nature portfolio

Peer Review File

Hierarchical and Distinct Biological Motion Processing in Macaque Visual Cortical Areas MT and MST

Corresponding Author: Professor Shenbing Kuang

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Communications Biology.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Review: Hierarchical and distinct motion processing……. Feng et al.

General:
A very interesting and well written paper that focuses -at least partially- on an all-important distinction about how the brain processes self- versus object motion. Most publications in this field of interest have focused on the self-motion aspect when dealing with motion-sensitive neurons in the parietal cortex and elsewhere.

In that context, the authors could discuss (and take suggestions for future experimental designs) a combination of linear motion and optic flow situations, e.g., if BM also involves expansion/contraction with different foci of expansion/contraction (i.e., simulating a real-life situation when a person/object moves towards/away from an observer). After all, most observed and executed movements consist of combinations of translations and rotations. Furthermore, the aspect of self-motion detection versus object motion was only briefly touched (l. 442 ff; see also l. 394 ff). In this broader context, the authors might want to consult the following publications:

Bremmer, F., S. Ben Hamed, J.-R. Duhamel and W. Graf. Heading encoding in the macaque ventral intraparietal area (VIP). Eur. J. Neurosci. 16: 1554-1568, 2002.

Bremmer, F., F. Klam, J.-R. Duhamel, S. Ben Hamed and W. Graf. Visual-vestibular interactive responses in primate ventral intraparietal area (VIP). Eur. J. Neurosci. 16: 1569-1586, 2002.

Bremmer, F., A. Schlack, W. Graf and J.-R. Duhamel. Multisensory self-motion encoding in parietal cortex. Visual Cognition 11: 161-172, 2004.

Fetsch C.R., DeAngelis, G.C., and Angelaki D.E. (2010) Visual-vestibular cue integration for heading perception: applications of optimal cue integration theory. Eur. J. Neurosci. 31: 1721-1729.

Angelaki, D.E., Gu, Y., and DeAngelis, G.C. (2011) Visual and vestibular cue integration for heading perception in extrastriate visual cortex. J. Physiol. (Lond.) 589: 825-833.

Figures:
Better color separation would be helpful especially for the raster and graph displays (Fgs. 2-4). Also Fig. 1a could be improved.

References:
The reference presentation needs to be cleaned up for uniformity: Some references are in upper case (e.g., ref 15 and others), others list as “author et al.” (e.g., reference 30 and others).

The all-important reference to previously analyzed data (#28; l. 114) cannot be accessed to an incomplete citation). Either the authors provide details in the manuscript, or they wait until a proper reference becomes available.

Reviewer #2

(Remarks to the Author)
In the current study, the authors investigated and compared two areas of MT and MST, about whether biological motion (BM) can be encoded in these areas. Two main control conditions were applied for identifying BM coding: scrambled image, or upside-down inversed image. The main finding is that compared to MT, the higher-level MST overall contains more proportion of neurons showing selective response to intact BM image relative to scrambled or inversed images. The paper then constructed a feedforward model to explain the neurophysiological findings. The paper is novel, providing insightful experimental findings about neural correlates of BM perception. The paper was wirtten clearly. Here I have some comments :

Major comments:
1. Can the authors provide evidence about their recording sites, for example, atlas locations, or receptive mapping showing they are in MT and MST respectively.
2. The authors have shown the scrambled and inversed control separately. The bias (selective BM coding) is clearer in scrambled control data, yet effective size is a bit weak in the inversed control data. Is there a correlation between the two dataset? That is, the authors could show a scatter plot of the data in Figure 3D and Figure 4D for the MST data. If there is a significant correlation between the two, the BM-specific coding in MST would be more persuasive.
3. I think BM perception is a relatively high cognitive function that needs to integrate more information. Is the BM-specific coding related with the size of receptive field? What about MT? The distribution in Figure 3-4D is broad for MT, indicating some cells also prefer BM. Are those cells with a large RF?

Reviewer #3

(Remarks to the Author)
This paper explores biological motion tuning in the MT and MST areas of the monkey visual cortex. The findings suggest that neural populations in MST, but not MT, show selectivity for biological motion (BM). The authors also present a feedforward spiking neural network model of MT and MST to investigate the connectivity patterns that drive BM selectivity in the MST population. The question is both interesting and important, and the approach skillfully integrates electrophysiology data with neural network modeling. That said, I believe the paper could benefit from more thorough analysis and some additional work, as I outline below.

Major Comments:

- Since one of the key pieces of evidence for MST neurons being sensitive to biological motion is the comparison between normal and inverted BM, it would be helpful if the authors confirm that the recorded population isn't biased in terms of motion direction tuning.

- The modeling section could be clearer with more details on the model structure and training. For example, the authors mention that the MST neurons respond to three types of biological motion (BM), but what exactly are these types? Also, what loss function was used during training? What behavioral output does the model produce? What learning rule was applied to adjust the weights, and what was the format of the input to MT neurons? These additional details would enhance the clarity of the model description.

- The accuracy reported in Figure 5 could be better explained—it’s not clear what task this accuracy is referring to. Additionally, it would be helpful to include the r-squared value to show how well the model neurons capture the activity of real neurons.

- The connectivity analysis in Figure 6 could use stronger support. In many of the plots, the differences between conditions seem to hinge on just a few data points. Including ablation experiments would make the findings more compelling. For example, if the authors argue that a specific type of MT neuron plays a critical role in a certain type of MST neuron, then lesioning this MT type should result in a corresponding loss of MST selectivity, as predicted by their theory. Could this be incorporated into the analysis?

