. Raiguel S, Hulle MM, Xiao DK, Marcar VL, & Orban GA (1995) Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *Eur. J. Neurosci.* 7(10):2064-2082.
16. Xiao DK, Raiguel S, Marcar V, & Orban GA (1997) The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cereb. Cortex* 7(7):662-662.
17. Heuer HW & Britten KH (2007) Linear Responses to Stochastic Motion Signals in Area MST. *J. Neurophysiol.* 98(3):1115-1124.
18. Britten KH, Shadlen MN, Newsome WT, & Movshon JA (1993) Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci.* 10(6):1157-1169.
19. Britten KH & Heuer HW (1999) Spatial Summation in the Receptive Fields of MT Neurons. *J. Neurosci.* 19(12):5074-5084.


A
translation
excitation only
translation
balanced inhibition
and excitation
expansion
excitation only
expansion
balanced inhibition
and excitation
Real receptive fields
B
Estimated receptive fields
30
0
-30
C
Likelihood
-380
-400
-420
-440
boosting iterations
Validated likelihood


A
x position (deg.)
-12
0
12
y position (deg.)
12
0
-12
Translation
Spirals
Deformation
B
15
0
-15
y position (deg.)
-15
0
15
x position (deg.)
-
0
+


C
50
40
30
20
10
0
# of cells
-50 -30 -10 10 30 ≥ 50
Difference in
CV log-likelihood (%)

