Systems/Circuits

Neural Mechanism for Coding Depth from Motion Parallax in Area MT: Gain Modulation or Tuning Shifts?

Zhe-Xin Xu and Gregory C. DeAngelis
Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York 14627

There are two distinct sources of retinal image motion: objects moving in the world and observer movement. When the eyes move to track a target of interest, the retinal velocity of some object in the scene will depend on both eye velocity and that object's motion in the world. Thus, to compute the object's velocity relative to the head, a coordinate transformation must be performed by vectorially adding eye velocity and retinal velocity. In contrast, a very different interaction between retinal and eye velocity signals has been proposed to underlie estimation of depth from motion parallax, which involves computing the ratio of retinal and eye velocities. We examined how neurons in the middle temporal (MT) area of male macaques combine eye velocity and retinal velocity, to test whether this interaction is more consistent with a partial coordinate transformation (for computing head-centered object motion) or a multiplicative gain interaction (for computing depth from motion parallax). We find that some MT neurons show clear signatures of a partial coordinate transformation for computing head-centered velocity. Even a small shift toward head-centered velocity tuning can account for the observed depth-sign selectivity of MT neurons, including a strong dependence on speed preference that was previously unexplained. A formal model comparison reveals that the data from many MT neurons are equally well explained by a multiplicative gain interaction or a partial transformation toward head-centered tuning, although some responses can only be well fit by the coordinate transform model. Our findings shed new light on the neural computations performed in area MT, and raise the possibility that depth-sign selectivity emerges from a partial coordinate transformation toward representing head-centered velocity.

Key words: **coordinate transformation; depth; motion; visual cortex**

Significance Statement

Eye velocity signals modulate the responses of neurons in the middle temporal (MT) area to retinal image motion. Two different types of interactions between retinal and eye velocities have previously been considered: a vector addition computation for computing head-centered velocity, and a multiplicative gain interaction for computing depth from motion parallax. Whereas previous work favored a multiplicative gain interaction in MT, we demonstrate that some MT neurons show clear evidence of a partial shift toward coding head-centered velocity. Moreover, we demonstrate that even a small shift toward head coordinates is sufficient to account for the depth-sign selectivity observed previously in area MT, thus raising the possibility that a partial coordinate transformation may also provide a mechanism for computing depth from motion parallax.

Introduction

Living in a dynamic, three-dimensional world, many animals rely heavily on visual information to infer the motion and depth of objects. The retinal image motion of an object generally reflects both self-motion (including translation and rotation of the head, as well as rotation of the eyes) and movement of the object in the world. Depending on the viewing context, retinal image motion can be used to compute different properties of an object, such as its depth or its velocity relative to the world.

When an observer is translating in the world and fixates a world-fixed target by making a compensatory eye rotation, the retinal velocity of a stationary object depends on the distance from the object to the observer's point of fixation (Howard and Rogers, 2002). In this case, retinal image motion relative to the point of fixation (or motion parallax [MP]) provides crucial information about the depth of the object (see Fig. 1*D*). Theoretical and psychophysical studies have shown that depth perception from MP depends on the ratio of retinal velocity to eye velocity, as specified by the motion-pursuit law (Nawrot and Stroyan, 2009). Moreover, previous studies show that humans use these signals to judge depth (Rogers and Graham, 1979, 1982; Ono et al., 1986, 1988; Rogers, 1993; Bradshaw and Rogers,

Received June 17, 2021; revised Dec. 5, 2021; accepted Dec. 9, 2021.
Author contributions: Z.-X.X. and G.C.D. designed research; Z.-X.X. performed research; Z.-X.X. analyzed data; Z.-X.X. wrote the first draft of the paper; Z.-X.X. and G.C.D. edited the paper.
This work was supported by National Institutes of Health Grant EY013644 to G.C.D.; and National Eye Institute CORE Grant EY001319.
The authors declare no competing financial interests.
Correspondence should be addressed to Gregory C. DeAngelis at gdeangelis@ur.rochester.edu.
https://doi.org/10.1523/JNEUROSCI.1240-21.2021

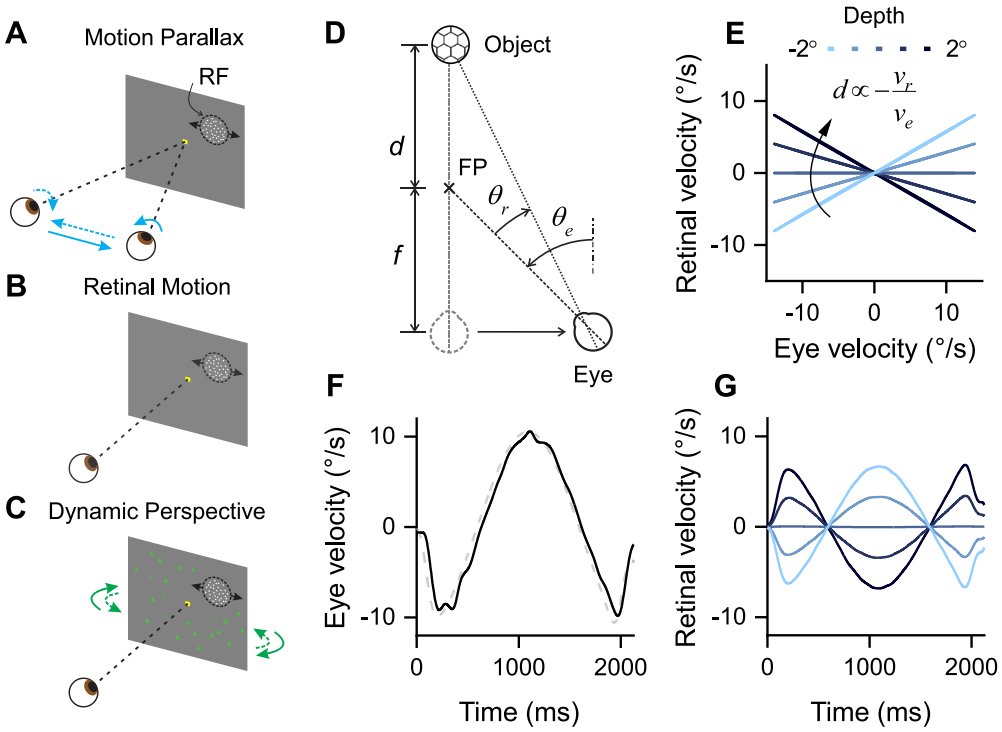

Figure 1. Schematic illustration of experimental conditions and viewing geometry. ***A***, MP condition. The animal was laterally translated by a motion platform, while maintaining fixation at a world-fixed target (yellow square) by smooth pursuit. A patch of random dots was presented monocularly and moved over the neuron's receptive field (RF), providing MP cues to simulate near or far depths. ***B***, RM condition. The animal remained stationary and fixated a static target at the center of the screen. The image motion of the patch replicates that in the MP condition. ***C***, DP condition. Conditions are the same as the RM condition, except that a large-field 3D background (green triangles) was present. The image motion of the background dots provided information about (simulated) eye movements. ***D***, Geometry of the viewing context typically used in studies of depth perception based on MP. An observer translates rightward while smoothly rotating their eye to track a stationary target, *FP*, at viewing distance *f*. The relative depth, *d*/*f*, of an object can be obtained from the ratio of its retinal image motion $v_r \equiv d\theta_r/dt$ and the observer's eye velocity, $v_e \equiv d\theta_e/dt$. ***E***, Mapping of different depths onto the retinal velocity versus eye velocity plane. In this space, depth is described by the slope of a line that passes through the origin. Depths that are closer than the fixation point ("near") have positive slopes (lighter blue lines), while far depths have negative slopes (darker blue lines). ***F***, Temporal profile of average eye velocity (solid black curve) and idealized target velocity (dashed gray curve) in an experimental session. ***G***, Temporal trajectories of retinal velocities corresponding to different depths.

1996; Nawrot, 2003). Neurophysiological studies have demonstrated that neurons in the middle temporal (MT) area of macaque monkeys are selective for the sign (near vs far) of depth defined by MP (Nadler et al., 2008), and that this selectivity is realized by combining retinal image motion with smooth eye movement command signals (Nadler et al., 2009). A more recent study reported that MT responses are gain modulated by the direction of eye movement to generate depth-sign selectivity (Kim et al., 2017). Moreover, previous work has also shown that MT neurons are selective for depth from MP when body translation and eye rotation are simulated by large-field visual motion, such that "dynamic perspective" (DP) cues are provided (Kim et al., 2015). Together, these findings suggest that signals related to eye velocity in MT may serve the goal of computing depth from MP. So far, however, there are no direct measurements of how MT neurons are jointly tuned for eye velocity and retinal velocity, which may clarify the exact nature of the interaction between these variables in the neural responses.

When an object is moving in the world and a stationary observer tracks a moving target with their eyes (see, e.g., Fig. 5*A*), the retinal velocity of the object is the difference between eye velocity and the object's velocity relative to the scene. Thus, to estimate the object's velocity in head-centered coordinates (or world-centered, if the head/body do not move), the visual system must vectorially subtract the image motion produced by eye movement, which is quite different from the computation involved in the motion-pursuit law. Because the observer is stationary in the world in this viewing geometry, eye rotation and the resultant retinal image motion do not provide information about depth (unlike the case of a translating observer illustrated in Fig. 1*D*). At the neural level, the interactions between retinal and eye velocity signals might need to be different for performing a coordinate transformation (CT) and computing depth from MP, but this issue has not been examined. Previous studies have examined how neurons in areas MT and MST represent object motion during smooth pursuit eye movements (Inaba et al., 2007, 2011; Chukoskie and Movshon, 2009). They measured neural speed tuning in both retinal and screen coordinates during pursuit eye movements, and investigated whether tuning in areas MT and MST was shifted toward screen coordinates when smooth eye movements were executed (screen and head coordinates are equivalent in their experiments). While these studies generally found much greater effects in MST neurons, the response shifts of some MT neurons were not negligible. Thus, whether or not MT neurons convey useful information about object velocity relative to the head remains an open question. Moreover, it remains unclear whether there is any relationship between selectivity for velocity in head-centered coordinates and selectivity for depth from MP.

It has long been established that neurons in area MT are sensitive to binocular disparity and local two-dimensional motion signals (Zeki, 1974; Maunsell and Van Essen, 1983a,b; Albright, 1984; Felleman and Kaas, 1984; Rodman and Albright, 1987; DeAngelis and Newsome, 1999; Born and Bradley, 2005), and

that their responses correlate with the perception of depth and motion (Britten et al., 1992, 1996; DeAngelis et al., 1998; DeAngelis and Newsome, 2004; Uka and DeAngelis, 2004). However, relatively little is known about how MT neurons contribute to the computation of higher-level perceptual variables, such as depth from MP and object motion in the world. Existing evidence, as described above, suggests a potential role of MT neurons that goes beyond representing retinal motion (RM), but the detailed mechanisms by which other variables, such as eye movements, modulate MT responses are not well understood.

In this study, we analyze data from previous experiments on depth from MP, and devise a more rigorous approach to examine the interactions between eye velocity and retinal image velocity in the responses of MT neurons. We formulate a model in which eye velocity modulates the gain of MT responses, substantially extending the work of Kim et al. (2017). We fit this model to MT spike trains and show that the joint tuning for eye and retinal velocities is well described by this gain modulation (GM) model for many MT neurons. However, the joint tuning for other neurons reveals a clear diagonal structure that cannot be explained by a mechanism of gain modulation (Kim et al., 2017). We further develop a computational model to account for responses that are shifted, to varying degrees, toward head-centered velocity tuning (i.e., tuning for object motion in the world). Our head-centered tuning (HT) model also successfully predicts the depth tuning curves of many MT neurons, including those with diagonal structure, and further predicts an empirically observed relationship between depth-sign selectivity and speed preference that was previously unexplained. Our simulations make specific testable predictions for experiments that could distinguish between mechanisms based on gain modulation versus partial coordinate transformation. Our findings suggest that the representation of velocity is transformed partially toward head-centered coordinates, and that this transformation could account for the genesis of depth-sign selectivity.

Materials and Methods

Experimental apparatus and stimuli

Electrophysiological data examined in this study were taken from a series of studies of depth selectivity based on MP in macaque area MT, which have been described in detail previously (Nadler et al., 2008, 2009; Kim et al., 2015). In brief, 4 male macaque monkeys sat in a chair mounted on top of a motion platform with 6 df (MOOG 6DOF2000E, Moog) that allows translation along any axis in 3D space (Gu et al., 2006; Nadler et al., 2008). The animals viewed stimuli presented on a 60 cm × 60 cm tangent screen at a viewing distance of 32 cm, such that the display subtended ~90° × 90°. Stimuli were presented stereoscopically as red-green anaglyphs using Kodak Wratten2 filters (red no. 29, green no. 61), or were presented monocularly by presenting just red or green elements on the display. A three-chip digital light projector (Mirage S + 3k, Christie Digital) with a resolution of 1280 × 1024 pixels was used to project visual stimuli onto the screen. Both the video display and movements of the motion platform were controlled and updated at 60 Hz. Visual stimuli were generated by a custom C++ software using the OpenGL graphics library.

The room was dark except for the faint backlight of the video projector. While the faint background texture on the display screen could weakly activate some MT neurons, as shown previously (Chowdhury et al., 2009), this was observed when there was no other stimulus in the receptive field. In our experiments, there was always a high-contrast motion stimulus in the receptive field, such that effects of the faint background texture are likely to be negligible. Motion of the faint background texture outside the receptive field could potentially affect MT responses through center-surround interactions, but we have shown previously that there is no significant correlation between the strength of depth-sign tuning and the strength of surround suppression (Kim et al., 2015).