Minor Comments:

- The tuning plots (e.g., Fig 2b-c) require error bars for better clarity of the statistical significance.
- It would be great if the authors could improve figure resolution. It would make the visuals much more clear.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The authors have addressed my comments well. I do not have other comments. Support for publication.

Reviewer #3

(Remarks to the Author)
The authors have thoroughly addressed all the comments I raised previously. They have incorporated several suggested analyses that, I believe, strengthen their findings, particularly the additional ablation studies on the neural network model. Based on these improvements, I recommend this manuscript for publication.

Point-by-point reply to reviewers' comments

We thank all reviewers for their time and efforts in reviewing our manuscript. In the following, we provided a point-by-point reply to address each of the reviewers' comments. To ease the read, we have color-coded the texts so that the original comments appear in black, our replies in blue, and quotations from the manuscript in blue italics.

Reviewer #1 (Remarks to the Author):

General:
A very interesting and well written paper that focuses -at least partially- on an all-important distinction about how the brain processes self- versus object motion. Most publications in this field of interest have focused on the self-motion aspect when dealing with motion-sensitive neurons in the parietal cortex and elsewhere.

We thank the reviewer for the encouraging and constructive feedback. We also appreciate the reviewer's perspective, which inspired us to discuss the relevance of our study in a broader context in the revision (see below).

In that context, the authors could discuss (and take suggestions for future experimental designs) a combination of linear motion and optic flow situations, e.g., if BM also involves expansion/contraction with different foci of expansion/contraction (i.e., simulating a real-life situation when a person/object moves towards/away from an observer). After all, most observed and executed movements consist of combinations of translations and rotations. Furthermore, the aspect of self-motion detection versus object motion was only briefly touched (l. 442 ff; see also l. 394 ff). In this broader context, the authors might want to consult the following publications:

Bremmer, F., S. Ben Hamed, J.-R. Duhamel and W. Graf. Heading encoding in the macaque ventral intraparietal area (VIP). Eur. J. Neurosci. 16: 1554-1568, 2002.

Bremmer, F., F. Klam, J.-R. Duhamel, S. Ben Hamed and W. Graf. Visual-vestibular interactive responses in primate ventral intraparietal area (VIP). Eur. J. Neurosci. 16: 1569-1586, 2002.

Bremmer, F., A. Schlack, W. Graf and J.-R. Duhamel. Multisensory self-motion encoding in parietal cortex. Visual Cognition 11: 161-172, 2004.

Fetsch C.R., DeAngelis, G.C., and Angelaki D.E. (2010) Visual-vestibular cue integration for heading perception: applications of optimal cue integration theory. Eur. J. Neurosci. 31: 1721-1729.

Angelaki, D.E., Gu, Y., and DeAngelis, G.C. (2011) Visual and vestibular cue integration for heading perception in extrastriate visual cortex. J. Physiol. (Lond.) 589: 825-833.

We thank the reviewer for the valuable suggestion and for pointing out these highly relevant papers. In the revised manuscript, we have substantially expanded the Discussion section to highlight the relevance of our study in this broader context of self-motion versus object motion distinction, citing these and several other related studies in the field (Page 23, Line 535-562).

"[…] Previous neurophysiological studies in dorsal stream extrastriate cortex have predominantly centered on optic flow and the perception of self-motion (or heading) (50, 51). Perceiving the direction of self-motion in the environment relies on visual optic flow and nonvisual cues such as vestibular input during navigation. It has been shown that neurons in MST and ventral intraparietal area (VIP) are multimodal, tuned to both visual optic flow and vestibular cues (43, 44, 52, 53), and integrate multisensory information in a statistically optimal manner to achieve accurate self-motion perception (54). However, motions in real-world scenarios are far more complex than self-motion. The ability to perceive the motion of other objects (or living animals) in the environment is also crucial for navigation and survival. So far, only a handful of studies in the field have focused on the perception of object motion, in particular the biological movement of other creatures (25, 41). In this regard, the current study marks an important endeavor in this direction. Our findings highlight new distinct functional roles of area MT and MST in biological movement perception and social cognition, beyond conventional visual optic flow analysis.

Moreover, since both self-motion and object motion can cause image motion on the retina, resolving the ambiguity between self- and object motions has become an important task for the brain to achieve perceptual stability. To date, a few studies have characterized neural responses to visual stimuli consisted of a combination of self-motion and object motion (55, 56) and examined neural computations underlying the dissociation of self-motion and object motion in MT and MST (57, 58). In this context, the current study investigated the neural responses to object motion (BM stimuli) and self-motion (optic flow stimuli) in isolation. But in real-life situations they often occur simultaneously. This complexity calls for a more comprehensive exploration that combines these different aspects of motion in future experimental designs, e.g., a point-light display portraying walking but at the same time involves expansion/contraction simulating movement towards/away from an observer. […]"

Figures:
Better color separation would be helpful especially for the raster and graph displays (Fgs. 2-4). Also Fig. 1a could be improved.

Done. We have redrawn Fig. 2-4 subplots with more distinctive color separations and thicker lines to enhance the visibility of these graphs. We have also improved the presentation of Fig. 1a.

References:
The reference presentation needs to be cleaned up for uniformity: Some references are in upper case (e.g., ref 15 and others), others list as “author et al.” (e.g., reference 30 and others).

Done. We have corrected these formatting inconsistencies for the full reference list.

The all-important reference to previously analyzed data (#28; l. 114) cannot be accessed to an incomplete citation). Either the authors provide details in the manuscript, or they wait until a proper reference becomes available.