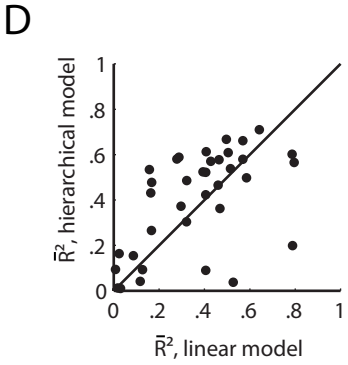
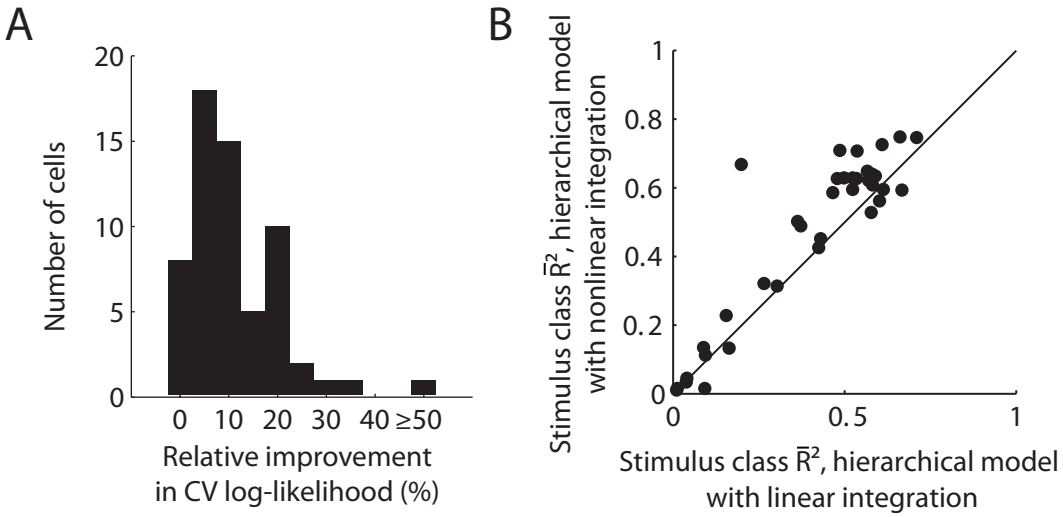
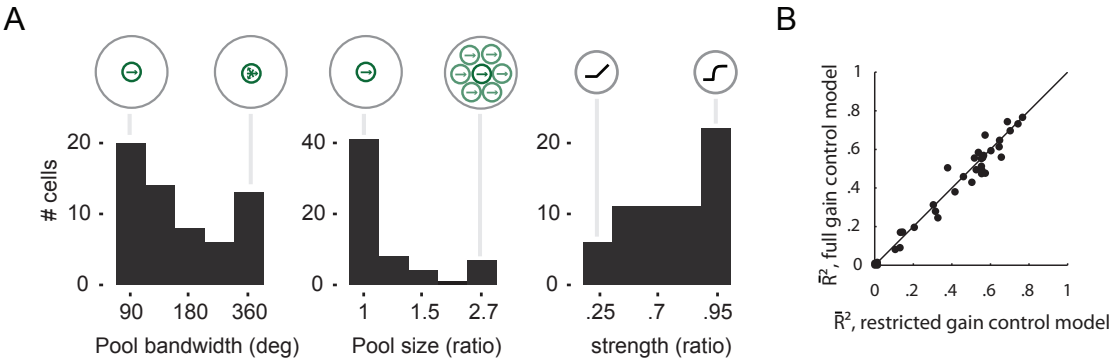
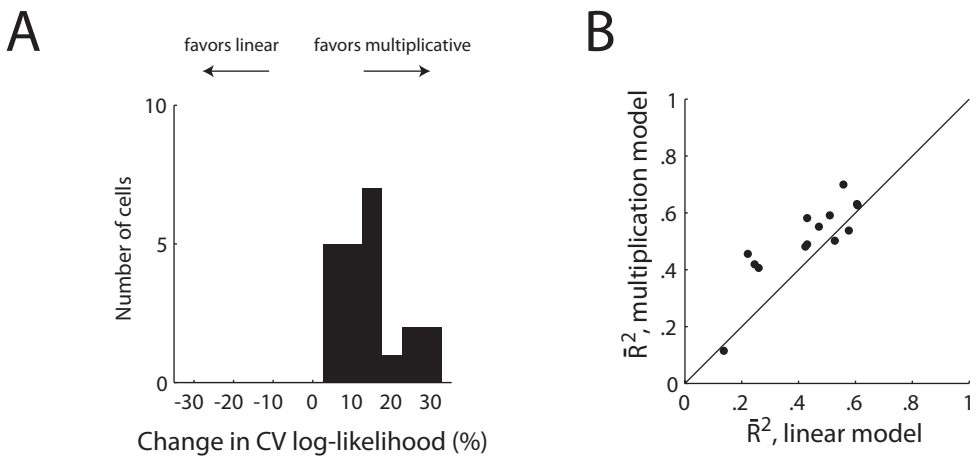
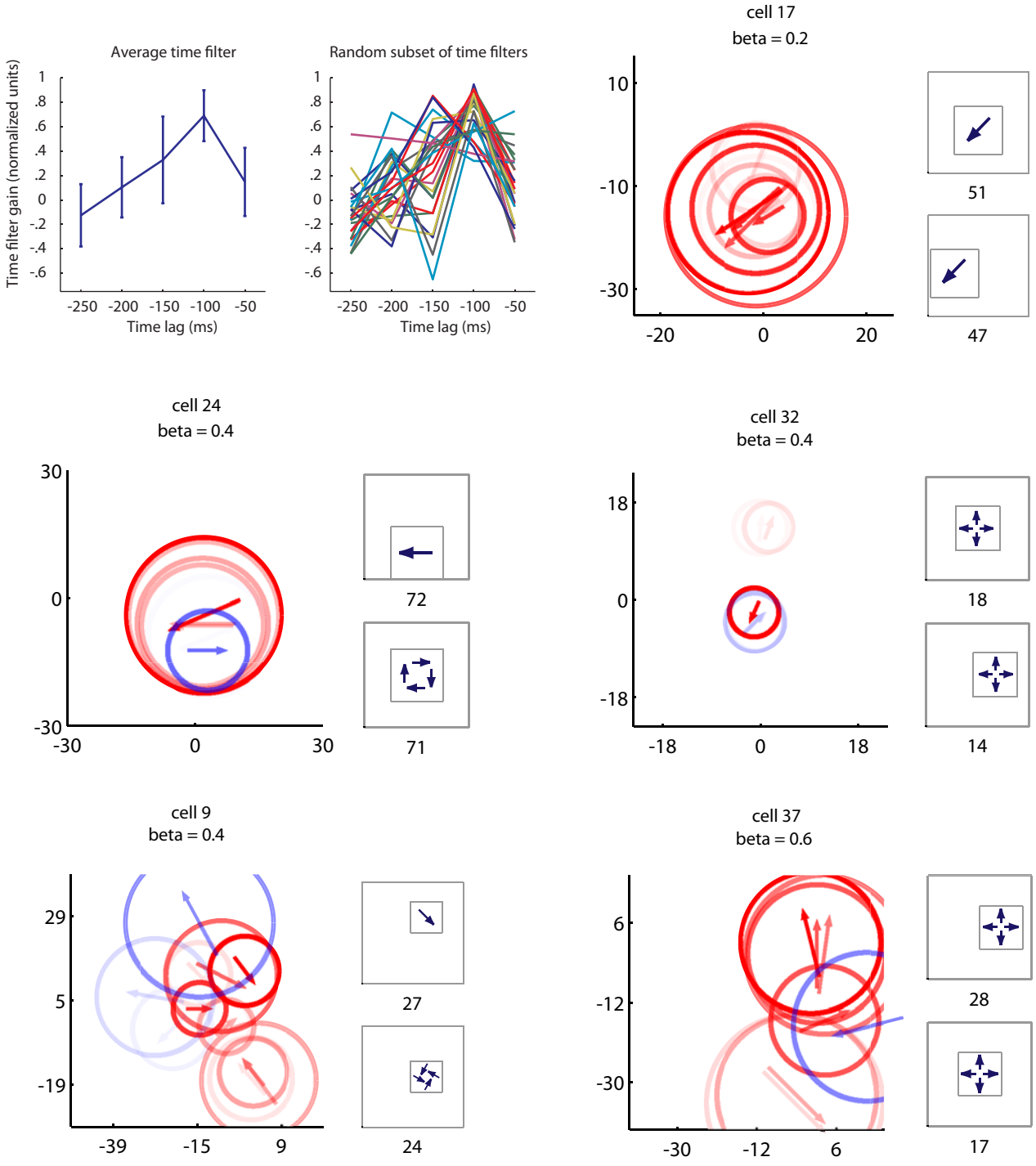

D
1
.8
.6
.4
.2
0
$\bar{R}^2$, hierarchical model
0 .2 .4 .6 .8 1
$\bar{R}^2$, linear model

Figure S3


A
Number of cells
20
15
10
5
0
0 10 20 30 40 ≥50
Relative improvement
in CV log-likelihood (%)
B
Stimulus class $\bar{R}^2$, hierarchical model
with nonlinear integration
1
0.8
0.6
0.4
0.2
0
0 0.5 1
Stimulus class $\bar{R}^2$, hierarchical model
with linear integration