Neural recording protocols

Extracellular single-unit responses were recorded using tungsten microelectrodes (FHC). For each neuron, basic tuning properties, including speed and direction tuning, were measured, followed by a main experimental protocol that involved three stimulus conditions: (1) In the MP condition (see Fig. 1*A*), a motion platform was used to laterally translate the animal sinusoidally along the preferred-null axis of each recorded neuron (for details, see Nadler et al., 2008). The movement was one cycle of a 0.5 Hz sinusoid multiplied by a windowing function to avoid sharp onset acceleration, and had two starting phases (0° and 180°). A world-fixed target was presented binocularly throughout the movement, and animals were trained to fixate this target by making smooth pursuit eye movements in the opposite direction to head translation. A random-dot patch was placed over the neuron's receptive field and viewed monocularly, with the motion of the patch computed to provide MP consistent with 1 of 9 simulated depths (0°, ±0.5°, ±1°, ±1.5°, ±2°). (2) In the RM condition (see Fig. 1*B*), the animal remained stationary while fixating on a static target. The same image motion over the receptive field as in the MP condition was generated by moving the OpenGL camera in the virtual environment along the same trajectory as in the MP condition. Consequently, this condition contained identical image motion of the patch relative to the fixation target as presented in the MP condition, but no pursuit eye movements were made. The size of the random-dot patch was small enough such that the visual cues that can provide information about eye rotation (i.e., DP cues) were negligible in this condition (Kim et al., 2015). (3) In the DP condition (see Fig. 1*C*), the image motion of the random-dot patch and the fixation target were the same as those in the RM condition. However, a 3D background of random dots was presented, whose image motion simulated the optic flow that results from lateral body translation and compensatory smooth pursuit eye movements, as made in the MP condition (for details, see Kim et al., 2015). Thus, the background dots in the DP condition provided robust information about eye rotation relative to the scene in the absence of real eye movements.

We excluded cells that did not show strong response modulations by retinal image motion. We first calculated the difference in firing rates between the two opposite movement phases (0° and 180°) for each depth in the RM condition. This subtraction canceled out baseline activity and stimulus onset transients that were not dependent on depth, allowing us to isolate the response modulation induced by image motion. A Fourier transform was then applied to obtain the amplitude spectrum of the neural response. We took the median amplitude at frequencies in the range from 0.5 to 1 Hz, and computed the significance of that amplitude by shuffling spike trains 10,000 times across time bins. A frequency range of 0.5-1 Hz was used because some neurons with very slow-speed preferences have a nearly frequency-doubled response to our modified 0.5 Hz stimulus. Significance level was Bonferroni-corrected for multiple comparisons. Cells that showed significant modulation amplitudes at less than two depths with the same sign (both near or both far) were excluded from analyses; ~8% of the cells were removed in this fashion, and most were neurons tuned to high speeds that were not well driven by our stimuli (which had a maximal retinal velocity of ~7°/s).

Joint tuning for retinal and eye velocities

Vertical and horizontal eye position signals were recorded by an eye coil system at a sampling rate of 200 Hz. The eye position signals were calibrated at the beginning of each recording session with an eye window of 1.5°-2°. We linearly interpolated the raw data to 1 kHz, then rotated it to obtain signals aligned with the preferred-null axis of the neuron under study. The position data were filtered by a Gaussian window having an SD, σ = 33 ms. The MATLAB function *filtfilt* was used to prevent phase shifts. The velocity of eye movements was then computed by taking the derivative of the filtered position signals. We calculated the theoretical retinal image position of the center of the random-dot patch at each video frame based on the simulated geometry and OpenGL projective parameters. We then calibrated this signal by adding the positional

difference between the eye and target during pursuit to find the actual retinal image position of the center of the patch at each time point. The first derivative of a Gaussian filter (σ = 33 ms) was then applied to the image positions to obtain retinal stimulus velocities.

To estimate the joint tuning of MT neurons for retinal and eye velocities, we discretized the joint instantaneous velocities in each trial into 1×1°/s bins, and we counted the number of spikes associated with each pair of velocities (see Fig. 2). The mean firing rate in each 1×1°/s bin was calculated as the total number of spikes divided by the number of samples of that particular pair of retinal and eye velocities, and then multiplied by the 1 kHz sampling rate. Bins with <200 velocity samples were discarded from the analysis. In the MP condition, the actual eye velocity at each time point in the trial was used. In contrast, for the RM and DP conditions, the simulated eye velocity (i.e., the idealized eye velocity that would be consistent with the visual images presented) was used to plot the joint velocity tuning profiles.

Simulation of head-centered speed tuning

To explore how depth-sign selectivity might arise for neurons that are tuned to velocity in head coordinates, we simulated neural responses with head-centered velocity tuning and examined their joint tuning for retinal velocity and eye velocity (see Fig. 5), as well as their corresponding depth tuning curves. We simulated 500 neurons with velocity tuning that could range from being in retinal to head-centered coordinates, and with a diverse range of speed preferences. Velocity coordinates can be specified as a weighted sum of retinal and eye velocities, with an adjustable weight on eye velocity. For the i-th simulated neuron, spike rates $\mathbf{r}_i$ were randomly sampled from a Poisson distribution as follows:

$$\mathbf{r}_i \sim \text{Poisson}(\lambda_i) \tag{1}$$

where λ_i is the mean firing rate for a given pair of retinal velocity, $\mathbf{v}_r$, and eye velocity, $\mathbf{v}_e$, values:

$$\lambda_i = A_i f_i(\mathbf{v}_r + \omega_i \mathbf{v}_e) + B_i \tag{2}$$

Here, A_i is a scaling factor on response amplitude (uniformly sampled from 50 to 100 spikes/s), B_i is the spontaneous firing rate (uniformly sampled from 0-20 spikes/s), and ω_i is the weight placed on eye velocity (uniformly sampled in a logarithmic space from 0.001 to 1). $f_i(\cdot)$ represents the velocity response function for each model neuron. It is well established that the speed and direction tuning curves of MT neurons are multiplicatively separable (Rodman and Albright, 1987), and can be expressed as log-Gaussian (Nover et al., 2005) and von Mises functions (Smolyanskaya et al., 2013), respectively. Therefore, we modeled $f_i(\cdot)$ as the product of these two functions as follows:

$$f_i(\mathbf{v}) = \exp\left(-\frac{1}{2\sigma_i^2}\left(\log\frac{\|\mathbf{v}\| + \delta_i}{s_i + \delta_i}\right)^2\right) \cdot \exp\kappa_i\left(\cos\left(\Phi(\mathbf{v}) - \varphi_i\right) - 1\right) \tag{3}$$

where $\|\mathbf{v}\|$ and $\Phi(\mathbf{v})$ indicate the magnitude (speed) and direction of a given 2D velocity vector $\mathbf{v} = \begin{pmatrix} v_x \\ v_y \end{pmatrix}$, respectively. s_i and φ_i denote the preferred speed and direction, respectively, where s_i is uniformly sampled in a logarithmic space from 0.1 to 30°/s. Because both retinal image motion and eye movements were aligned with each neuron's preferred-null motion axis in each experiment, there is no need to simulate neurons with different direction preferences; we therefore simplified the model such that the preferred direction of each neuron was $\varphi_i = 0°$. σ_i and κ_i determine the width of speed and direction tuning curves, respectively, with $\sigma_i \in [0.5, 1.5]$ and $\kappa_i \in [1, 2]$. δ_i is a small constant that keeps the logarithm from being undefined as the speed $\|\mathbf{v}\|$ goes to zero, and was drawn from $\delta_i \in [0.001, 3]$. All parameter ranges were chosen based on empirical data from area MT (Nover et al., 2005).

We repeated the simulations using a Gaussian speed tuning profile instead of the log-Gaussian function in Equation 3, and we obtained very similar results. Therefore, the pattern of depth selectivity in our simulated neurons is not attributable to the particular nonlinearity of the log-Gaussian function.

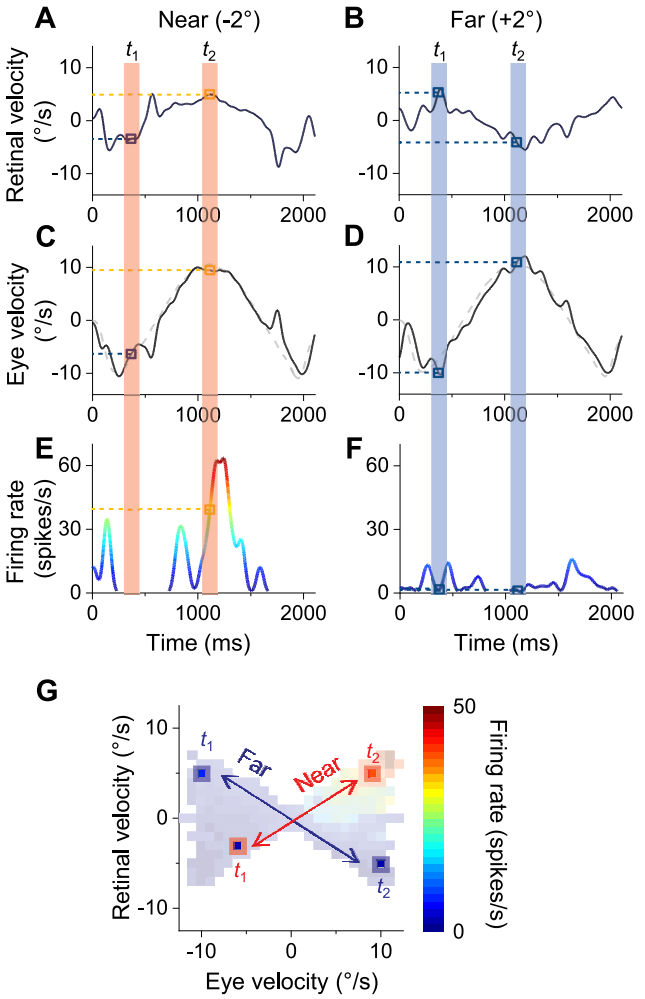

Figure 2. Mapping temporal responses onto the joint retinal and eye velocity tuning profile. ***A***, ***B***, Retinal velocity as a function of time for near (−2°) and far (2°) depths, respectively, in example trials. Shaded bands represent two particular time points, t_1 and t_2. ***C***, ***D***, Trajectories of eye velocity (black curve) and target velocity (dashed gray curve) for the same example trials at near and far depths. ***E***, ***F***, PSTHs of an MT neuron's response for the same example trials in the MP condition for near and far depths. Response at each time point (color coded based on firing rate) corresponds to a pair of retinal and eye velocities. ***G***, Average firing rate in each time bin is mapped to the retinal-eye velocity space according to the corresponding velocities at that time point. Responses in near-depth trials transform to lines with positive slopes (e.g., pink arrow and squares); responses to far depths transform to lines with negative slopes (e.g., blue arrow and squares).

Given the fact that velocity in head coordinates is the vector sum of retinal and eye velocities, $\mathbf{v}_h \equiv \mathbf{v}_r + \mathbf{v}_e$, we can rewrite the input velocity of $f_i(\cdot)$ in Equation 2 for $\omega_i \in [0, 1]$ as follows:

$$\begin{aligned} \mathbf{v}_r + \omega_i \mathbf{v}_e &= \omega_i(\mathbf{v}_r + \mathbf{v}_e) + (1 - \omega_i)\mathbf{v}_r \\ &= \omega_i \mathbf{v}_h + (1 - \omega_i)\mathbf{v}_r \end{aligned} \tag{4}$$

Hence, ω_i is also the weight on velocity in head-centered coordinates; the input velocity becomes completely retinocentric when $\omega_i = 0$, and purely head-centered when $\omega_i = 1$.

To obtain the joint velocity tuning and depth tuning for each simulated neuron, we presented pairs of retinal and eye velocities in a range from −12 to 12°/s, with 0.1°/s spacing. The direction of both velocities was along each neuron's preferred-null axis (0° by convention). For each given depth, d, the depth tuning curve was computed as the averaged

firing rate $\langle \mathbf{r}_i \rangle$ across all stimulus combinations that are selected based on the motion-pursuit law (Nawrot and Stroyan, 2009) as follows:

$$\frac{d}{l} \approx -\frac{\mathbf{v}_r}{\mathbf{v}_e} \tag{5}$$

where l is the viewing distance, and $l = 38$ cm. The motion-pursuit law here contains a negative sign to be compatible with our definition that near depth is negative. We generated 20 simulated responses (with Poisson noise) for each velocity pairing and computed the depth-sign discrimination index (DSDI, described below) from the responses. We also ran these simulations with wider ranges of depths and velocities, and obtained similar results.

It should be noted that the motion-pursuit law (Nawrot and Stroyan, 2009) assumes perfect ocular tracking of the fixation target during translation of the observer, and pursuit eye movements cannot be perfect. We have performed extensive analyses in previous studies to demonstrate that the depth-sign selectivity of MT neurons is not correlated with modest imperfections in pursuit behavior (Nadler et al., 2008, 2009; Kim et al., 2015).

To examine whether depth information can be extracted from model neurons with speed tuning shifted toward head coordinates, we trained a linear decoder to estimate depth d from the responses $\boldsymbol{r}$ of the simulated population as follows:

$$\hat{d} = \mathbf{r}^T \boldsymbol{\beta} = \sum_i r_i \beta_i + \beta_0 \tag{6}$$

where the weights $\boldsymbol{\beta}$ were given by the solution for ordinary least squares as follows:

$$\boldsymbol{\beta} = (\mathbf{r}^T \mathbf{r})^{-1} \mathbf{r}^T \mathbf{d} \tag{7}$$

The stimulus parameters used for these simulations were the same as those described in the simulation section of Kim et al. (2017). We generated responses for 500 simulated trials for each stimulus condition to train the decoder, and used responses from another 500 simulated trials as a test set to evaluate decoding performance.

Depth-sign discrimination index

For both simulated and real neurons, we quantified their depth-sign selectivity using DSDI (Nadler et al., 2008) as follows:

$$DSDI = \frac{1}{n} \sum_{i=1}^{n} \frac{r_{far}^{(i)} - r_{near}^{(i)}}{|r_{far}^{(i)} - r_{near}^{(i)}| + \sigma_{avg}^{(i)}}, n = 4 \tag{8}$$

Here, the difference in average firing rate between each pair of far and near depths was calculated relative to the SD of the response to that pair. DSDI was then obtained by averaging across all four pairs of depths, resulting in DSDI $\in [-1, 1]$. Negative DSDI values indicate near-preferring neurons, and positive DSDIs indicate far-preferring neurons. Statistical significance of DSDI values was computed by a permutation test in which the stimulus labels (near or far) for each depth magnitude were shuffled 1000 times. CIs for DSDI values were obtained by bootstrapping, based on 1000 iterations of resampling with replacement.