Done. We now provided the detailed DOI information for reference #28 in the bibliography. The preprint has been published in bioRxiv, and currently it is freely accessible at the website (https://doi.org/10.1101/2023.10.30.564682).

Reviewer #2 (Remarks to the Author):

In the current study, the authors investigated and compared two areas of MT and MST, about whether biological motion (BM) can be encoded in these areas. Two main control conditions were applied for identifying BM coding: scrambled image, or upside-down inversed image. The main finding is that compared to MT, the higher-level MST overall contains more proportion of neurons showing selective response to intact BM image relative to scrambled or inversed images. The paper then constructed a feedforward model to explain the neurophysiological findings. The paper is novel, providing insightful experimental findings about neural correlates of BM perception. The paper was written clearly. Here I have some comments:

We thank the reviewer for the overall positive assessment of our manuscript. We also appreciate the concerns and valuable suggestions made by the reviewer, which are very helpful in improving our manuscript (see below).

Major comments:
1. Can the authors provide evidence about their recording sites, for example, atlas locations, or receptive mapping showing they are in MT and MST respectively.

Area MT and MST were localized via a combination of structural magnetic resonance imaging (MRI) scans, stereotaxic coordinates, and functional mapping of their physiological responses to optic flow and translational motion stimuli.

Specifically, we did two MRI scans for each monkey, one pre-surgery to guide chamber implantation and one post-surgery to guide the coordinates of electrode penetration in the grid. The post-surgery scan used a modified grid with dye markers, and under Brainsight Navigator Software the markers extended into the sagittal plane in stereotaxic coordinates. We targeted mainly the dorsal subdivision of area MST located more posteriorly and medially in the upper bank of superior temporal sulcus, and area MT located in the lower bank of superior temporal sulcus.

Furthermore, we did two functional mappings for each recorded MT and MST neuron before measuring their responses to BM stimuli. One was the RF mapping in a relatively coarse manner; we mapped the visual field in a 9x7 space lattice with five repetitions at each spatial location, and this helped us to quickly decide where to place the motion stimuli on the screen in subsequent experiments (see below our reply to your third question). The other was the more quantitative mapping of each neuron's responses to optic flow patterns (expansion/contraction/CW/CCW rotations) and eight translational motions. As in many previous studies, we also confirmed in our data that MST neurons were more responsive and more selective to optic flow patterns than MT neurons.

In the revised manuscript, we now provided these further details in the Supplement for interested readers, without overloading the primary narrative of the main text.

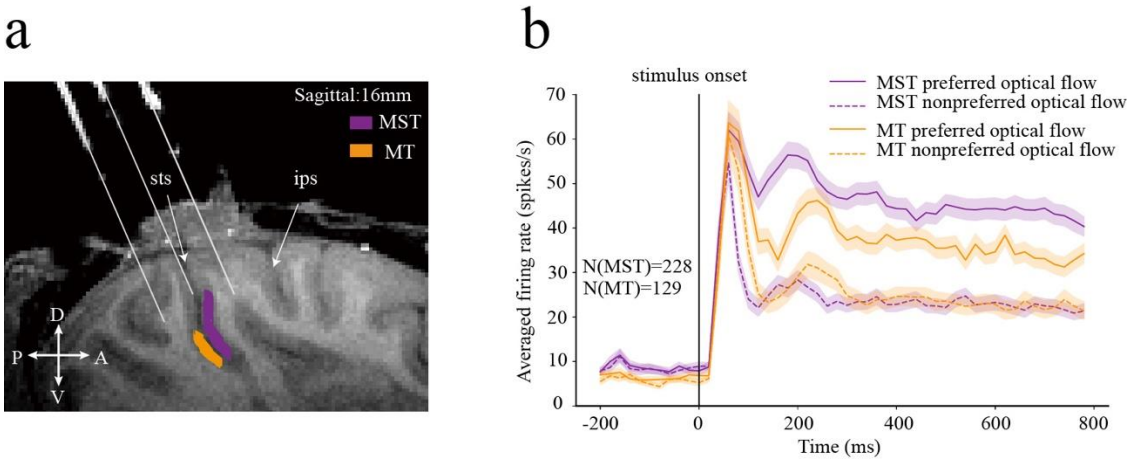


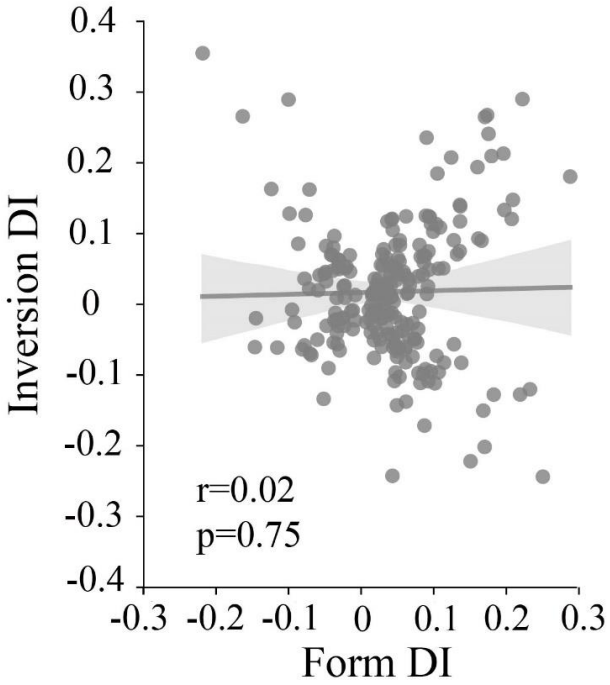
Figure S1: *(a) Chamber-aligned MRI image (scanned with markers) in the sagittal plane showing the regions of interest in one monkey, as implemented by Brainsight Navigator Software; (b): Neurophysiological responses of MST and MT neurons to optic flow patterns.*

2. The authors have shown the scrambled and inversed control separately. The bias (selective BM coding) is clearer in scrambled control data, yet effective size is a bit weak in the inversed control data. Is there a correlation between the two dataset? That is, the authors could show a scatter plot of the data in Figure 3D and Figure 4D for the MST data. If there is a significant correlation between the two, the BM-specific coding in MST would be more persuasive.