A
cells
20
10
0
90
180
360
Pool bandwidth (deg)
40
20
0
1
1.5
2.7
Pool size (ratio)
20
10
0
.25
.7
.95
strength (ratio)
B
$\bar{R}^2$, full gain control model
1
.8
.6
.4
.2
0
0
.2
.4
.6
.8
1
$\bar{R}^2$, restricted gain control model


A
favors linear
favors multiplicative
10
5
0
Number of cells
-30 -20 -10 0 10 20 30
Change in CV log-likelihood (%)
B
1
.8
.6
.4
.2
0
$\bar{R}^2$, multiplication model
0 .2 .4 .6 .8 1
$\bar{R}^2$, linear model



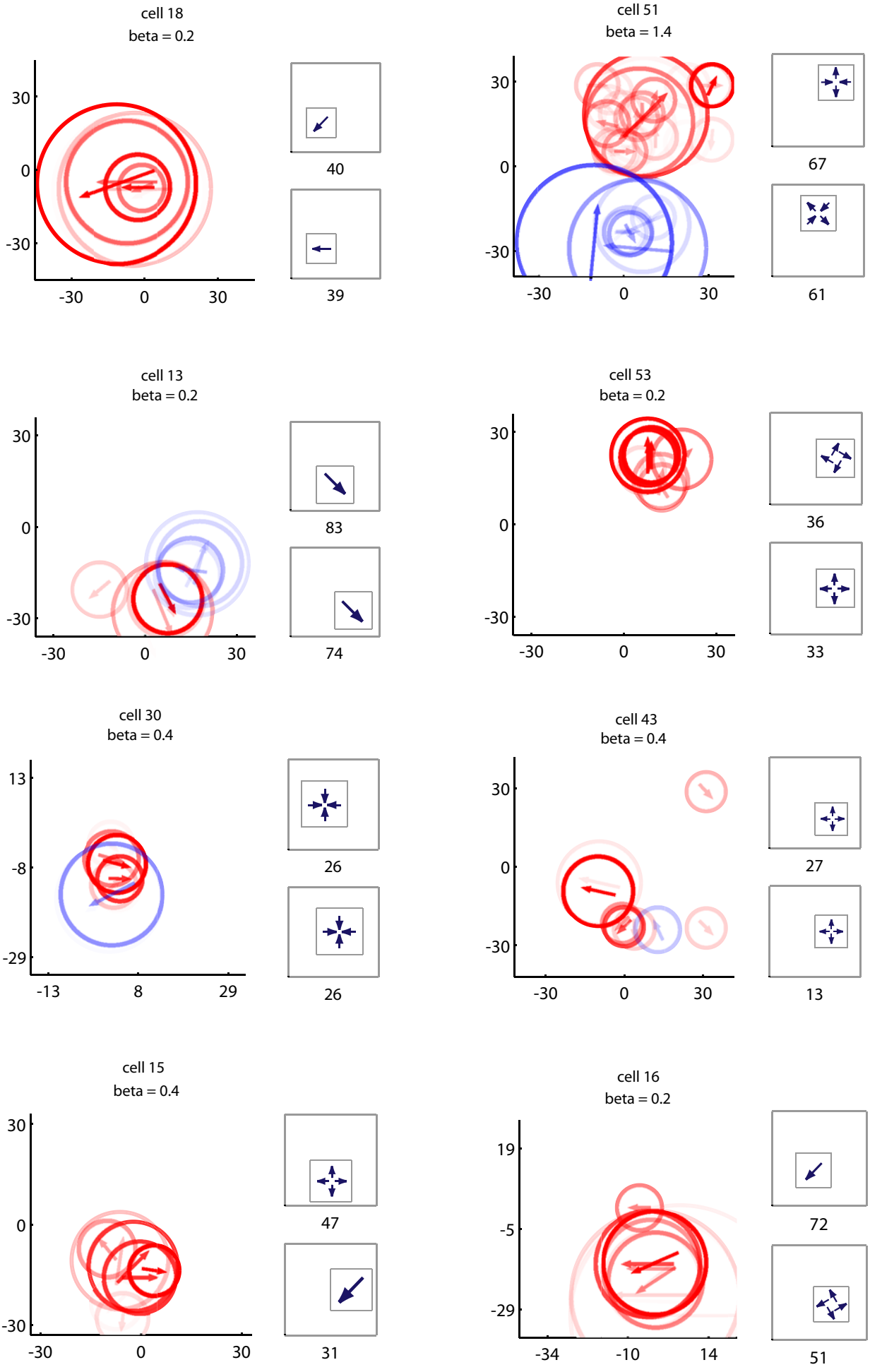
Supplementary Figure S6: Subunits of sample cells