Description of models

We developed a series of models that could be directly fit to the spike trains of each recorded MT neuron. We considered three different extraretinal components that could modulate the neural response to RM: gain (or multiplicative) modulation (GM), offset (or additive) modulation (OM), and HT. Because these three components can operate independently, we are able to examine all possible combinations of them and conduct a factorial model comparison. For all models described in this section, the speed and direction tuning function f were the same as in Equation 3.

Control model. As a control, we included a model in which the response is only driven by retinal image velocity as follows:

$$\boldsymbol{\lambda} = A \cdot f(\mathbf{v}_r) + B \tag{9}$$

where $\boldsymbol{\lambda}$ represents the firing rates over stimulus duration T for N trials, A is a scaling factor on response amplitude, and B is the spontaneous firing rate. The dimensions of the input retinal velocity $\mathbf{v}_r$ and eye velocity $\mathbf{v}_e$ variables are $N \times T \times 2$. This control model has 7 parameters in total.

GM model. We constructed a GM model in which the response to retinal velocity is multiplicatively modulated by a function of eye velocity (see Fig. 8*B*) as follows:

$$\boldsymbol{\lambda} = A \cdot g(\mathbf{v}_e) f(\mathbf{v}_r) + B \tag{10}$$

Based on our observations over the range where we have data, the effect of eye velocity is generally monotonic; therefore, we modeled the gain, g, as a sigmoid function with one free parameter α that controls the slope, as follows:

$$g(\mathbf{v}) = \frac{2}{1 + \exp(-\alpha \mathbf{v})} \tag{11}$$

The output of this function is in the range from 0 to 2. When the input eye velocity is zero, $g(0) = 1$, and the neural response is therefore determined solely by retinal velocity tuning. This gain function only requires one extra parameter over the Control model, yielding 8 free parameters in total.

OM model. We developed another model in which eye velocity has an additive effect on the response to retinal velocity (see Fig. 8*C*) as follows:

$$\boldsymbol{\lambda} = A \cdot \left(f(\mathbf{v}_r) + o(\mathbf{v}_e) \right)^+ + B \tag{12}$$

where o is an additive offset term and $(\cdot)^+$ denotes a rectifier that prevents negative firing rates. Similar to the multiplicative gain function, g, we modeled the additive effect of eye velocity as a sigmoid with one free parameter β as follows:

$$o(\mathbf{v}) = \frac{2}{1 + \exp(-\beta \mathbf{v})} - 1 \tag{13}$$

This offset term lies within the range from −1 to 1, thus allowing both suppression and facilitation. This model also has 8 free parameters.

HT model. To test whether MT responses could be explained by velocity tuning in head coordinates compared with the GM and OM models, we fit data with the same model used in the simulations described above (Eq. 2) (see Fig. 8*A*) as follows:

$$\boldsymbol{\lambda} = A \cdot f(\mathbf{v}_r + \omega \mathbf{v}_e) + B \tag{14}$$

where ω is the weight on eye velocity that shifts tuning toward head coordinates, as described in Equations 2 and 4. As for the GM and OM models, this HT model has 8 free parameters.

Full model and two-component models. To further investigate the role of each model component in accounting for MT responses, we integrated the multiplicative gain $g(\cdot)$ term, the additive offset term $o(\cdot)$, and the weighting toward head coordinates, ω, into one comprehensive model with 10 parameters as follows:

$$\boldsymbol{\lambda} = A \cdot \left(g(\mathbf{v}_e) f(\mathbf{v}_r + \omega \mathbf{v}_e) + o(\mathbf{v}_e) \right)^+ + B \tag{15}$$

In addition, we can remove each of the components, one at a time, from the Full model to create two-component models with 9 free parameters each. We refer to these two-component models as –GM, –OM, and

–HT, where, for example, the –GM model is the Full model without the GM component, such that it contains the OM and HT components.

GM-sign model. To explore whether the direction (i.e., sign) of eye movement alone can explain the observed modulations of MT responses (without explicit representation of eye speed), we replaced the gain function $g(\mathbf{v})$ in Equations 10 and 11 with the following:

$$g(\mathbf{v}) = \alpha \cdot \text{sign}(\mathbf{v}) + 1 \tag{16}$$

Similar to the GM model, this model also has 8 free parameters.

Model fitting and comparison

For each neuron, we used spike trains from the main depth-tuning experiments (RM and MP conditions, or RM and DP conditions), as well as from the speed and direction tuning protocols, to fit the above models in two consecutive steps. In the first step, the retinal velocity tuning function (Eq. 3) was fit to spike counts from the speed and direction tuning curves. We used bootstrapping (with 1000 iterations of resampling with replacement) to estimate the CIs of each parameter in the retinal velocity tuning function. In the second step, models described in the previous section were fit to spike trains from the main experiments at 1 kHz resolution (which was the native resolution of the data); 68% and 95% CIs of each tuning parameter estimated in the first step were used here as plausible bounds and hard bounds to constrain the parameters of Equation 3. This fitting procedure acknowledges that there may be overall response differences between direction and speed tuning measurements and the main depth-tuning experiment because of stimulus differences or nonstationary neural responses, while constraining the model fits based on information about tuning curve shapes and stimulus preferences from the independent measurements. We used the MATLAB (The MathWorks) function *fmincon* in the first step of the fitting procedure, and Bayesian adaptive direct search (BADS) (Acerbi and Ma, 2017) in the second step. BADS was used to find the global maximum of the log-likelihood $\log L(\boldsymbol{\theta})$ over N trials and stimulus duration T as follows:

$$\underset{\boldsymbol{\theta}}{\text{argmax}} \log L(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\text{argmax}} \sum_{i=1}^{N} \sum_{t=1}^{T} r_i(t) \log \lambda_i(t) - \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda_i(t) \tag{17}$$

where $\boldsymbol{\theta}$ represents the free parameters in the model being fit, $r_i(t)$ represents the actual spikes at time t on the i-th trial (0 or 1 for each 1 ms time bin), and $\lambda_i(t)$ represents the firing rate predicted by the model.

To assess model performance in terms of fitting raw spike trains at 1 kHz resolution, we computed the Bayesian information criterion (BIC) (Schwarz, 1978) for each model as follows:

$$BIC = -\log L(\hat{\boldsymbol{\theta}}) + \frac{k}{2} \log(N \times T) \tag{18}$$

where k is the number of free parameters and $N \times T$ is the number of data points to be fit.

To quantify the relative contribution and uniqueness of each model component, we examined the change of BIC caused by adding or removing each of the three model components (GM, OM, HT). By comparing the GM, OM, and HT models to the Control model, we were able to examine how much each component contributes to the fit. By removing one component from the Full model (–GM, –OM, –HT models), we were able to measure the uniqueness of the contribution of each component. Peristimulus time histograms (PSTHs) were obtained, for plotting purposes, by applying a moving Gaussian window (σ = 33 ms) to the raw spike data and then averaging across repetitions.

The GM, OM, and HT models have the same number of parameters; therefore, comparing the BIC values is the same as directly comparing the negative log-likelihood of the models, $-\log L(\hat{\boldsymbol{\theta}})$, and the temporal resolution T in the penalty term, $\frac{k}{2}\log(N \times T)$, does not have a significant impact on our results. We also explored fitting the data after binning spikes into 20 ms bins, and our results remained unchanged.

Critically, the models were always fit to the spike trains of MT neurons, and we then examined how well the models predicted the depth tuning curves of the MT neurons. Model performance on predicting depth tuning curves was measured by computing the absolute error of the DSDI prediction. Because we only modeled how the mean firing rate of each neuron depends on eye velocity and retinal velocity, our models do not directly capture the response variability required for the DSDI calculation (σ_{avg} in Eq. 8). To compute DSDI for each model, we first injected Poisson noise into the model estimates of instantaneous firing rates for each trial, and we then used the simulated Poisson responses to calculate the DSDI for each model from Equation 8. The uncertainty of DSDI predictions was estimated by bootstrapping (1000 resamples with replacement). Since the DSDI predictions were generated solely from the fits to retinal and eye velocities in each trial, the models do not have direct information about depth from MP.

Results

We analyzed and modeled data from a series of previous experiments on the neural representation of depth from MP in area MT, with the goal of understanding whether the joint neural tuning for retinal and eye velocities is more consistent with gain modulation or head-centered velocity tuning. We examined data from three experimental conditions. In the MP condition, the animal was laterally translated by a motion platform and counter-rotated their eyes to maintain fixation at a world-fixed target (Fig. 1*A*,*D*). A patch of random dots was presented monocularly over the receptive field of an isolated MT neuron, with its motion simulating 1 of 9 depths. In the RM condition, the animal remained stationary and fixated a static target, while a patch of random dots reproduced the retinal image motion that was experienced in the MP condition (Fig. 1*B*). Since the random-dot stimulus was generated to be depth-sign ambiguous (Nadler et al., 2008) (see Materials and Methods), neural responses are not expected to have depth-sign selectivity in the RM condition. In the DP condition, the animal's posture and the random-dot patch stimulus were the same as in the RM condition, except that a full-field 3D cloud of random-dots was presented in the background (Fig. 1*C*). The image motion of these background dots provided an optic flow field that simulated eye rotation (Kim et al., 2015).

We start by exploring the joint tuning of MT neurons for retinal and eye velocities, followed by simulations which show that depth selectivity can emerge from even a small shift toward velocity tuning in head coordinates. Next, we fit computational models to MT responses to assess the mechanisms that might account for depth-sign selectivity in both the MP and DP conditions.

Joint tuning for retinal and eye velocities

As shown by Nadler et al. (2009), the combination of smooth eye movement command signals and responses to image motion is sufficient to generate depth-sign selectivity from MP in MT neurons. To better understand exactly how these two signals interact, we visualized how depth from MP maps onto a 2D space having dimensions of retinal velocity and eye velocity. According to the motion-pursuit law (Nawrot and Stroyan, 2009), when an observer is translating laterally and fixating a world-fixed target by compensatory pursuit (Fig. 1*D*), depth from MP is approximately proportional to the ratio of retinal velocity and eye velocity (Eq. 5). Hence, depth is defined by the slope of a line that passes through the origin in the 2D velocity plane (Fig. 1*E*).

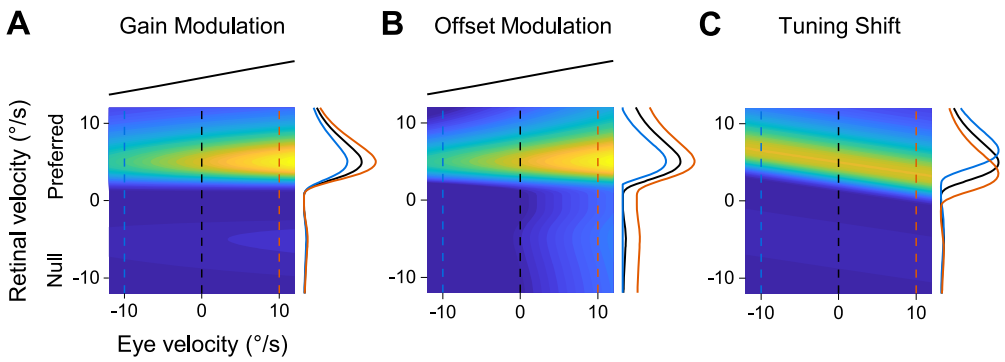

Figure 3. Possible types of interaction between retinal and eye velocities that could produce a preference for depth sign. ***A***, Multiplicative modulation of retinal velocity tuning by eye velocity. ***B***, Additive offset interaction between retinal velocity and eye velocity. ***C***, Retinal velocity tuning that shifts as a function of eye velocity (i.e., velocity tuning that is transformed toward head-centered coordinates). Tuning curves along the margins show retinal velocity responses that are modulated by three different eye velocities (dashed lines).

In the MP condition, retinal and eye velocities covary in a quasi-sinusoidal fashion (Figs. 1*F*,*G*, 2*A*-*D*). Therefore, the idealized trajectory (assuming accurate pursuit) of each trial forms a line (iso-depth line; Fig. 1*E*) whose slope describes the simulated depth of the patch in this trial. Following this logic, we mapped the temporal response of a neuron in each trial (Fig. 2*E*,*F*) onto the 2D velocity space. Joint tuning for retinal and eye velocities was then estimated by mapping the responses for all trials onto this velocity plane (Fig. 2*G*). The eye velocity (Fig. 2*G*, *x* axis) for the RM and DP conditions represents the eye velocity that was visually simulated while the real eye was stationary. It is also worth noting that the magnitude of MP-based depth scales up rapidly and eventually becomes undefined as the iso-depth line approaches vertical (Eq. 5). Since the range of simulated depths in the experiments was limited to the range from −2° to 2° of equivalent disparity, the available data only sample a portion of this joint velocity tuning space. This results in the "bow-tie" shape of the data (Fig. 2*G*; see Fig. 4).

We consider three general types of interaction between retinal and eye velocities (Fig. 3) that could describe the joint tuning of MT neurons: multiplicative gain modulation (Fig. 3*A*), additive offset modulation (Fig. 3*B*), and a gradual shift in retinal velocity tuning with eye velocity (Fig. 3*C*). Both multiplicative and additive modulations have been observed in neuronal responses in the visual cortex (e.g., Smolyanskaya et al., 2013; Arandia-Romero et al., 2016), and we include both effects in our models such that their influences can be distinguished. Both multiplicative and additive modulations would predict responses to be suppressed when the eye moves in a particular direction and facilitated when the eye moves in the opposite direction (Fig. 3*A*,*B*). On the other hand, when eye velocity causes a shift in the retinal velocity tuning curve, the joint tuning will exhibit a diagonal structure (Fig. 3*C*).

Figure 4*A*–*C* shows data from the most typical type of neuron that we found in MT. In the RM condition, firing rate primarily varies with retinal velocity but not with (simulated) eye velocity (Fig. 4*A*). The corresponding depth tuning curve for the RM condition is thus approximately symmetric around a depth of zero (Fig. 4*C*, solid squares). In the MP condition, the response to retinal image motion was facilitated by positive eye velocities and suppressed by negative eye velocities (Fig. 4*B*). In line with previous findings, this pattern is well described by a multiplicative modulation of responses to image motion by eye velocity (Kim et al., 2017), such that the joint tuning profile is multiplicatively separable (Fig. 3*A*). This gain modulation leads to a highly asymmetric depth tuning curve (Fig. 4*C*, open circles), yielding a near preference for depth.