We agree with the reviewer that it is interesting to examine whether there exists a correlation between the two biased BM encoding. We followed the suggestion and performed the Pearson correlation analysis between form DI and inversion DI in the MST data. As we can see from the scatter plot (see below), the form and inversion selectivity were not significantly correlated (r(226)=0.02, p=0.75), implying independent representations of these two BM features in area MST.

In the revised manuscript, we reported the result of this analysis and added a few remarks regarding this point in the Results section (Page 12, Line 260-272):

"[…] It is also worth noting that, the two biased encodings in MST, i.e., the BM-specific processing in the form and inversion selectivity, were also independent from each other (r(226)=0.02, p=0.75, Pearson correlation). These independent representations of BM features in area MST suggested that they were probably encoded in distinct neural subpopulations, or in shared neural population but through distinct computations. Either way, these independent encodings do not refute the idea of BM-specific processing in MST. Meanwhile, we should point out that the neural bias being larger in the form selectivity than in the inversion selectivity was not unexpected. In fact, it aligned well with their effect size differences at the behavioral level. Psychophysical studies have shown that scrambling does more damage to BM recognition than inversion (entirely lost by scrambling vs. partially impaired by inversion). It is hence reassuring for us to observe neural effect size differences, i.e., relatively large form bias but smaller inversion bias in the neurophysiological data. […]"



The Pearson correlation between form DI and inversion DI in MST.

3. I think BM perception is a relatively high cognitive function that needs to integrate more information. Is the BM-specific coding related with the size of receptive field? What about MT? The distribution in Figure 3-4D is broad for MT, indicating some cells also prefer BM. Are those cells with a large RF?

We thank the reviewer for these thoughtful questions. In the neurophysiological experiment when we could keep stable recordings of an isolated neuron, we intended to focus on measuring neural responses to eight BM stimuli, four optic flow patterns, and eight translational motions. Thus, for each neuron we did the RF mapping briefly in a relatively coarse manner; the dot patch was repeated only five times at each location in a 9x7 space lattice on the screen (see Methods). In most cases, five repetitions at each spatial location were insufficient for us to obtain a reasonably smooth receptive field map. This precluded us from analyzing the size of RF in a meaningful way. Yet, the center of RF was relatively less susceptible to neural noise, and in practice we have used this coarse estimation of RF eccentricity to guide the placement of visual stimuli in subsequent formal experiments.

Therefore, we examined the correlations between BM-specific encoding and RF eccentricity in area MST and MT respectively, and these analyses yielded intriguing results. As we can see from the figure below, there were weak but significant negative correlations between RF eccentricity and BM-specific encoding (form & inversion DI), and this pattern emerged only in area MST but not in MT. These results suggested that BM-specific processing in the brain was stronger in the fovea than in the periphery, which is highly compatible with the psychophysical studies showing that the performance of BM detection was significantly worse in the periphery than in the central visual field (Ikeda et al 2005; Ceple et al 2023). Our current findings provided the neural correlate of eccentricity-dependent BM perception, indicating that the neural resources for BM perception are primarily concentrated on the central region of the visual field.

In the revised manuscript, we presented the results of these analyses and discussed its implications in the Supplement for interested readers. Our goal was to present analyses in the main text that are directly tied to our primary research questions, while including additional analyses in the Supplement for interested readers.

Reference:

Ikeda H, Blake R, & Watanabe K (2005) Eccentric perception of biological motion is unscalably poor. Vision Res 45(15):1935-1943.

Ceple I, Skilters J, Lyakhovetskii V, Jurcinska I, & Krumina G (2023) Figure-ground segmentation and biological motion perception in peripheral visual field. Brain Sci 13(3).