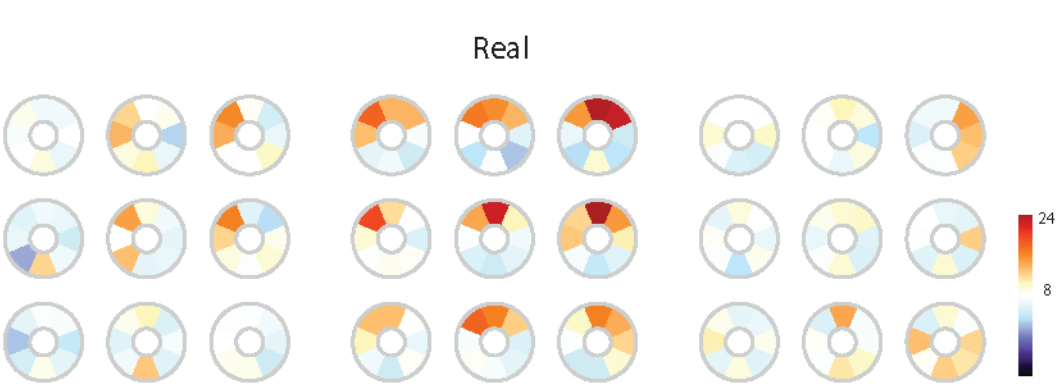
Page 1:

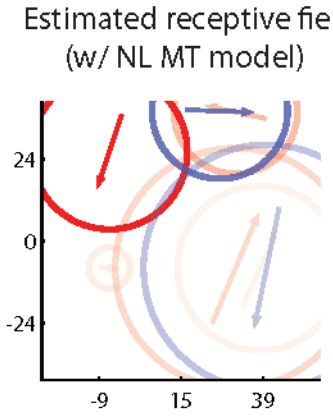
Top left: average and sample timefilters found by the estimation procedure. Other positions: subunits of 5 sample cells. To the right of the receptive fields are pictured the two tuning curve stimuli eliciting the greatest response in the cell; the number underneath these diagrams is the measured firing rate in Hertz.

Page 2:

Subunits of 8 other sample cells

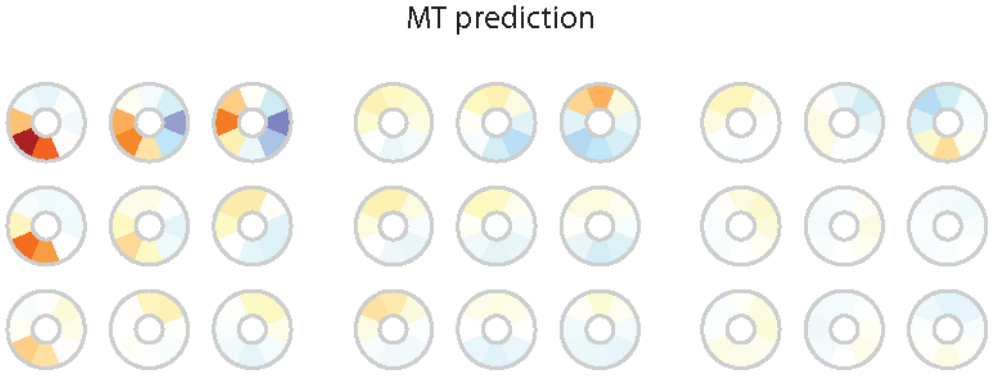

cell 18
beta = 0.2
30
0
-30
-30
0
30
40
39
cell 51
beta = 1.4
30
0
-30
-30
0
30
67
61
cell 13
beta = 0.2
30
0
-30
-30
0
30
83
74
cell 53
beta = 0.2
30
0
-30
-30
0
30
36
33
cell 30
beta = 0.4
13
-8
-29
-13
8
29
26
26
cell 43
beta = 0.4
30
0
-30
-30
0
30
27
13
cell 15
beta = 0.4
30
0
-30
-30
0
30
47
31
cell 16
beta = 0.2
19
-5
-29
-34
-10
14
72
51



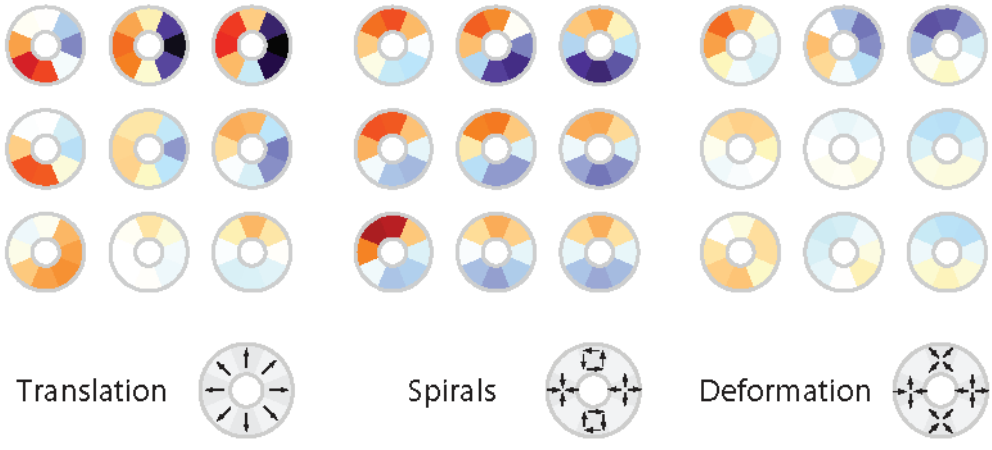


Salient features:

Highly invariantly tuned for counterclockwise rotation



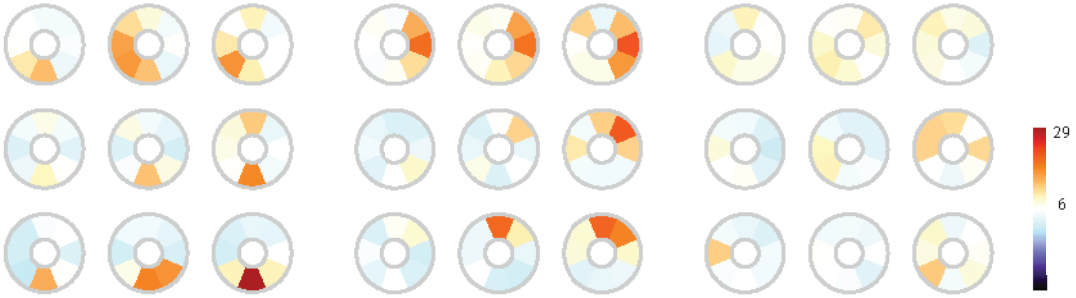
$\bar{R}^2$ =0.18



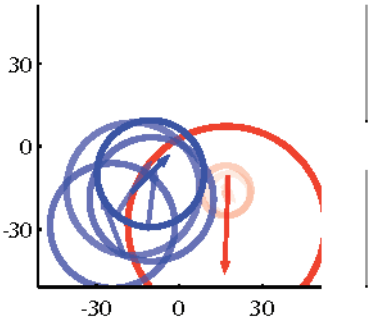
$\bar{R}^2$ =0.34

Figure S7: Tuning mosaics of 6 sample cells and predictions of the hierarchical models with linear and nonlinear integration

Real



Estimated receptive field
(w/ NL MT model)

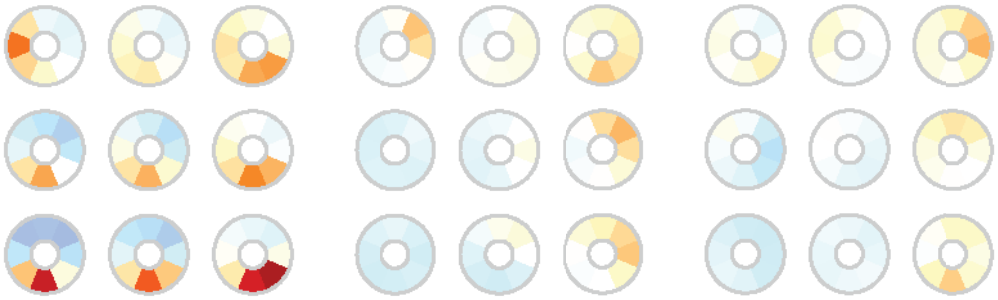


Salient features:

Strong variant tuning to spirals (expansion at top/ccw rotation at bottom)
Responses below median on left-hand side for most stimulus types

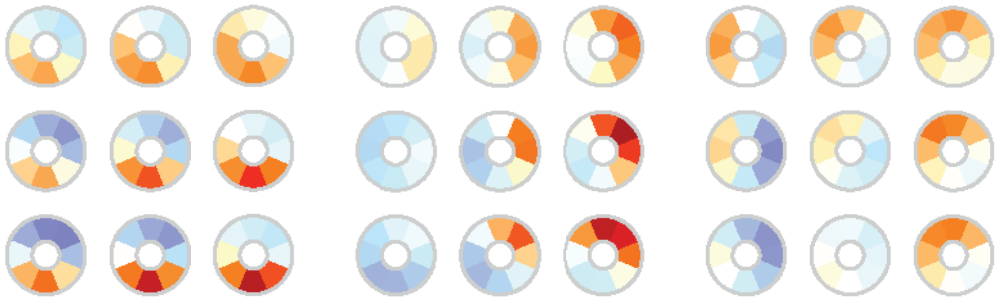
MT prediction

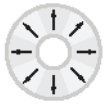
$\bar{R}^2$ =0.44

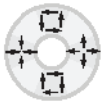


MT + nonlinear integration prediction beta = 0.2

$\bar{R}^2$ =0.58

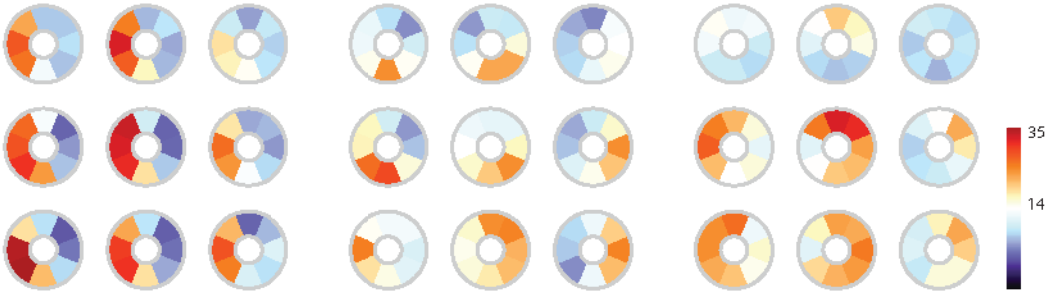


Translation 

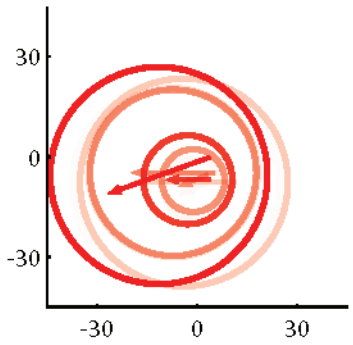
Spirals 

Deformation 

Real



Estimated receptive field
(w/ NL MT model)