We found that other neurons, which prefer very slow retinal velocities and are substantially less common, show a strikingly different pattern of joint velocity tuning. These slow-speed-preferring cells exhibit RM responses that peak near zero retinal speed (Fig. 4*D*,*G*), producing symmetric depth tuning curves with a peak at zero depth (Fig. 4*F*,*I*, solid squares). In the MP condition, the retinal velocity tuning of these neurons is neither uniformly suppressed nor facilitated by eye velocity, as expected from gain modulation. Instead, the peak of the retinal velocity tuning shifts systematically with eye velocity, causing a negative diagonal structure in the joint tuning profile (Fig. 4*E*,*H*). This negative diagonal structure, which is a clear deviation from multiplicative separability, produces a preference for far depths (Fig. 4*F*,*I*, open circles).

This new observation of negative diagonal structure in the joint tuning suggests an alternative explanation for MP-based depth tuning: some MT neurons might be tuned to velocity in head coordinates (the sum of retinal and eye velocities), instead of being gain-modulated by eye velocity. To better understand the link between this nonseparable joint velocity tuning and head-centered velocity coding, consider a scenario in which a stationary observer is pursuing a moving target by eye rotation, while another object is moving in the world (Fig. 5*A*). In this case, the retinal velocity of a moving object reflects both the object motion relative to the world and the image motion induced by the pursuit eye movement. Object velocity in head-centered coordinates can be obtained by simply adding the eye velocity to the retinal velocity. As a result, each negative diagonal line in the retinal velocity versus eye velocity plane corresponds to a given head-centered velocity (Fig. 5*B*). Therefore, a joint tuning velocity tuning profile that follows a negative diagonal (e.g., Fig. 4*E*,*H*) implies velocity tuning in head coordinates.

Emergent selectivity for depth from head-centered velocity tuning

To examine how depth selectivity could emerge from neurons tuned to motion in head coordinates, we simulated neurons with log-Gaussian speed tuning in a space ranging from retinal to head coordinates, and we calculated depth tuning curves from these model neuron responses.

For simulated neurons that prefer fast speeds (Fig. 5*C*–*F*), the joint velocity tuning profile in the MP condition produces a broad diagonal band of response (Fig. 5*D*,*E*). Over the range of retinal and eye velocities sampled in the experiments analyzed here (Fig. 5*C*–*E*, lower color intensity shows regions of the velocity space that were not sampled in the experiment), the predicted

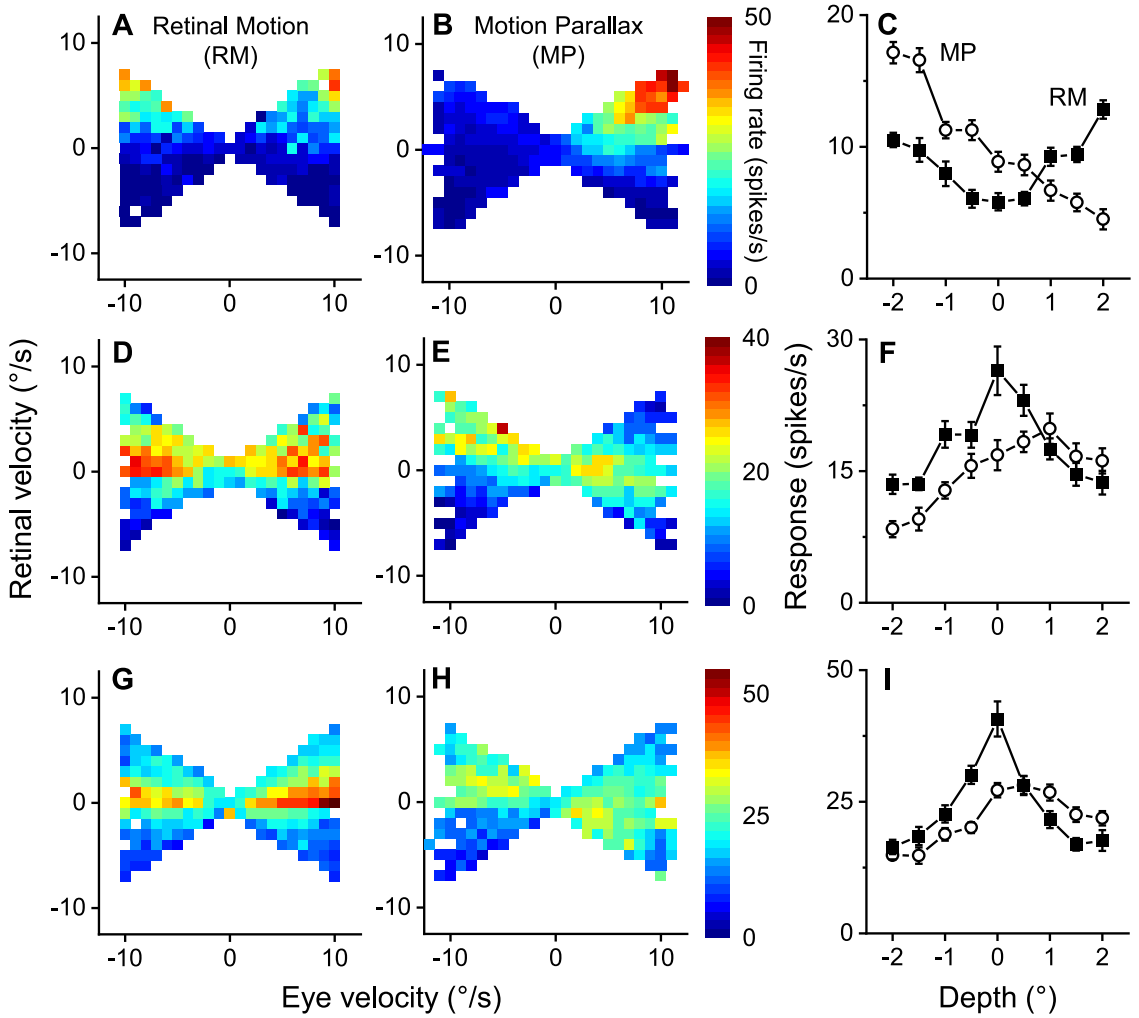

Figure 4. Examples of joint velocity tuning profiles and depth tuning curves of MT neurons. Each row corresponds to one example neuron. ***A***, Joint tuning of an MT neuron, which prefers fast speeds, for retinal and eye velocities in the RM condition. ***B***, Joint tuning of the same neuron in the MP condition. The response to image motion is suppressed for negative eye velocities (top left), and facilitated for positive eye velocities (top right). ***C***, Depth tuning curve for both the RM (solid squares) and MP (open circles) conditions. In the RM condition, the depth tuning curve is symmetric around zero, showing no depth-sign preference. In the MP condition, the firing rate is higher for near than far depths. ***D***, ***G***, Joint tuning in the RM condition for two example neurons with slow-speed preferences. ***E***, ***H***, Joint tuning of the same two slow-speed-preferring neurons in the MP condition, revealing a diagonal structure which indicates that the preferred retinal velocity changes with eye velocity. ***F***, ***I***, Depth tuning curves for the two slow-speed neurons. Depth tuning in the RM condition has a symmetric peak around zero depth (solid squares). In the MP condition, the diagonal structure of joint velocity tuning leads to a far preference for depth (open circles). Error bars indicate ±1 SD.

result resembles data observed for many of the recorded MT cells (compare Fig. 5*D*,*E* with Fig. 4*B*), and the corresponding depth tuning curve has a near preference (compare Fig. 5*F* with Fig. 4*C*). Notably, both low and high weights on eye velocity (ω = 0.25 and 0.75, respectively) can produce depth selectivity (Fig. 5*F*, blue and orange dashed lines). Consequently, model neurons with tuning shifted toward head coordinates and fast speed preferences produce tuning patterns that are difficult to distinguish from the multiplicatively separable profiles expected from a gain mechanism, given the limited range of the data available.

For model neurons with slow-speed preferences (Fig. 5*K*–*N*), the negative diagonal structure in the joint velocity tuning profile is evident within the range of the measured data (e.g., Fig. 5*L*). This structure produces depth tuning with a far preference (Fig. 5*N*, dashed lines), even for relatively small shifts toward head coordinates (low weights on eye velocity). When model neurons have an intermediate speed preference (Fig. 5*G*–*J*), the preferred depth transitions from near to far as the weight on eye velocity increases (Fig. 5*J*, dashed lines). These results indicate that velocity tuning that is even partially transformed toward head coordinates can predict a wide variety of depth tuning patterns from MP, qualitatively similar to those observed in area MT (Fig. 4).

To examine the quantitative relationships between shifts toward head coordinates and depth-sign selectivity, we sampled the parameters in our model within a biologically plausible range, and we systematically varied the weight on eye velocity to control the degree of shift in reference frame toward head coordinates. We generated Poisson spikes from the simulated neurons, and used them to compute DSDI (for details, see Materials and Methods) as a quantitative measurement of depth-sign selectivity, as used previously in empirical studies (e.g., Nadler et al., 2008, 2009). Neurons that prefer near depths have negative DSDI values, and far-preferring neurons have positive DSDI values. Surprisingly, significant depth-sign selectivity ($p < 0.05$, permutation test; Fig. 6*A*) begins to emerge when the weight on eye velocity is <0.1, and this selectivity saturates when the weight on eye velocity is ~0.2, indicating that robust depth tuning can arise from even a modest shift toward head-centered velocity tuning. The intuition for this high sensitivity to the weight on eye

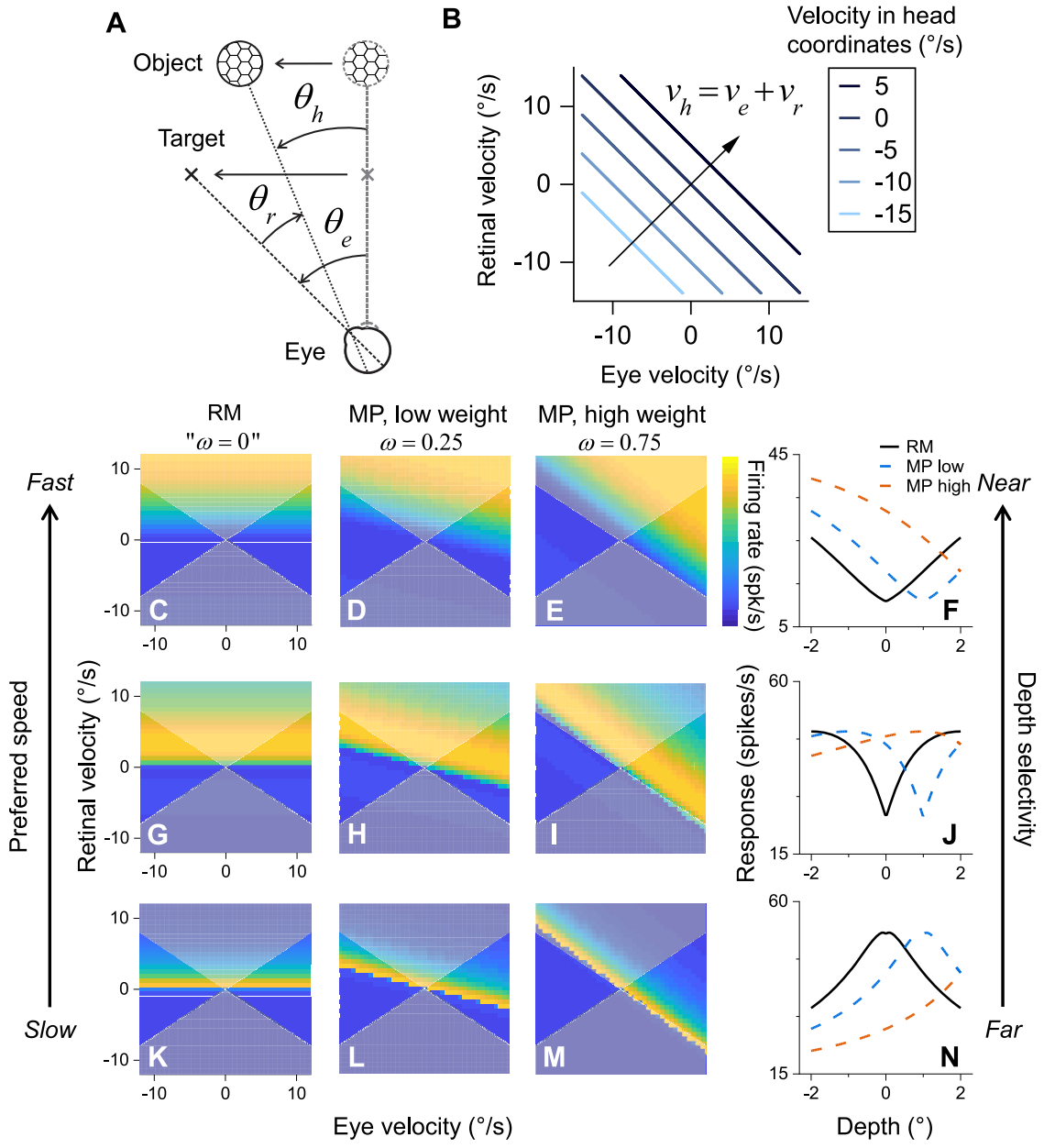

Figure 5. Velocity tuning in head coordinates and predicted depth selectivity. ***A***, Geometry of object motion relative to the head. While a stationary observer makes smooth pursuit eye movements to track a moving target (denoted by ×), an object (soccer ball shape) also moves independently to the left. The velocity of the object in head-centered coordinates, $v_h \equiv d\theta_h/dt$, is the vector sum of its retinal velocity $v_r \equiv d\theta_r/dt$ and the observer's eye velocity $v_e \equiv d\theta_e/dt$. ***B***, Head-centered velocity in the joint domain of retinal and eye velocities. Each diagonal line indicates a locus of retinal and eye velocities that correspond to a particular velocity in head coordinates. ***C***, The joint velocity tuning in the RM condition for a simulated neuron with a high-speed preference of 15°/s. White lines indicate the range over which data were collected in the experiments. Portions of the velocity space that were not sampled in experiments are shown in lower intensity. ***D***, ***E***, Joint velocity tuning of the same simulated neuron in the MP condition for low ($\omega = 0.25$) and high ($\omega = 0.75$) weights on eye velocity. ***F***, Depth tuning curves in the RM and MP conditions for the same simulated neuron. Selectivity for depth-sign emerges in the MP condition for both low and high weights on eye velocity (blue and orange dashed curves), but not in the RM condition (solid black curve). ***G–J***, Joint velocity tuning and depth tuning curves for a model neuron with a moderate speed preference (3.3°/s). Format the same as in ***C–F***. ***K–N***, Joint velocity tuning and depth tuning for a model neuron that prefers very slow speeds (0.16°/s preference). For this unit, low and high weights on eye velocity both yield a far preference for depth (***N***, dashed curves). Format the same as in ***C–F***.

velocity is that DSDI is a measure of the asymmetry of the integrated neural response along lines that have positive and negative slopes in the joint velocity tuning domain (Fig. 1*E*). Thus, even a slight tilt of the joint tuning function away from a horizontal band will create a large asymmetry between the integrated responses for positive and negative slopes in model neurons with head-centered velocity tuning.