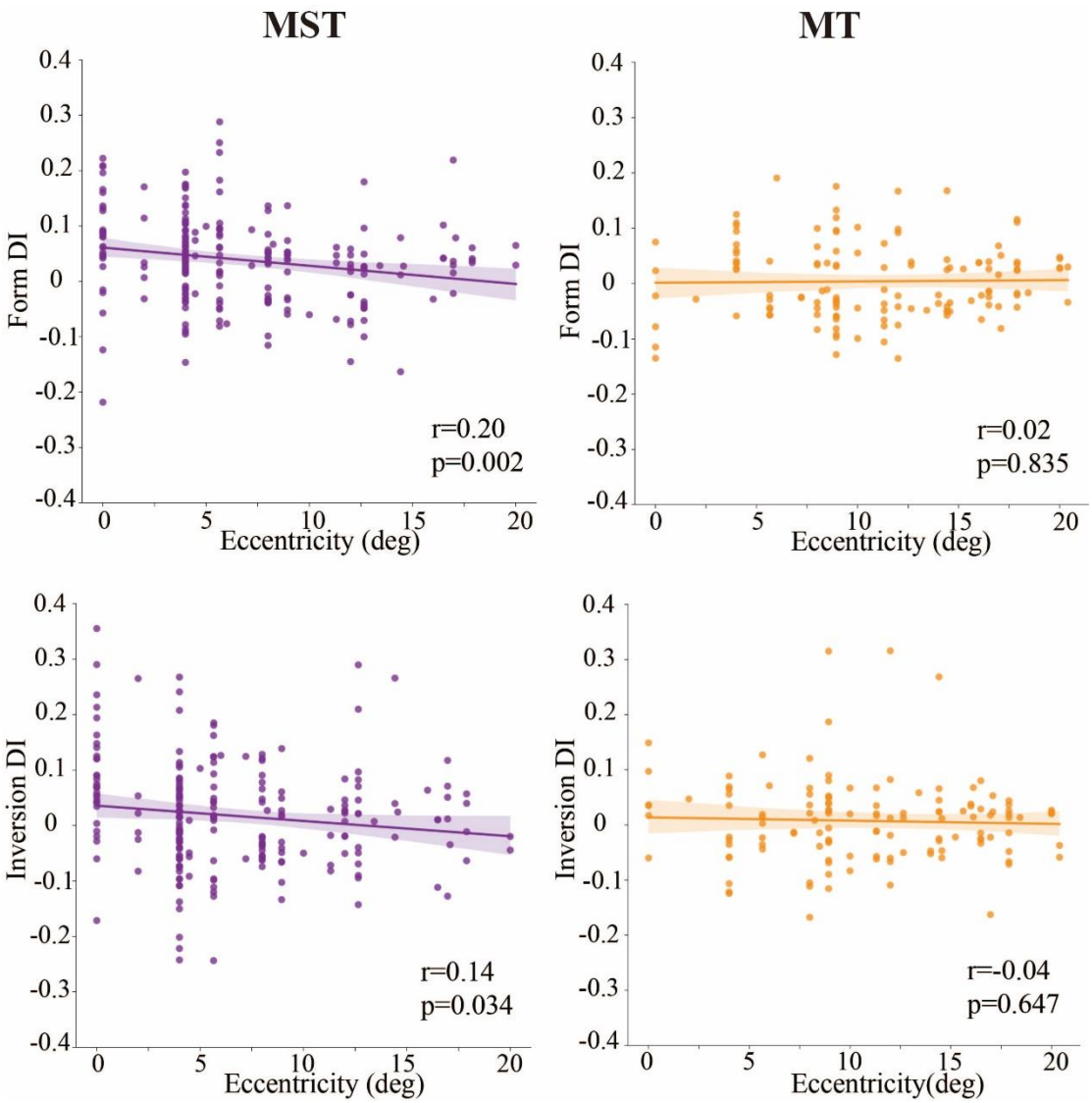


Figure S2: *The Pearson correlation between RF eccentricity and form/inversion DI in area MST (purple) and MT (orange). The results indicated that RF eccentricity was negatively correlated with both form and inversion selectivity in MST but not in MT.*

Reviewer #3 (Remarks to the Author):

This paper explores biological motion tuning in the MT and MST areas of the monkey visual cortex. The findings suggest that neural populations in MST, but not MT, show selectivity for biological motion (BM). The authors also present a feedforward spiking neural network model of MT and MST to investigate the connectivity patterns that drive BM selectivity in the MST population. The question is both interesting and important, and the approach skillfully integrates electrophysiology data with neural network modeling. That said, I believe the paper could benefit from more thorough analysis and some additional work, as I outline below.

We thank the reviewer for the favorable evaluation of our manuscript. We are also grateful for the thoughtful suggestions of additional analysis and experiments, which are enormously instrumental in strengthening our paper.

Major Comments:

- Since one of the key pieces of evidence for MST neurons being sensitive to biological motion is the comparison between normal and inverted BM, it would be helpful if the authors confirm that the recorded population isn't biased in terms of motion direction tuning.

The reviewer made an important point regarding whether the biased form and inversion selectivity in MST are arising from biased sampling in the recorded population. We consider this possibility very unlikely. First, we confirmed that the direction selectivity in area MST (and in area MT as well) was perfectly balanced between left- and right-walking directions, arguing against biased sampling. Second, we found that neither the form selectivity nor the inversion selectivity in MST was correlated with its direction selectivity, indicating that these biased encoding in MST reflects neural features independent from the basic motion direction representations, and should be better interpreted as the neural correlates of the inversion and scrambling effects observed during BM recognition.

In the revised manuscript, we now reported the results of these additional analyses and added a new paragraph to clarify this point in the Results section (Page 11, Line 243-258).

"[…] To sum up, we have so far characterized and compared the selectivity responses of MT and MST neurons to BM stimuli. Our results revealed that, while cells in both brain areas were selective for each of the BM features, cells in MST but not MT showed BM-specific processing. In other words, MST cells had biased form and inversion selectivity, preferring intact (vs. scrambled) and upright (vs. inverted) BM. It should be noted that, the biased form and inversion selectivity in MST could not be trivially explained by biased sampling in the recorded population for two reasons. First, as we showed already in the previous section, that the direction selectivity in area MST (and in area MT as well) was perfectly balanced between left- and right-walking directions (Fig. 2b-2d), arguing against biased sampling. Second, the Pearson correlation analysis revealed that neither the form selectivity nor the inversion selectivity in MST was correlated with its direction selectivity (Inversion DI vs. direction DI: $r(226)=-0.07$, $p=0.28$; Form DI vs. direction DI: $r(226)=0.01$, $p=0.83$). This suggested that these biased encodings in MST reflected neural features independent from the basic motion direction representations, and should be better interpreted as serving the neural correlates of the inversion and scrambling effects observed behaviorally during BM recognition. […]"

- The modeling section could be clearer with more details on the model structure and

training. For example, the authors mention that the MST neurons respond to three types of biological motion (BM), but what exactly are these types? Also, what loss function was used during training? What behavioral output does the model produce? What learning rule was applied to adjust the weights, and what was the format of the input to MT neurons? These additional details would enhance the clarity of the model description.