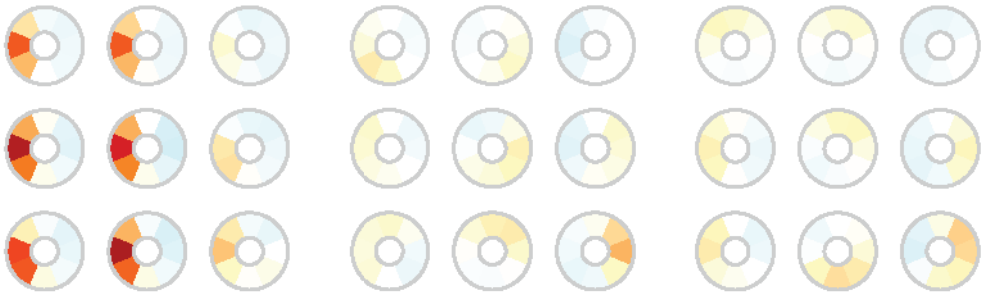


Salient features:

Strong leftward tuning

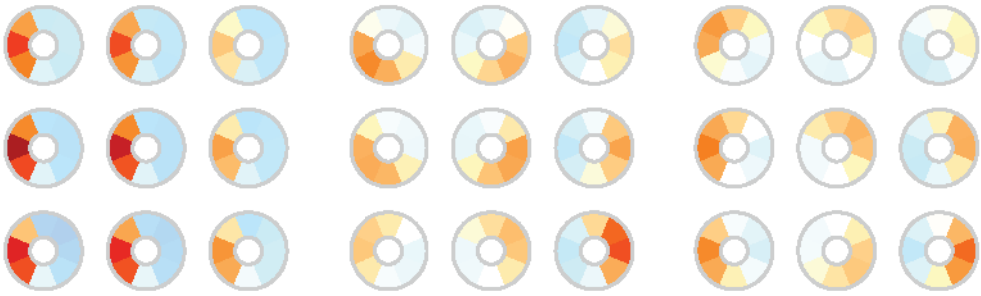
MT prediction

$\bar{R}^2$ =0.61

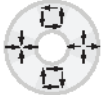


MT + nonlinear integration prediction beta = 0.2

$\bar{R}^2$ =0.74

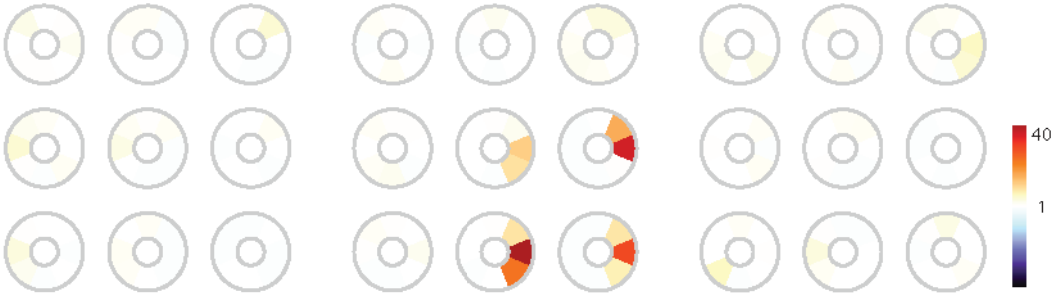


Translation 

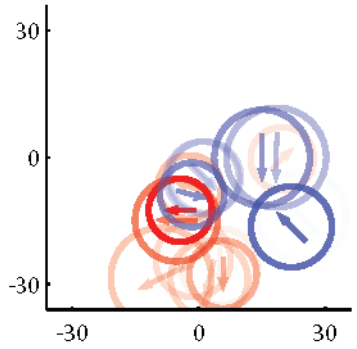
Spirals 

Deformation 

Real



Estimated receptive field
(w/ NL MT model)



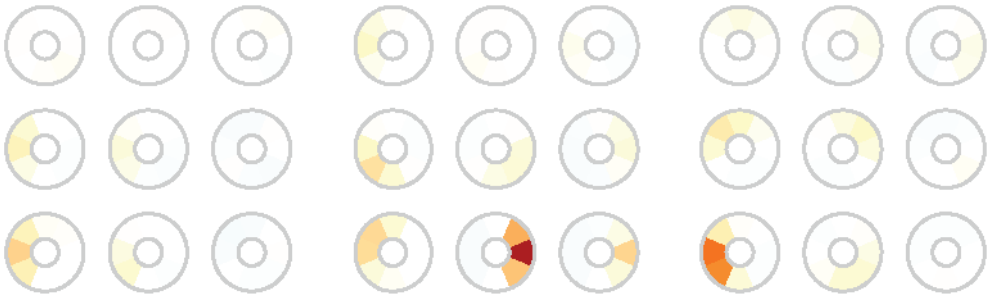
Salient features:

Exquisite sensitivity to expansion at bottom right

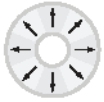
MT prediction

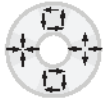
$\bar{R}^2$ =0.21

MT + nonlinear integration prediction beta = 0.4



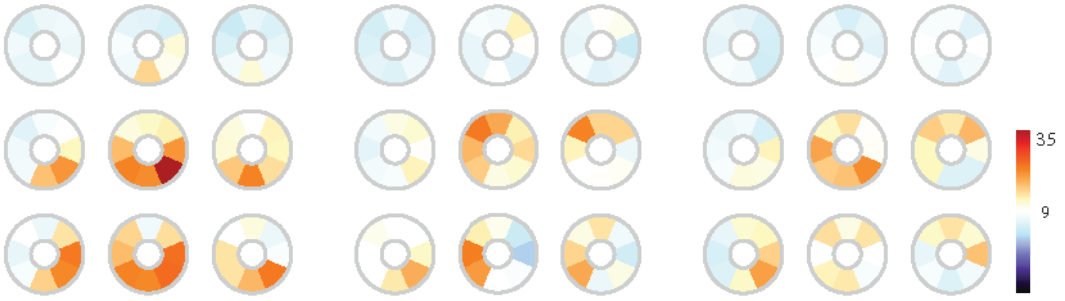
$\bar{R}^2$ =0.36

Translation 

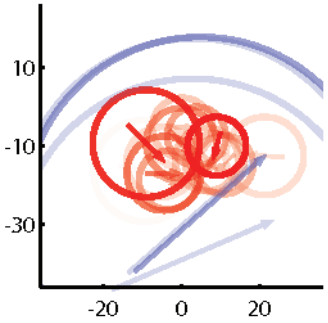
Spirals 

Deformation 

Real



Estimated receptive field
(w/ NL MT model)



Salient features:

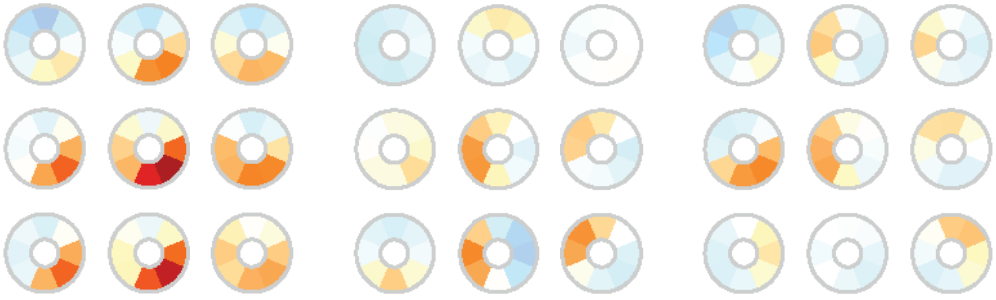
Mostly downwards right selective, almost directionally non-selective in middle and bottom center, some contraction tuning

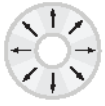
MT prediction

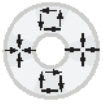
$\bar{R}^2$ =0.55

MT + nonlinear integration prediction beta = 0.6

$\bar{R}^2$ =0.61

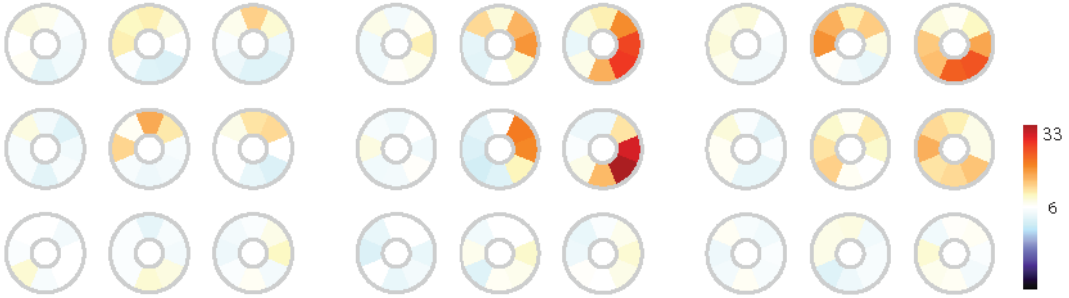


Translation 

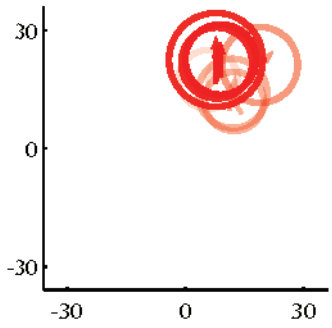
Spirals 

Deformation 

Real



Estimated receptive field
(w/ NL MT model)

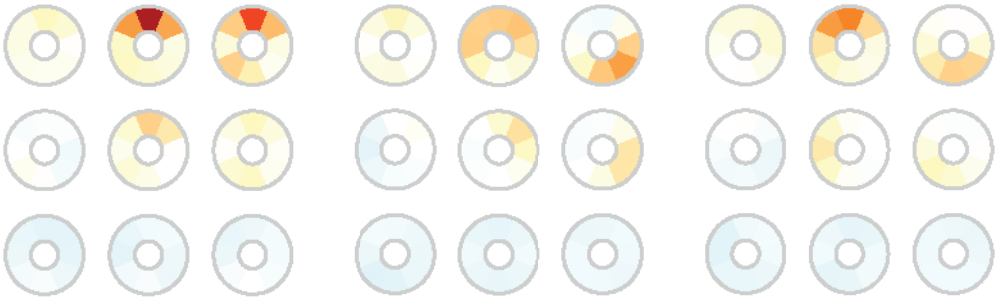


Salient features:

Expansion, spiral and shear tuning

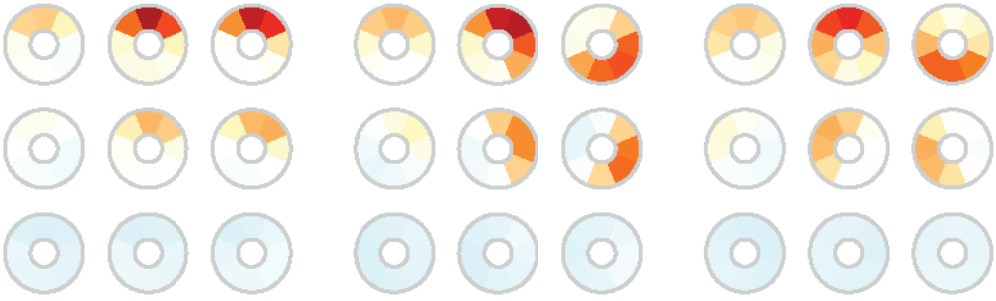
MT prediction

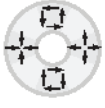
$\bar{R}^2$ =0.27

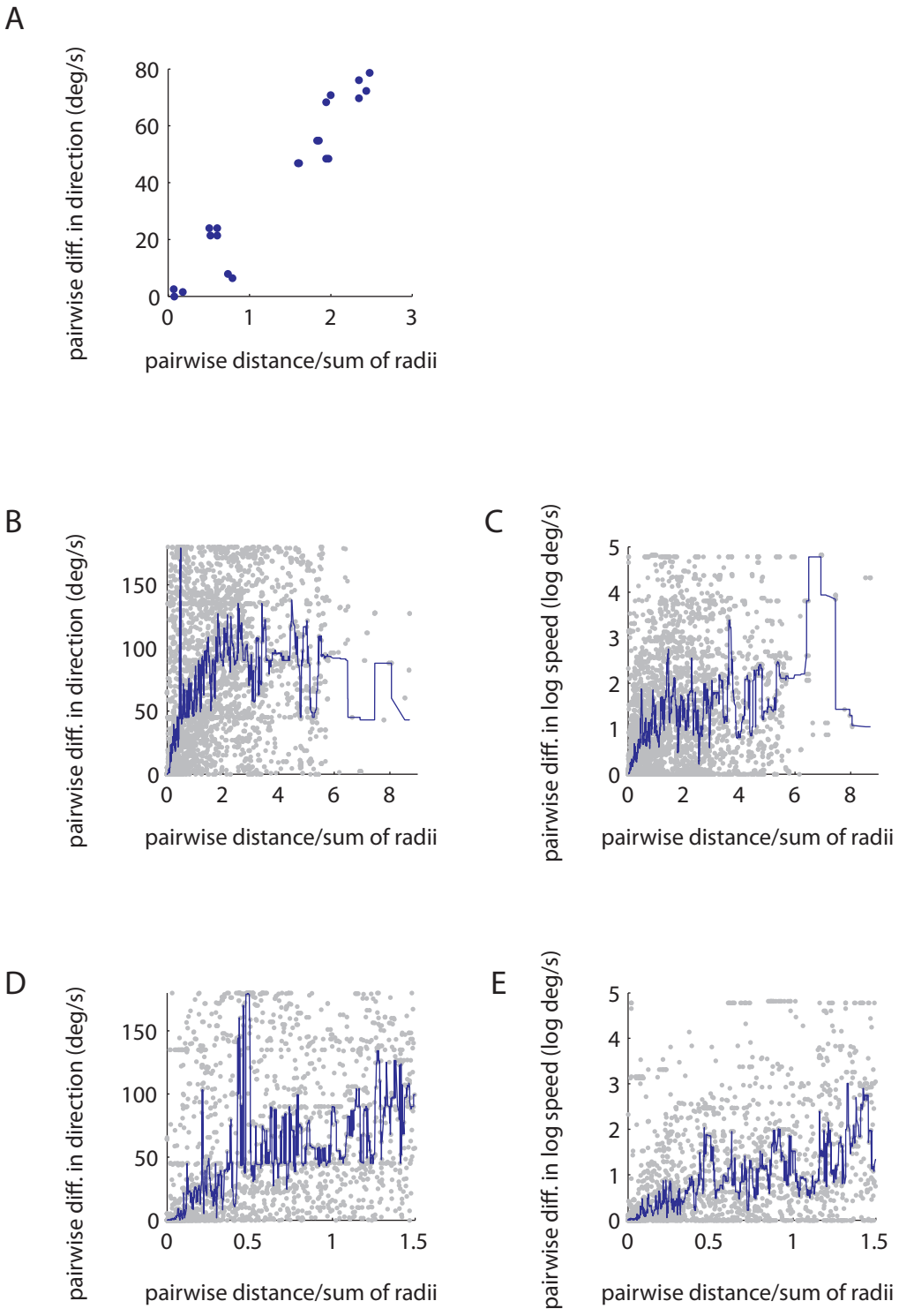


MT + nonlinear integration prediction beta = 0.2

$\bar{R}^2$ =0.50



Translation  Spirals  Deformation 


A
pairwise diff. in direction (deg/s)
80
60
40
20
0
0
1
2
3
pairwise distance/sum of radii
B
pairwise diff. in direction (deg/s)
150
100
50
0
0
2
4
6
8
pairwise distance/sum of radii
C
pairwise diff. in log speed (log deg/s)
5
4
3
2
1
0
0
2
4
6
8
pairwise distance/sum of radii
D
pairwise diff. in direction (deg/s)
150
100
50
0
0
0.5
1
1.5
pairwise distance/sum of radii
E
pairwise diff. in log speed (log deg/s)
5
4
3
2
1
0
0
0.5
1
1.5
pairwise distance/sum of radii