Our simulation also reveals a striking relationship between preferred speed and depth-sign preference, as suggested by the example units of Figure 5. To further investigate this relationship, we simulated a population of neurons with preferred speeds that are equally spaced on a logarithmic scale from 0.1 to 30°/s. We find a strong relationship between the preferred speed in head coordinates and preferred depth sign (Fig. 6*B*), such that neurons with slow-speed preferences tend to prefer far depths (Spearman's $r = -0.748$, $p = 8.47 \times 10^{-91}$). A very similar pattern was reported in the empirical data of Nadler et al. (2008) (Spearman's $r = -0.561$, $p = 3.06 \times 10^{-13}$; Fig. 6*C*), which

previously had no clear explanation (by design, near and far stimuli with the same magnitude of equivalent disparity have identical speeds in our experiments). This pattern arises naturally from a head-centered representation for the following reasons. First, a shift toward head-centered velocity tuning always entails a diagonal band of response that slopes downward from left to right (Fig. 5*B*). When the speed preference of the neuron is slow in head coordinates (e.g., Fig. 5*L*), this entire diagonal band lies within the range of tested retinal and eye velocities (saturated bow-tie shape in Fig. 5*C–M*), leading to a preference for a far depth (compare with the slopes in Fig. 1*E*). When the speed preference is fast in head coordinates (e.g., Fig. 5*D*), then only the right portion of the diagonal band of responses lies within the tested range of retinal and eye velocities. In this case, within the range of joint velocities tested experimentally, there is more total response along positively sloped diagonal lines corresponding to near depths (Fig. 1*E*) than along negatively sloped diagonal lines corresponding to far depths. This interaction between response profile of the HT model neurons and the tested range of velocities accounts for the empirically observed relationship between depth-sign preference and speed preference (Fig. 6*C*).

To examine whether this relationship between speed and depth-sign preferences could also be predicted by a multiplicative or additive modulations, we simulated these mechanisms based on Equations 10 and 12 using the same range of parameters. We find no relationship between preferred speed and DSDI in either case (Spearman's $r = -0.035$, $p = 0.436$ for gain modulation; Spearman's $r = 0.076$, $p = 0.089$ for offset modulation; Fig. 7). Thus, we show that this systematic dependence of DSDI on speed preference only emerges naturally from a representation that is partially transformed toward head-centered velocity tuning, which suggests that this mechanism might underlie the depth-sign selectivity described previously.

We further explored whether a neural population with head-centered velocity tuning could be used to accurately estimate depth from MP. Poisson spikes were generated from neurons with velocity tuning that was shifted toward head coordinates, with the preferred speeds and weights on eye velocity sampled from a logarithmic space having ranges of [0.1, 30] and [0.001, 1], respectively. Other parameters, including tuning width, response amplitude, and baseline firing rate, were randomly sampled from ranges that are constrained by empirical data (for details, see Materials and Methods). We trained a simple linear decoder on the population response for 500 simulated trials to recover the magnitude and sign of depth. The performance of the decoder was then evaluated by predicting depth from the responses in another 500 test trials. The linear decoder successfully estimated depth in the MP condition, while it performed very poorly in the RM condition (Fig. 6*D*). The decoder also failed when model responses for the MP condition were shuffled across combinations of eye velocity and retinal velocity (Fig. 6*D*, purple dashed line). These results demonstrate that a population of neurons with tuning that is shifted toward head-centered velocity can reliably represent depth from MP.

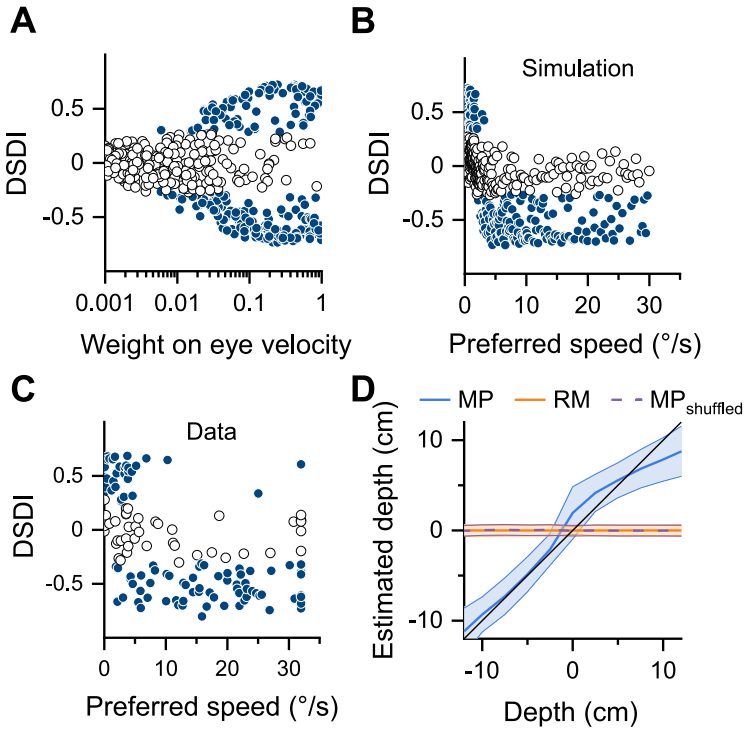

Figure 6. Summary of depth selectivity arising from head-centered velocity tuning: comparison of simulations and data. ***A***, Relationship between DSDI and the weight on eye velocity for a population of simulated neurons based on the HT model. Significant DSDI values ($p < 0.05$, permutation test, filled blue circles) begin to emerge when the weight on eye velocity is >0.01, and DSDI values saturate around a weight of 0.2. Open circles represent nonsignificant DSDI values. ***B***, Simulated DSDI is negatively correlated with preferred speed in head coordinates (Spearman's $r = -0.748$, $p = 8.47 \times 10^{-91}$). Format the same as in ***A***. ***C***, A similar negative correlation between DSDI values and preferred speeds in the empirical data reported by Nadler et al. (2008) (Spearman's $r = -0.561$, $p = 3.06 \times 10^{-13}$). ***D***, Decoding performance based on simulated population responses for the RM and MP conditions. The decoder successfully estimates depth from MP responses (blue curve) but fails when trained on responses from the RM condition (orange curve). Dashed purple curve indicates results for a control in which responses to the MP condition are shuffled across trials. Shaded areas represent ±1 SD around the mean.

Modeling neural interactions between retinal velocity and eye velocity

While head-centered velocity tuning could account for the empirical findings of depth selectivity, it remains unclear whether the data generally support this hypothesis over the previous theory of gain modulation (Kim et al., 2017). To quantitatively compare these hypotheses across a population of MT neurons, we evaluated models that incorporate three components that describe different types of interactions between eye velocity and retinal velocity: a shift toward HT, a multiplicative GM, and an additive OM (Fig. 8). As described above, the HT model incorporates a weighted sum of retinal and eye velocities as the input to the velocity tuning function, $f(\cdot)$ (Fig. 8*A*). By contrast, the tuning function, $f(\cdot)$, in both GM and OM models only takes retinal velocity as input, while its output is modulated by eye velocity in either a multiplicative or additive manner (Fig. 8*B*,*C*). These three components are not mutually exclusive and were also combined together in different ways. Indeed, we considered all eight combinations of these three model components, including a control model in which there is no influence of eye velocity, three models with only one interaction component (HT, GM, and OM), a full model with all three components (Full), and three models with one of the components removed (–HT, –GM, and –OM). The velocity tuning function in each model was first fit to speed and direction tuning measurements from each

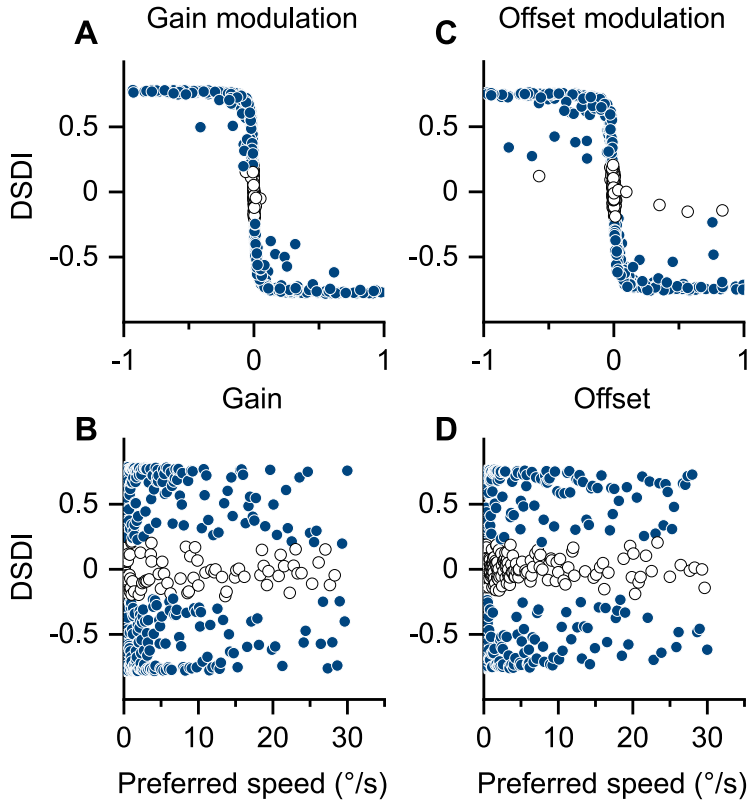

Figure 7. GM and OM models do not predict a dependence of DSDI on speed preference. ***A***, DSDI values vary systematically with the amount of multiplicative gain applied to simulated responses. Positive (negative) gain values correspond to greater responses for positive (negative) eye velocities. Format the same as in Figure 6*A*. ***B***, DSDI values of gain-modulated model neurons do not depend on preferred speed. Format the same as in Figure 6*B*. ***C***, DSDI as a function of additive offset. Positive (negative) offset values correspond to greater responses for positive (negative) eye velocities. ***D***, Preferred speed of offset-modulated model neurons also does not predict DSDI values.

neuron (e.g., Fig. 9*A*) to determine parameter bounds, and then each model was fit to spike trains from the main experimental conditions (RM and MP conditions). We then used the estimated parameters of each model to predict depth tuning curves and DSDI values. Our examination of all combinations of model components allows us to assess the commonalities and uniqueness of each hypothesized mechanism.

Figure 9*A–J* shows the fitting results for a typical near-preferring neuron with a relatively high-speed preference. We found that all models fit to spike trains captured the temporal structure of response histograms reasonably well, and all models predicted the depth tuning reasonably well (Fig. 9*B*,*C*). To quantify the uniqueness of each model component, we computed the change of the BIC (ΔBIC) in two different ways: (1) by adding each model component to the Control model (Fig. 9*D*, gray bars); and (2) by removing each component from the Full model (Fig. 9*D*, white bars). Positive ΔBIC values indicate a reduction in goodness of fit, whereas negative values reflect improvements in fit quality. For this type of neuron, adding any of the three components to the Control model greatly improved the goodness of fit of spike trains, whereas removing any individual component from the Full model showed little change in BIC (Fig. 9*D*). This indicates that each of the three model components (HT, GM, and OM) is largely capable of capturing the data for this neuron. Analysis of the model predictions of DSDI reveals a similar result: all models except the Control model predicted the neuron's depth-sign selectivity pretty accurately (Fig. 9*E*). One can appreciate why all of the models fit the data well by examining the raw data and fits in the joint velocity space (Fig. 9*F–J*). Because the speed preference of this neuron is relatively high, the empirical data are missing in the region of the joint velocity space (high retinal speeds and low eye speeds) that would best distinguish among models, especially HT (Fig. 9*G*) versus GM (Fig. 9*H*). Thus, HT, GM, and OM models are equally likely to explain this type of joint tuning pattern given the limitations of the empirical data.

In contrast, model fits to responses of neurons with a clear diagonal structure in their joint velocity profile show a different pattern of results (Fig. 9*K–T*). For these cells, the HT model clearly outperformed the GM and OM models both in fitting spike trains and in predicting depth tuning curves (Fig. 9*L–O*). In addition, the HT model uniquely accounted for the diagonal structure in the joint velocity tuning (Fig. 9*P–T*), such that removing the HT component resulted in a large increase in BIC and DSDI prediction error (Fig. 9*N*,*O*). This result further supports our hypothesis that some neurons in MT have velocity tuning that is shifted toward head-centered coordinates, and that this shift produces a far preference for depth defined by MP.

We compared the models at the population level based on both ΔBIC values and DSDI prediction errors. When adding one of the model components (HT, GM, or OM) to the Control model, a negative ΔBIC indicates that a better fit is obtained when the model component is added. In general, ΔBIC was negative when a component was added to the Control model, with significant variation across components ($\chi^2_{(3)} = 307.02$, $p = 3.01 \times 10^{-66}$, Kendall's W = 0.463, Friedman's test; Fig. 10*A*). Across the population, the ΔBIC value obtained by adding either the HT or OM component is significantly more negative than that obtained by adding the GM component ($Z = -9.306$, $p = 1.33 \times 10^{-20}$ for HT-GM; $Z = -9.157$, $p = 5.34 \times 10^{-20}$ for OM-GM; Wilcoxon signed-rank test; Bonferroni-corrected $\alpha = 0.0167$; Fig. 10*A*), and no significant difference in ΔBIC was found between the HT and OM models ($Z = 1.850$, $p = 0.0643$, Wilcoxon signed-rank test). The relative contribution of each model component to ΔBIC for individual neurons is visualized in Figure 10*B*, where each vertex represents a dominant contribution by one component. Most neurons show substantial contributions of the HT and OM components, while fewer cells have a strong contribution of the GM component. As a result, the largest density of neurons lies in the lower-middle region of the triangle (Fig. 10*B*).