Thanks for the comment. We have taken this opportunity to improve the clarity of model descriptions in two ways. (1) In the Methods section, we restructured the model descriptions to follow a more logical and cohesive format. We better streamlined this part of the texts with subheadings to improve the clarity of the presentation. (2) In the Results section, we now provided more detailed information about the model and clarify these issues before reporting the statistics to ease the readability. For example,

On model neuron and input (Page 13, Line 286-298, Results)

"[…] In brief, The SNN consisted of an input MT layer and an output MST layer that were fully connected. The number of MT and MST neurons, and their neuron types matched the electrophysiological data. The spike trains of the collected 129 MT neurons during stimulus presentation period (0ms-1500ms, binned at 10ms in each epoch) were provided as the inputs to the model. On a given trial, each MT neuron i (i=1, 2, 3,…129) transmits a spike train S_i^{MT} *to each output MST neuron j (j=1, 2, 3, …, 168), inducing a postsynaptic potential ϵi(t) on MST neuron j. The membrane voltage μj(t) of MST neuron j is the weighted sum of wji·ϵi(t) from all MT neurons. If the μj(t) exceeds a threshold θ (red dotted line), the MST neuron j will emit an output spike (*S_j^a*) and enters the refractory period. The initial connectivity weights wji between MT and MST neurons were randomly and were updated during training. After training, the model reproduced the spike trains of the collected 168 BM-selective MST neurons under each stimulus condition. […]"*

On learning rule, cost function, and model output (Page 14, Line 317-322, Results)

"[…] We used the tempotron learning rule to modify the synaptic weights, minimizing a cost function that measures the amount of timing deviations between the actual and desired output spikes (32). In an iterative learning scheme, synaptic weights were updated (either increasing or decreasing) by an amount proportional to the cost at each iteration (see Methods). The output of the model was expected to reproduce the recorded MST responses after training. […]"

On learning rule, cost function, and model output (Page 32, Line 768-789, Methods)

"[…] ***Synaptic plasticity*** *(subheading)*

We used the tempotron algorithm to modify the connectivity weights in the model (32). The purpose of weights update was to minimize the amount of spike timing deviations between the actual and desired output spikes. When a BM stimulus (such as intact-left-

upright) is presented to the neural network, MST neurons in the target populations ('intact', 'left', and 'upright') are regarded as target neurons, whereas those in the non-target populations ('scramble', 'right' and 'inverted') are designated as non-target neurons. Each target neuron will have a target spike train S = [t1, t2, t3] (output spikes that the MST neuron needs to learn), which are randomly generated within the specified three windows t1 ∈T1win = [9ms, 29ms], t2 ∈T2win = [60ms, 80ms], t3 ∈T3win = [110ms, 130ms]. The time windows were determined by the peak responses of MST neurons in each oscillation cycle. On the contrary, each non-target neuron keeps silent for the entire simulation duration.

During learning, synaptic weights were updated by an amount proportional to the spike timing deviations. Specifically, the weights were modified depending on whether any actual output spikes occur within the target time windows. There were three scenarios. If the target neuron fires within in the time window, which means that it recognizes the BM feature correctly and its weights remain unchanged. If the target neuron keeps silent, its weights will be increased to make this neuron fire at this window in the future. Conversely, if the target neuron emits a spike outside the time window or the non-target neuron emits a spike, which means the neuron incorrectly fires and its weights will be decreased to make this neuron silent in the future. […]"

- The accuracy reported in Figure 5 could be better explained—it's not clear what task this accuracy is referring to. Additionally, it would be helpful to include the r-squared value to show how well the model neurons capture the activity of real neurons.

We now clarified the notion of task accuracy in both the Results and Methods sections:

(1) In the Results, we introduced it before reporting its statistic (Page 14, Line 322-325):

"[…] To evaluate the performance of model output, we computed the accuracy as a measure of similarity between biological and model MST neuron pools using Hamming distance (see Methods). The training and test accuracy were computed at every time epoch. […]"

(2) In the Methods, we described it with mathematical details (Page31, Line 744-765):

"[…] For each recorded MT and MST neuron, there are a total of 160 samples (trials) of BM stimuli (8 BM stimuli X 20 repetitions at each stimulus condition), 120 motion samples are used for training and 40 for testing. The model learned the features of training samples with 200 epochs. For a given input sample (e.g., the kth input sample), there is a target label yk (actual classification), and a predicted label zk from model output (predicted classification), they are both 6-dimensional binary vectors (because 6 pools of MST neurons were used to classify 6 opposing BM features). For yk, the target population of each opposing population is labeled 1 and the non-target population is labeled 0. For example, when the input spike trains of kth sample ('intact-right-upright') are fed the neural network, the target labels yk of these populations intact-scramble-left-right-upright-inverted are yk = [1, 0, 0, 1, 1, 0]. The model used a

winner-take-all strategy in the output of each opposing pools to signify the stimulus. For zk, each opposing populations have at most one winner population labeled 1, i.e., the population with the higher average firing rate is the winner. The accuracy of a sample k is to calculate the Hamming distance-based similarity between these two binary vectors (yk and zk,), representing the probability of correct feature classification based on the responses of 6 pools of MST neurons (Eq.4):

$$Accuracy_k = 1 - \frac{\sum_{i=1}^{6} |y_k^i - z_k^i|}{6} \quad (4)$$

where y_k^i and z_k^i are the ith index in yk and zk, respectively, and $|y_k^i - z_k^i|$ measures the Hamming distance between the two binary vectors (ranging between 0 and 6). The training (test) accuracy during each epoch is the average accuracy of all training (test) samples. […]"

In addition, to quantify the similarity between the model and biological neuron we computed R^2 values from the Pearson correlation analysis and reported these statistics in the revised text as suggested (Page 14, Line 332-337).

"[…] As an illustration, Fig. 5d plotted the spiking responses of a model MST neuron (left panel) and a biological MST neuron (right panel). The Pearson correlation analysis indicated that the model neuron and the biological neuron had very similar response profiles (Upright-intact-left: R2=0.82, p=5.8x10-20; Inverted-intact-left: R2=0.35, p=2.4x10-7), and both of them were inversion-selective and preferred upright BM […]"

- The connectivity analysis in Figure 6 could use stronger support. In many of the plots, the differences between conditions seem to hinge on just a few data points. Including ablation experiments would make the findings more compelling. For example, if the authors argue that a specific type of MT neuron plays a critical role in a certain type of MST neuron, then lesioning this MT type should result in a corresponding loss of MST selectivity, as predicted by their theory. Could this be incorporated into the analysis?

We thank the reviewer for this good idea. We closely followed this suggestion and implemented the ablation experiment in the model. The results from the ablation experiment provided additional compelling evidence supporting our argumentation.

In the revised manuscript, we expanded the original Fig. 6 with three more subpanels showing new evidence. Correspondingly, we added two new paragraphs in the Results section to describe these results (Page 18, Line 402-434).

"[…] The connectivity pattern implicated from the model suggested that, while linear motion selective MT neurons projected to each MST cell type equally in an undifferentiated manner, nonlinear optic flow selective MT neurons projected

preferentially to MST intact-preferring form cells (Fig. 6g). It seemed that the nonlinear selectivity in the MT population was the critical driving force for form selectivity in area MST. We tested this idea directly with an "ablation" experiment in the model. The reasoning behind was, if nonlinear selectivity in MT plays a critical role in shaping form selectivity in MST, then "lesioning" nonlinear selective MT neurons would impair mainly the form selectivity in MST, while leaving the inversion and direction selectivity relatively intact. In the ablation experiment, we first trained the model with the full MT-MST connectivity as we did previously, we then "lesioned" the model by removing MT neurons whose nonlinear selectivity index were greater than 0.33 (i.e., when responding to optic flow, the preferred response was twice as large as the non-preferred response). This allowed us to assessed the form, inversion and direction selectivity of the simulated MST population before and after lesion (Fig. 6d-f). There were two noteworthy observations. First, similar to neurophysiological results, the simulated MST neurons in the full model exhibited biased form selectivity (intact-preferring) and biased inversion selectivity (upright-preferring), but balanced direction selectivity. Second, "lesioning" MT neurons with strong nonlinear optic flow selectivity resulted in a dramatic loss of form selectivity in area MST (Form DI difference = 0.42, $t(167)=8.31$, $p=2.15x10-14$, paired t-test), while the inversion selectivity was well preserved (Inversion DI difference = 0.0045, $t(167)=0.11$, $p=0.91$, paired t-test) and the direction selectivity only marginally impacted (Direction DI difference=0.09, $t(167)=2.09$, $p=0.04$, paired t-test).

To conclude, the results from these two modeling experiments (Fig. 6a-c & Fig. 6d-f) provided converging and compelling evidence that MT subpopulations selective for linear and nonlinear motion projected preferentially to distinct MST cells that support BM recognition. To be more specific, the ability to detect BM form in MST neurons depended mainly on inputs from MT cells selective for nonlinear optic flow, while the ability to detect BM direction and inversion (horizontal and vertical spatial transformations) in MST neurons might be driven primarily by inputs from MT cells selective for linear translational motion (Fig. 6g). […]"