When removing a model component from the Full model, a positive ΔBIC indicates that the goodness of fit is impaired by removing the component, indicating a unique contribution of that component to the fit (after accounting for the change in the number of free parameters). Across the population of neurons, there was a significant difference in ΔBIC depending on which model component was removed from the Full model ($\chi^2_{(3)} = 383.37$, $p = 8.84 \times 10^{-83}$, Kendall's W = 0.578, Friedman's test; Fig. 10*A*). We found a significantly greater ΔBIC when removing the HT component compared with GM ($Z = 7.259$, $p = 3.91 \times$

10^{-13}, Wilcoxon signed-rank test; Bonferroni-corrected $\alpha = 0.0167$), and also for OM compared with GM ($Z = 8.066$, $p = 7.29 \times 10^{-16}$, Wilcoxon signed-rank test). In addition, ΔBIC for removing the OM component is slightly greater than that for HT ($Z = 2.738$, $p = 0.0062$, Wilcoxon signed-rank test). Together, the ΔBIC data of Figure 10*A*, *B* indicate comparable contributions of HT and OM components to fitting MT responses, with a somewhat weaker contribution from the GM component.

To examine the contribution of model components to predicting depth-sign selectivity, we quantified their performance as the absolute error between the predicted and measured DSDI values, which we denote |ΔDSDI|. We find that median |ΔDSDI| values differ significantly across models ($\chi^2_{(4)} = 112.46$, $p = 2.18 \times 10^{-23}$, Kendall's W = 0.127, Friedman's test; Fig. 10*C*). We conducted a Wilcoxon signed-rank test for each pair of models and applied a Bonferroni correction, resulting in a significance level $\alpha = 0.005$ (Table 1). All models predicted depth tuning significantly better than the Control model, and the Full model significantly outperformed all single-component models. In addition, we find that both HT and GM models predict depth tuning better than the OM model, while there is no significant difference between HT and GM models (Table 1; Fig. 10*C*). The relative proportions of |ΔDSDI| from each model are shown in Figure 10*D*. There is a higher density of cells around the vertices of GM and HT components compared with OM, indicating that GM and HT components predicted depth selectivity more accurately for some neurons (Fig. 10*D*). Finally, the Full model performed well in predicting the depth selectivity for most MT neurons (Spearman's $r = 0.803$, $p = 3.87 \times 10^{-51}$; Fig. 11*A*, blue dots), which is especially notable given that the model was fit to the spike trains of neurons and was not optimized to predict depth-sign selectivity per se.

Kim et al. (2017) showed that a significant amount of extraretinal modulation observed in the responses of MT neurons can be explained by the direction of eye movements alone, without factoring in the magnitude of eye velocity (i.e., speed). Here, we explored whether a model that considers only the sign of eye velocity (i.e., direction) can perform comparably to the models described above. We fit MT responses with a modified GM model that only takes the sign of eye velocity as extraretinal input (denoted as the "GM-sign" model, see Materials and Methods). While this model significantly outperforms the Control model (which only operates on retinal velocity) in fitting PSTHs ($Z = -8.326$, $p = 8.40 \times 10^{-17}$, Wilcoxon signed rank test on ΔBIC) and in predicting depth-sign selectivity ($Z = -2.937$, $p = 0.003$, Wilcoxon signed rank test on |ΔDSDI|), it is substantially outperformed by both the GM and Full models at fitting response PSTHs (GM vs GM-sign: $Z = 3.228$, $p = 0.001$, Wilcoxon signed rank test; Full vs GM-sign: $Z = 4.22$, $p = 2.393 \times 10^{-5}$) and in predicting depth-sign selectivity (GM vs GM-sign: $Z = 6.261$, $p = 3.825 \times 10^{-10}$, Wilcoxon signed rank test; Full vs GM-sign: $Z = 7.206$, $p = 5.747 \times 10^{-13}$). This demonstrates that both the direction and speed of eye velocity make significant contributions to MT responses, and that just the direction of eye movement is not sufficient to predict depth-sign selectivity.

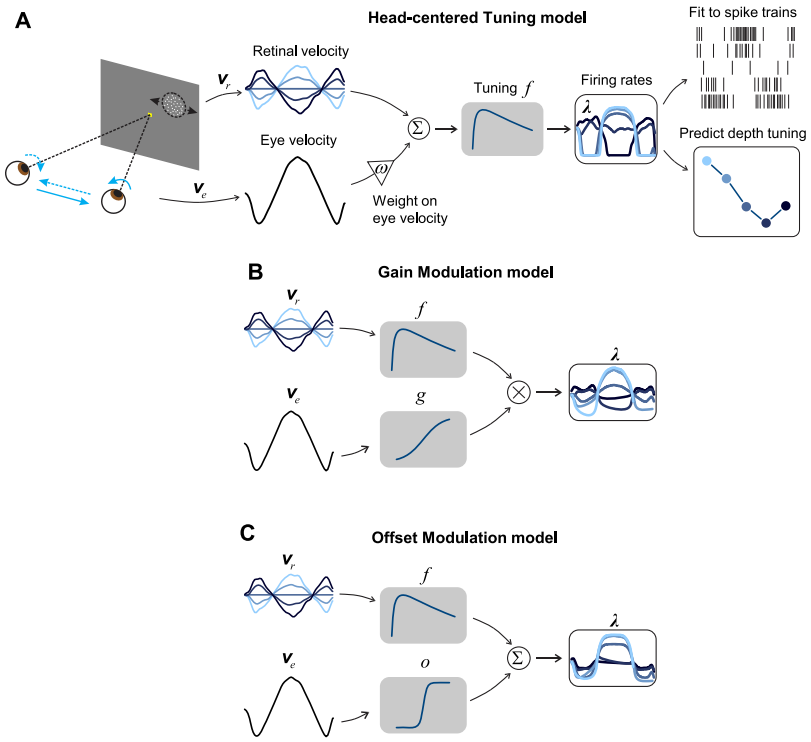

Figure 8. Schematic illustration of computational models. ***A***, Approach to model fitting illustrated for the head-centered velocity tuning (HT) model. Retinal and eye velocities, $\mathbf{v}_r$ and $\mathbf{v}_e$, are linearly combined with a weight ω on eye velocity, and fed into a velocity tuning function, f (for details, see Materials and Methods). The output firing rates, $\boldsymbol{\lambda}$, are then fit to spike trains to estimate model parameters. After fitting, the modeled response is used to predict the neuron's depth tuning curve. ***B***, GM model. Only retinal velocity is fed into the velocity tuning function, while eye velocity multiplicatively modulates the tuning response through a gain function, g. ***C***, OM model is similar to the GM model, except that eye velocity modulates the response in an additive fashion by an offset function, o. All models are fit to spike trains and then used to predict depth tuning curves. In addition to the single-component models shown here, we also fit models comprised of pairs of these components, as well as a Full model embodying all three components (for details, see Materials and Methods).

During the fitting of the HT component, we constrained the weight on eye velocity to the range from −1 to 1. As shown in Equation 4, this parameter reflects the shift of tuning toward head coordinates when it has a positive value. Indeed, we found that most of the estimated weights are biased toward positive values (median = 0.122, $Z = 7.830$, $p = 2.44 \times 10^{-15}$, one-tailed Wilcoxon signed-rank test; Fig. 11*B*, blue curve), indicating that this weight term is generally meaningful for representing head-centered velocity, rather than overfitting an arbitrary data pattern.

Modulation by simulated eye velocity from large-field background motion

In this section, we examine whether the mechanisms of interaction between retinal velocity and eye velocity are similar when eye movements are simulated by large-field visual background motion while the eye remains stationary. Specifically, we analyzed data from the DP condition (Fig. 1*C*) in the same way as we did for the MP condition.

Figure 12 shows data from an example neuron and the corresponding model fits. This example neuron exhibits a clear additive contribution of eye velocity (evidenced as a vertical stripe-like pattern on the left side of the joint velocity tuning profiles; Fig. 12*F*, *K*), which is captured by models with an OM component. As a consequence, the OM component contributes most to ΔBIC values in both the MP and DP conditions (Fig. 12*D*,*E*).

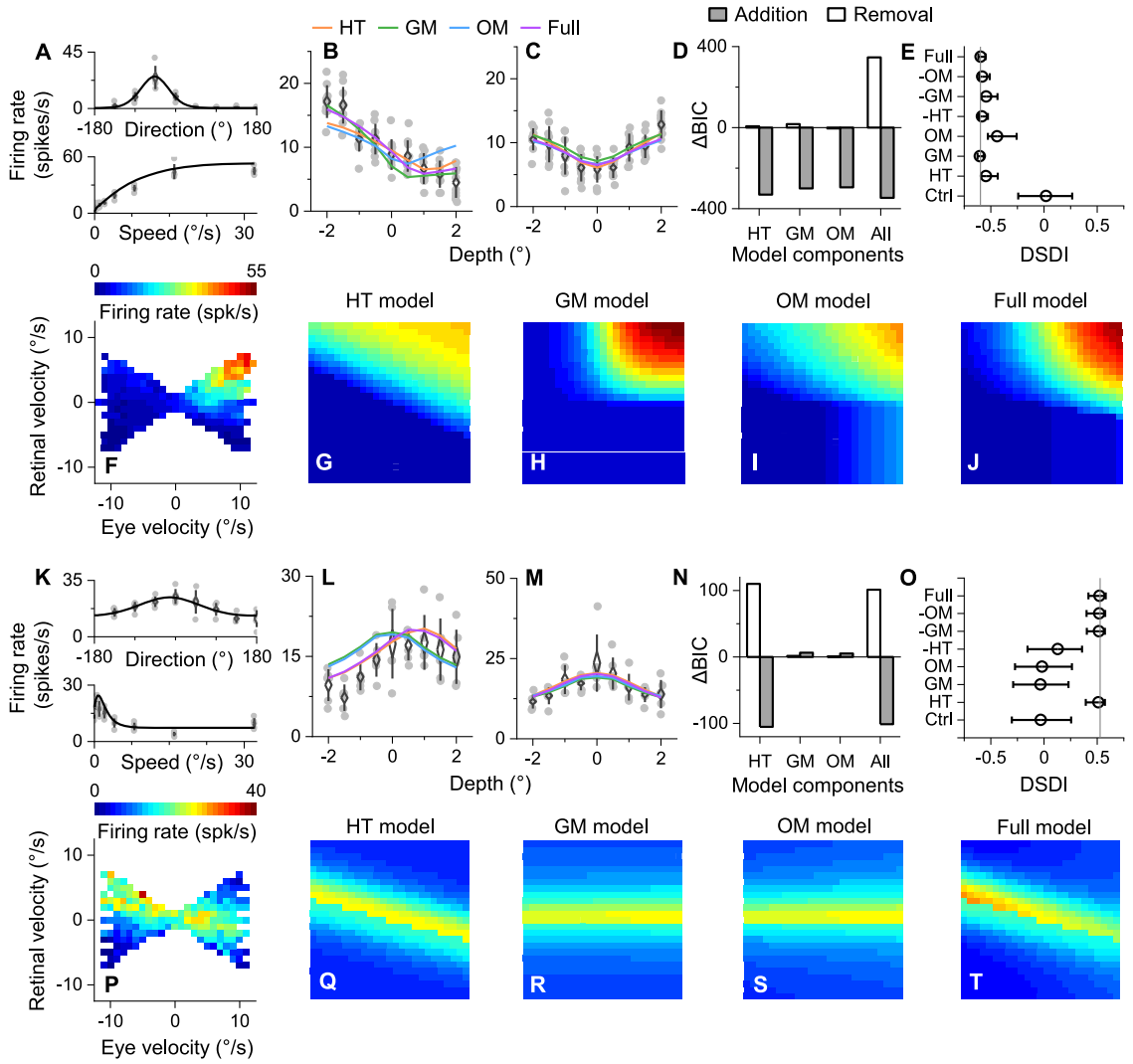

Figure 9. Model fitting results for two example MT neurons. ***A–J***, Model fits to data for a neuron that prefers fast speeds (same neuron as in Fig. 4*A–C*). ***A***, Retinal direction (top) and speed (bottom) tuning functions were fit to data from separate tuning measurements. Gray dots indicate mean firing rates for each trial. Black curve indicates the best fit of the model. Error bars indicate ±1 SD around the mean of the data. ***B***, ***C***, Depth tuning data for both MP (***B***) and RM (***C***) conditions, along with the corresponding model predictions (orange curve, HT model; green curve, GM model; blue curve, OM model; purple curve, Full model). Gray dots represent the mean firing rate for each trial. Open circles represent the average firing rates across repetitions. Error bars indicate ±1 SD. ***D***, Change of BIC after adding each component to the Control model (gray bars) or removing them from the Full model (white bars). Negative change of BIC indicates an improved model fit (accounting for degrees of freedom). ***E***, Open circles represent DSDI values predicted by each model, with error bars indicating 95% CIs estimated by bootstrapping. Vertical line indicates the measured DSDI in the MP condition. ***F***, Joint velocity tuning profile in the MP condition. Firing rate is color coded from low (blue) to high (red). ***G–J***, Model predictions of joint velocity tuning for the MP condition (***G***, HT model; ***H***, GM model; ***I***, OM model; ***J***, Full model). Color coding as in ***F***. ***K–T***, Model fitting results for a neuron that prefers slow speeds and far depths (the same neuron as shown in Fig. 4*D-F*). Format the same as in ***A–J***.

However, the OM component alone could not explain the observed depth tuning curves (Fig. 12*A*,*B*). Only models with both OM and one other component (either HT or GM) were able to correctly predict depth-sign selectivity (Full, –HT, and –GM models in Fig. 12*C*). On the other hand, data from a second example neuron depicted in Figure 13 show a more gain-like pattern of modulation in both MP and DP conditions, with no clear signatures that would suggest the necessity of either an OM or HT component (Fig. 13). Thus, all three model components fit the data similarly well, and are mostly interchangeable (Fig. 13*A–E*). For both example neurons, the joint velocity tuning and depth tuning are quite similar between the MP and DP conditions, consistent with previous findings (Kim et al., 2015). However, we were not able to identify any neurons with a clear negative diagonal structure in the joint velocity tuning profile for the DP condition. This might be because fewer neurons were found to prefer far depths and slow speeds in this dataset, or because of a difference between real and simulated eye movements (see Discussion).

In general, population results for the DP condition (Fig. 10*E–H*) are similar to those for the MP condition (Fig. 10*A–D*). We found a significant dependence of ΔBIC on which model component was added or removed ($\chi^2_{(3)} = 40.11$, $p = 1.01 \times 10^{-8}$, Kendall's W = 0.138 for component addition; $\chi^2_{(3)} = 46.14$, $p = 5.30 \times 10^{-10}$, Kendall's W = 0.159 for component removal; Friedman's test; Fig. 10*E*,*F*). ΔBIC values were significantly different between the HT and GM components ($Z = -4.939$, $p = 7.87 \times 10^{-7}$ for addition; $Z = 4.546$, $p = 5.46 \times 10^{-6}$ for removal; Wilcoxon signed-tank test; Bonferroni-corrected α = 0.0167), but not between other pairs of components (Z =

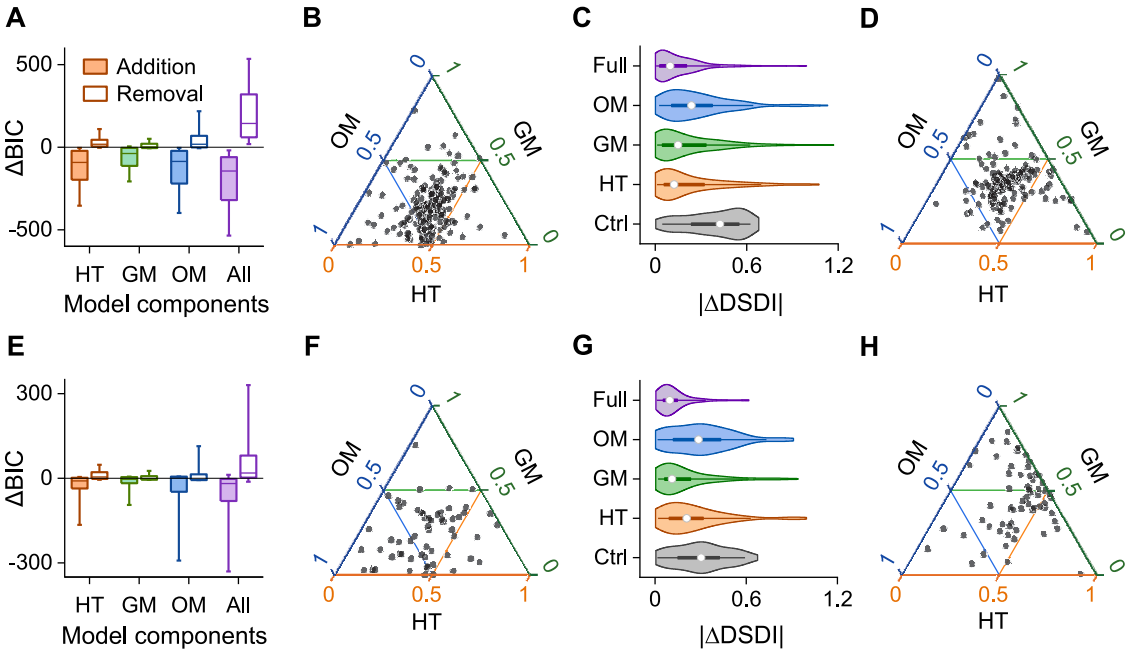

Figure 10. Summary of model fitting results for the MP and DP conditions. ***A–D***, Summary of results for the MP condition. ***A***, Change of BIC by addition (filled bars) and removal (open bars) of model components for all neurons. Horizontal lines indicate median values. Boxes represent interquartile ranges. Error bars indicate 90% CIs. ***B***, Ternary plot showing the proportion of |ΔBIC| produced by adding one of the three model components to the Control model for the MP condition. Because greater |ΔBIC| values correspond to a better fit relative to the Control model, dots closer to one of the vertices show a greater contribution of that model component to the fit. ***C***, Distributions of differences in DSDI, |ΔDSDI|, between model predictions and data. Open circles represent median values. Thick black lines indicate interquartile ranges. Whiskers represent 95% CIs. Smaller |ΔDSDI| values indicate better prediction performance. ***D***, Proportion of the improvement of |ΔDSDI| produced by adding one of the three model components to the Control model in the MP condition, $|\Delta DSDI|_{control} - |\Delta DSDI|_{model}$. Because greater improvement of |ΔDSDI| values corresponds to a better model prediction of depth tuning curves relative to the Control model, dots located closer to one of the vertices indicate that model component predicted DSDI more accurately. ***E–H***, Summary of fitting results for the DP condition. Format the same as in ***A–D***.

Table 1. Results of Wilcoxon signed-rank tests comparing |ΔDSDI| values for pairs of models in the MP condition[a]

Model pairs	*Z*	*P*
HT – Ctrl	−6.948	3.72×10^{-12}*
GM – Ctrl	−6.861	6.82×10^{-12}*
OM – Ctrl	−6.145	8.01×10^{-10}*
Full – Ctrl	−10.013	1.34×10^{-23}*
HT – Full	4.249	2.15×10^{-5}*
GM – Full	4.126	3.69×10^{-5}*
OM – Full	7.254	4.04×10^{-13}*
HT – GM	0.988	0.323
HT – OM	−3.701	0.0002*
GM – OM	−4.694	2.69×10^{-6}*

[a]Significance level after Bonferroni correction is 0.005.
*Statistical significance at the Bonferroni-corrected level.

−1.355, $p = 0.176$ for additions of HT and OM; $Z = 1.589$, $p = 0.112$ for removals of HT and OM; $Z = 1.204$, $p = 0.229$ for additions of GM and OM; $Z = -0.272$, $p = 0.786$ for removals of GM and OM; Wilcoxon signed-tank test).

Significant differences between models were also found for |ΔDSDI| values in the DP condition ($\chi^2_{(4)} = 60.08$, $p = 2.79 \times 10^{-12}$, Kendall's W = 0.155, Friedman's test; Fig. 10*G*,*H*). Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons were again performed to compare |ΔDSDI| values between each pair of models (Table 2). In contrast to what we found for the MP condition, only the GM and Full models showed significantly better predictions of DSDI than the Control model, while the Full model significantly outperformed all other single-component models. Our Full model also performed very well in predicting the depth-sign selectivity of MT neurons in the DP condition (Spearman's $r = 0.884$, $p = 4.56 \times 10^{-33}$; Fig. 11*A*, orange squares).

We also compared model performance in predicting DSDIs between the MP and DP conditions. The performance of the Full model, based on |ΔDSDI| values, did not differ significantly between these two conditions ($Z = -0.235$, $p = 0.815$, Wilcoxon rank-sum test). A modestly significant difference between MP and DP conditions was only found for the HT model ($Z = 2.137$, $p = 0.033$, Wilcoxon rank-sum test), but not for the GM or OM models ($Z = -0.289$, $p = 0.773$ for GM; $Z = 1.488$, $p = 0.137$ for OM; Wilcoxon rank-sum test). We also found that the weights on eye velocity in the HT component are closer to zero and less biased toward positive values in the DP condition (median = 0.0083, $Z = 1.979$, $p = 0.024$, one-tailed Wilcoxon signed-rank test for zero medians; $Z = 5.360$, $p = 4.17 \times 10^{-8}$, one-tailed Wilcoxon rank-sum test for a smaller median in DP compared with MP condition; Fig. 11*B*, orange curve). This indicates that shifts toward head-centered velocity tuning are less prominent in the DP condition.

Discussion

Our findings show that depth selectivity from MP in area MT can arise from even a modest shift in velocity tuning from retinal to head coordinates. While the joint velocity tuning of many MT neurons is consistent with the previous suggestion of a gain modulation mechanism (Kim et al., 2017), other neurons with slow speed preferences show a clear shift in retinal speed preference with eye velocity that manifests as a diagonal structure. Our simulations reveal that a range of depth tuning properties can be explained by a modest partial shift of velocity tuning toward head coordinates. We further demonstrate that a population of

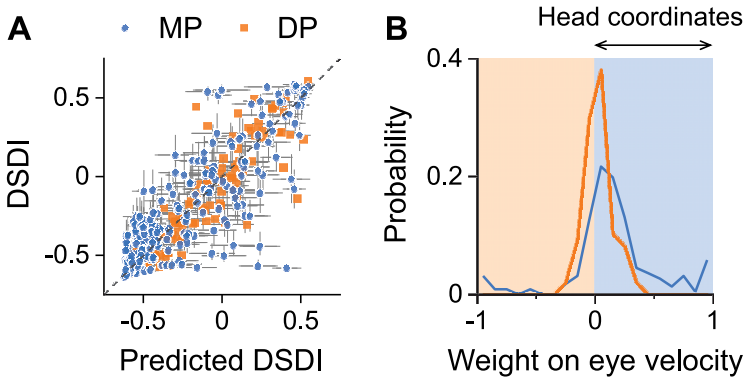

Figure 11. Summary of the Full model performance and distribution of weights on eye velocity. ***A***, Scatter plot showing the relationship between DSDI values predicted by the Full model (*x* axis) and experimentally measured DSDI values (*y* axis). Each dot represents one neuron. Blue dots represent data from the MP condition. Orange squares represent data from the DP condition. Error bars indicate 68% CIs. ***B***, Distribution of the weight on eye velocity in the Full model (blue curve, MP condition; orange curve, DP condition). Positive weights indicate a shift toward representing velocity in head coordinates.

such neurons can be simply decoded to estimate depth, and that the previously observed relationship between depth-sign selectivity and speed preference (Nadler et al., 2008) emerges naturally from a tuning shift toward head coordinates. Our models perform well in fitting spike trains and predicting depth-sign selectivity of MT neurons, and the HT component uniquely contributes to the fitting of a subset of MT neurons. However, the empirical data are not sufficient to distinguish among models for many MT neurons because of the limited range of retinal and eye velocities tested. Together, our results open the possibility that depth-sign selectivity in area MT might arise from a mechanism that is involved in transforming the representation of object motion toward head-centered coordinates during self-motion.

Relationship to previous studies on depth coding from MP

This study extends previous work on MP-based depth coding in the following ways. First, we analyzed raw spike trains at millisecond resolution, whereas previous analyses were generally performed on mean spike rates over a 1 or 2 s time period (Nadler et al., 2008; Kim et al., 2015, 2017). This fine-scale analysis allowed us to examine in detail the response modulations induced by eye velocity, thus providing estimates of joint tuning for retinal velocity and eye velocity. Second, we formalized a GM model based on previous theory, and addressed its strengths and weaknesses in explaining joint velocity tuning patterns. Third, we proposed a new computational model of head-centered speed tuning based on our observations, and compared it with a variety of other model variants in a rigorous manner. Our HT model not only successfully accounts for specific patterns of joint velocity tuning that the GM and OM models fail to explain, but it also naturally predicts a previously unexplained population relationship between preferred speed and depth-sign selectivity across a population of MT neurons (Nadler et al., 2008) that the GM and OM models do not predict.

The HT, GM, and OM models performed similarly well in fitting the responses of many MT neurons, suggesting the possibility of multiple mechanisms that may be somewhat redundant. In Kim et al. (2017), a large portion of MT neurons was found to be gain modulated by the direction of eye movement in a multiplicative fashion. However, that study did not explicitly consider other models nor compare models in a systematic fashion. Indeed, for many neurons, our results show that the HT model can produce joint tuning profiles similar to those predicted by GM, especially given the limited range of the experimental data. These results imply that the gain-modulation effects reported previously by Kim et al. (2017) might also arise from a shift in velocity tuning toward head-centered coordinates. In addition, for a subset of MT neurons with slow-speed preferences, our findings demonstrate clearly that depth tuning arises from a shift toward head-centered velocity tuning, not gain modulation. However, it remains to be determined whether the HT model also accounts for the selectivity of most MT neurons with higher speed preferences. It is also possible that multiple mechanisms coexist in MT and produce similar depth-sign selectivity for MP.

Differences between the MP and DP conditions

In this study, we fit the same set of models to neural responses recorded in both the MP (real eye movement) and DP (simulated eye movement) conditions. While the pattern of results across the population is similar between these two conditions (Fig. 10), we find that weights on eye velocity in the HT model for the DP condition concentrate more strongly around zero and are less biased toward positive values (Fig. 11*B*, orange vs blue curves). This implies that a shift toward a head-centered representation of velocity is less prominent in the DP condition. We speculate that this arises from the different levels of ambiguities of extraretinal signals in the MP and DP conditions. As shown by Nadler et al. (2009), a smooth pursuit command signal, but not a vestibular signal, is used to disambiguate MP-defined depth. When an eye movement signal is the only extraretinal signal in use, the source of retinal image motion is ambiguous: it could be the result of self-motion (Fig. 1*D*) or object motion in the world (Fig. 5*A*). When the scene is interpreted in the latter way, head-centered speed tuning may naturally occur as an effort to infer object motion in the world. However, the large-field visual background motion in the DP condition provides an unambiguous cue to distinguish between these two scenarios. In the DP condition, optic flow simulates a rotation of the scene around the point of fixation; this flow pattern lacks a laminar component of optic flow that would be produced when a stationary observer pursues a moving target. This means that the extraretinal signal in the DP condition is not ambiguous, and the animal can infer the true geometry (self-motion with eye rotation) without misinterpreting the scene. Thus, in the DP condition, a head-centered speed representation may not be required by the computation. Future studies that systematically manipulate subjects' inferences about viewing context will be needed to test this hypothesis.