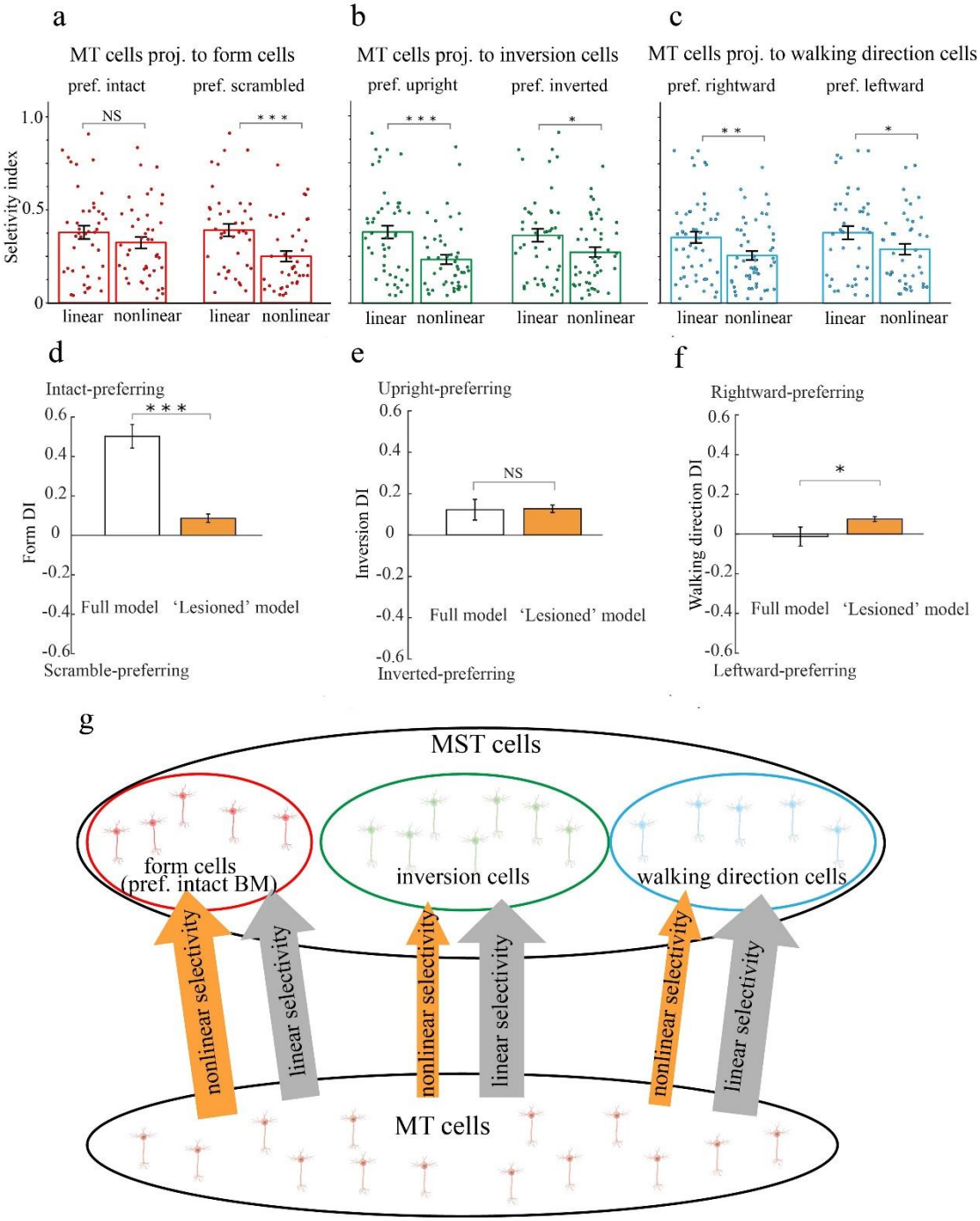


*Figure 6 with three new subpanel **d-f** showing the results of the "lesioning" experiment: Removing nonlinear optic flow selective MT neurons in the model resulted in dramatic loss of form selectivity but relatively preserved inversion and direction selectivity.*

Minor Comments:

- The tuning plots (e.g., Fig 2b-c) require error bars for better clarity of the statistical significance.

Done. We added error bars in these tuning plots (Fig. 2-4, panel b & c) as suggested.

- It would be great if the authors could improve figure resolution. It would make the visuals much more clear.

Done. We now used more distinctive color separations and thicker lines in the figures to enhance their visibility.

Review: Hierarchical and distinct motion processing……. Feng et al.

General:
A very interesting and well written paper that focuses -at least partially- on an all-important distinction about how the brain processes self- versus object motion. Most publications in this field of interest have focused on the self-motion aspect when dealing with motion-sensitive neurons in the parietal cortex and elsewhere.

In that context, the authors could discuss (and take suggestions for future experimental designs) a combination of linear motion and optic flow situations, e.g., if BM also involves expansion/contraction with different foci of expansion/contraction (i.e., simulating a real-life situation when a person/object moves towards/away from an observer). After all, most observed and executed movements consist of combinations of translations and rotations. Furthermore, the aspect of self-motion detection versus object motion was only briefly touched (l. 442 ff; see also l. 394 ff). In this broader context, the authors might want to consult the following publications:

Bremmer, F., S. Ben Hamed, J.-R. Duhamel and W. Graf. Heading encoding in the macaque ventral intraparietal area (VIP). Eur. J. Neurosci. 16: 1554-1568, 2002.

Bremmer, F., F. Klam, J.-R. Duhamel, S. Ben Hamed and W. Graf. Visual-vestibular interactive responses in primate ventral intraparietal area (VIP). Eur. J. Neurosci. 16: 1569-1586, 2002.

Bremmer, F., A. Schlack, W. Graf and J.-R. Duhamel. Multisensory self-motion encoding in parietal cortex. Visual Cognition 11: 161-172, 2004.

Fetsch C.R., DeAngelis, G.C., and Angelaki D.E. (2010) Visual-vestibular cue integration for heading perception: applications of optimal cue integration theory. Eur. J. Neurosci. 31: 1721-1729.

Angelaki, D.E., Gu, Y., and DeAngelis, G.C. (2011) Visual and vestibular cue integration for heading perception in extrastriate visual cortex. J. Physiol. (Lond.) 589: 825-833.

Figures:
Better color separation would be helpful especially for the raster and graph displays (Fgs. 2-4). Also Fig. 1a could be improved.

References:
The reference presentation needs to be cleaned up for uniformity: Some references are in upper case (e.g., ref 15 and others), others list as “author et al.” (e.g., reference 30 and others).

The all-important reference to previously analyzed data (#28; l. 114) cannot be accessed to an incomplete citation). Either the authors provide details in the manuscript, or they wait until a proper reference becomes available.