Comparison with previous work on pursuit eye movements and reference frames of speed tuning in MT

The effect of smooth pursuit on responses of MT neurons has been investigated in a number of previous studies. Komatsu and Wurtz (1988) showed that neurons in the foveal part of MT respond to pursuit eye movements. Later studies reported a sharp decrease in MT responses when the motion of a pursuit target was removed by blinking off or stabilizing the target

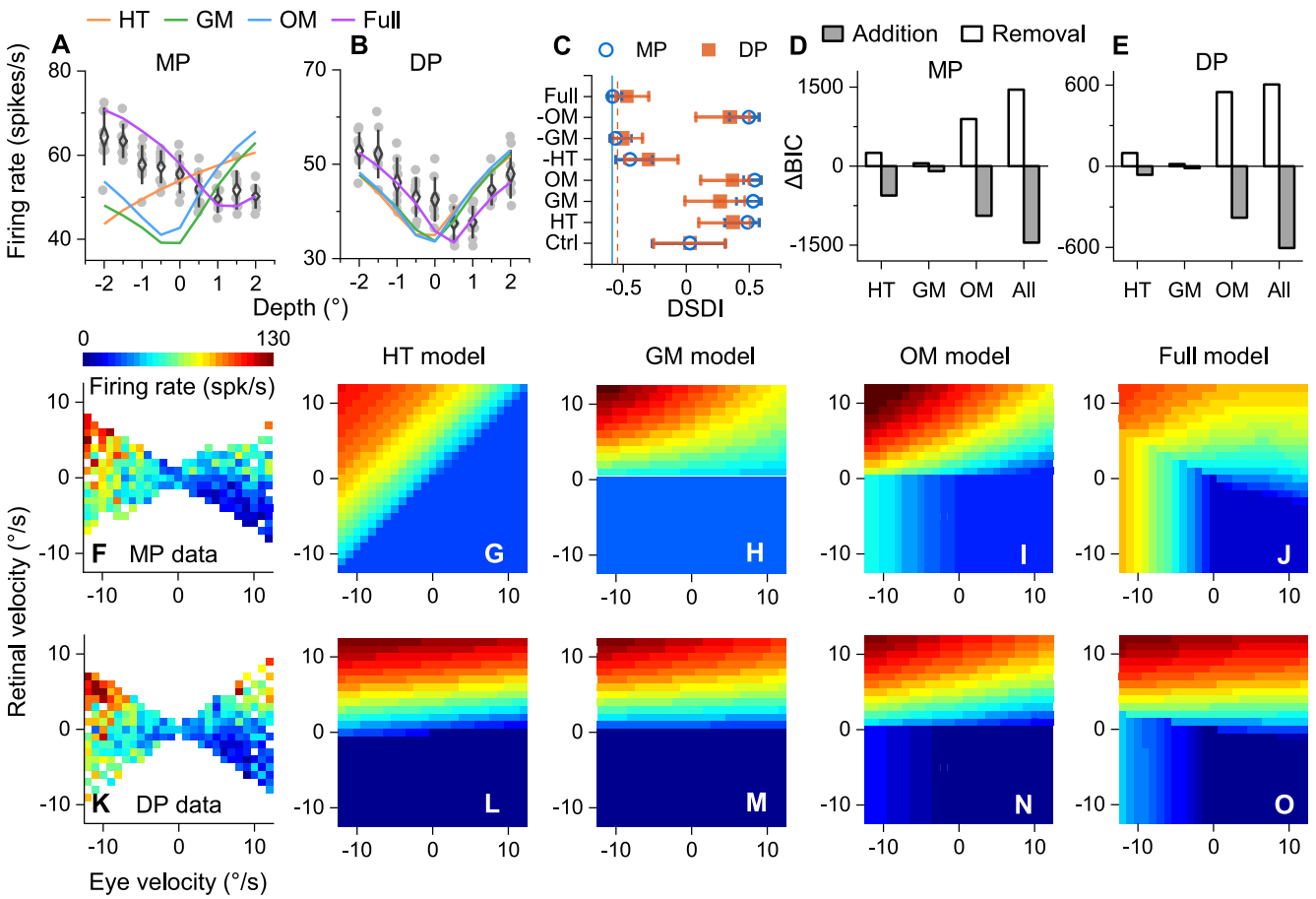

Figure 12. Joint velocity tuning, depth tuning, and model fitting results for an example neuron in the MP and DP conditions. ***A***, ***B***, Depth tuning curves for the MP and DP conditions, respectively. Format the same as in Figure 9*B*, *C*. ***C***, Comparison of measured and predicted DSDI values for the same example neuron in both MP (open blue circles) and DP (solid orange squares) conditions. Format otherwise as in Figure 9*E*. ***D***, ***E***, Change of BIC values associated with addition (gray bars) and removal (white bars) of model components for the MP and DP conditions, respectively. ***F–J***, Joint velocity tuning in the MP condition (***F***) and model fits (***G***, HT model; ***H***, GM model; ***I***, OM model; ***J***, Full model). Format the same as in Figure 9*F–J*. ***K–O***, Joint velocity tuning (***K***) and model fits for the same neuron in the DP condition (***L–O***).

(Newsome et al., 1988), or when the animal tracked an imaginary target (Ilg and Thier, 2003). These results suggested a potential role of foveal MT neurons in encoding retinal slip during pursuit, but extraretinal signals related to pursuit eye movements were not apparent in those experiments. In contrast, we have observed robust modulations of MT responses by smooth pursuit in the viewing geometry of Figure 1*D* (Nadler et al., 2008, 2009; Kim et al., 2017). We reason that the nonlinear interactions between retinal and eye velocity that are captured by our models, combined with the relatively small weights on eye velocity needed to fit the data (Fig. 11*B*), are likely to explain why pursuit-related extraretinal signals were not prominent in traditional experimental paradigms involving a single small pursuit target against a dark background.

Lagae et al. (1993) reported that MT neurons with receptive fields within a few degrees of the fovea tended to have low-pass speed tuning, although they did not report a significant correlation between speed preference and receptive field eccentricity. They speculated that these low-pass cells may be those that were found to be active during pursuit maintenance in foveal MT (Komatsu and Wurtz, 1988). In our study, the majority of the recorded neurons had receptive field eccentricities in the range from 10 to 30 degrees, and we did not observe a statistically significant correlation between eccentricity and preferred speed (Spearman's $r = 0.130$, $p = 0.057$), which is consistent with a previous report involving a larger sample (Palanca and DeAngelis, 2003). Therefore, our dataset likely did not include many foveal MT neurons that may be involved in pursuit maintenance.

Both Chukoskie and Movshon (2009) and Inaba et al. (2011) measured the speed tuning of neurons in MT and MST during fixation and smooth pursuit eye movements. They reported that MT neurons generally represent motion in retinal coordinates, whereas MST neurons are typically tuned to screen coordinates. However, at least some MT neurons in their data reveal a clear response shift toward screen coordinates, which is equivalent to head coordinates since neither the head nor the screen was moving in their experiment. Thus, their data also demonstrated an influence of eye velocity on MT responses. Our simulations demonstrate that even a slight shift toward head coordinates can produce robust selectivity for depth from MP, and our model-fitting results confirm that the shifts toward head coordinates are generally quite small in MT. Thus, while the extent of tuning shifts toward head coordinates in our MT data is compatible with these previous findings, these small shifts appear to be sufficient to account for the depth-sign selectivity of MT neurons.

Limitations of the present study

Since the magnitude of depth from MP is unrealistically large when the iso-depth line is close to vertical (Fig. 1*E*), the empirical data that we analyzed have limited samples from the full joint velocity space, especially the regions corresponding to high retinal speeds and low eye speeds. Our model fitting results reveal that

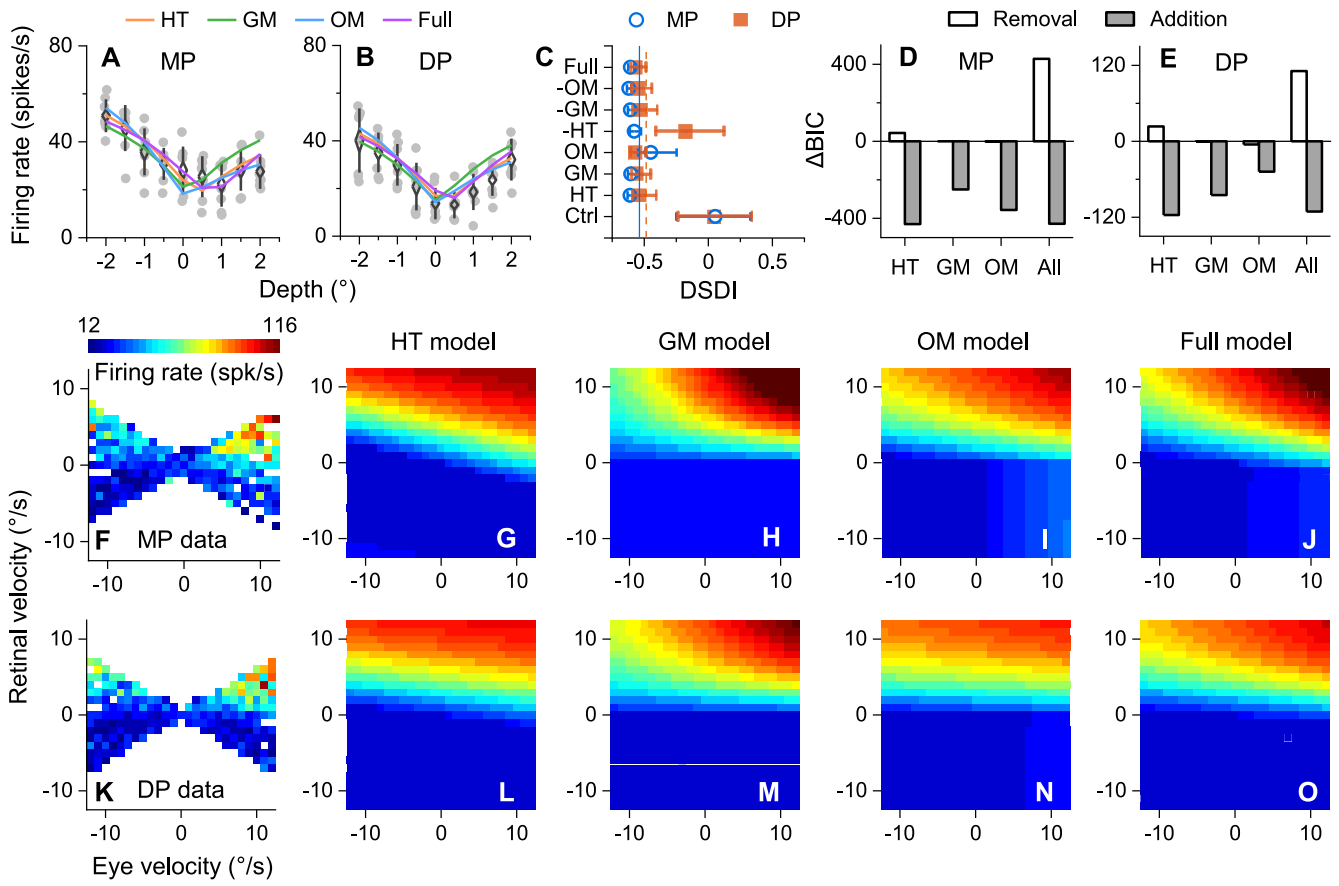

Figure 13. Model fits to the DP and MP conditions for an additional example neuron. Format the same as in Figure 12. ***A***, ***B***, Depth tuning curves and model predictions for the MP and DP conditions, respectively. ***C***, Model predictions of DSDI values for the MP (open blue circles) and DP (filled orange squares) conditions. ***D***, ***E***, Change of BIC by addition (gray bars) and removal (white bars) of each model component for the MP (***D***) and DP (***E***) conditions, respectively. ***F–J***, Joint velocity tuning profile (***F***) and model predictions (***G–J***) for the MP condition. ***K–O***, Corresponding results for the DP condition.

Table 2. Results of Wilcoxon signed-rank tests comparing |ΔDSDI| values for pairs of models in the DP condition[a]

Model pairs	*Z*	*P*
HT – Ctrl	−2.747	0.0060
GM – Ctrl	−4.313	1.61×10^{-5}*
OM – Ctrl	−1.148	0.251
Full – Ctrl	−6.633	3.28×10^{-11}*
HT – Full	5.439	5.37×10^{-8}*
GM – Full	3.024	0.00249*
OM – Full	6.399	1.56×10^{-10}*
HT – GM	3.891	9.96×10^{-5}*
HT – OM	−2.909	0.00362*
GM – OM	−5.108	3.26×10^{-7}*

[a]Significance level after Bonferroni correction is 0.005.
*Statistical significance at the Bonferroni-corrected level.

differences among the HT, GM, and OM models are subtle for many neurons over the range of available data. Therefore, a more extensive sampling of the joint tuning for retinal and eye velocities will be necessary to distinguish these different mechanisms for many MT neurons. Experiments that are not constrained by the geometry of MP (Fig. 1*D*,*E*) may be useful for measuring the interaction between retinal and eye velocities across much wider ranges. Our models were fit to spike trains without accounting for delays or dynamics in the neural responses. We explored modeling the temporal dynamics of neural responses but found that it made little difference to the fits since the temporal structure of our stimuli is slow (0.5 Hz movement). From a system identification perspective, uncorrelated white noise signals may be better suited for measuring MT responses to a larger range of retinal and eye velocities (Brostek et al., 2015), as well as for building a model that fully describes the bottom-up mapping from velocity signals to neural responses in MT.

We observed that a small proportion of the recorded neurons (<5%) showed joint velocity tuning profiles that were dominated by eye velocity, and this could be observed for both the DP and MP conditions (Fig. 14*A* and Fig. 14*F*, respectively). While these neurons show clear retinal speed tuning during a fixation task, this retinal velocity contribution to the joint tuning is overwhelmed by the effect of (real or simulated) eye velocity for this small subset of neurons. None of our model components could explain this observed pattern adequately (Fig. 14*B–E*,*G–J*), and we speculate that a normalization mechanism might be needed to account for these data, as previous work has shown that normalization can approximate a winner-take-all interaction when inputs are imbalanced (Busse et al., 2009).

We have only explored the neural interactions between retinal velocity and eye velocity in the context of an observer who translates laterally while maintaining fixation at a world-fixed target (Fig. 1*D*). We have suggested that this viewing context could potentially be misinterpreted as a coordinate transform in which a stationary observer tries to infer object motion in the world by compensating for their eye movements (Fig. 5*A*). However, it is unclear how exactly MT neurons jointly encode retinal and eye velocities when directly tested in the CT context. While we showed that some neurons might represent velocity in head

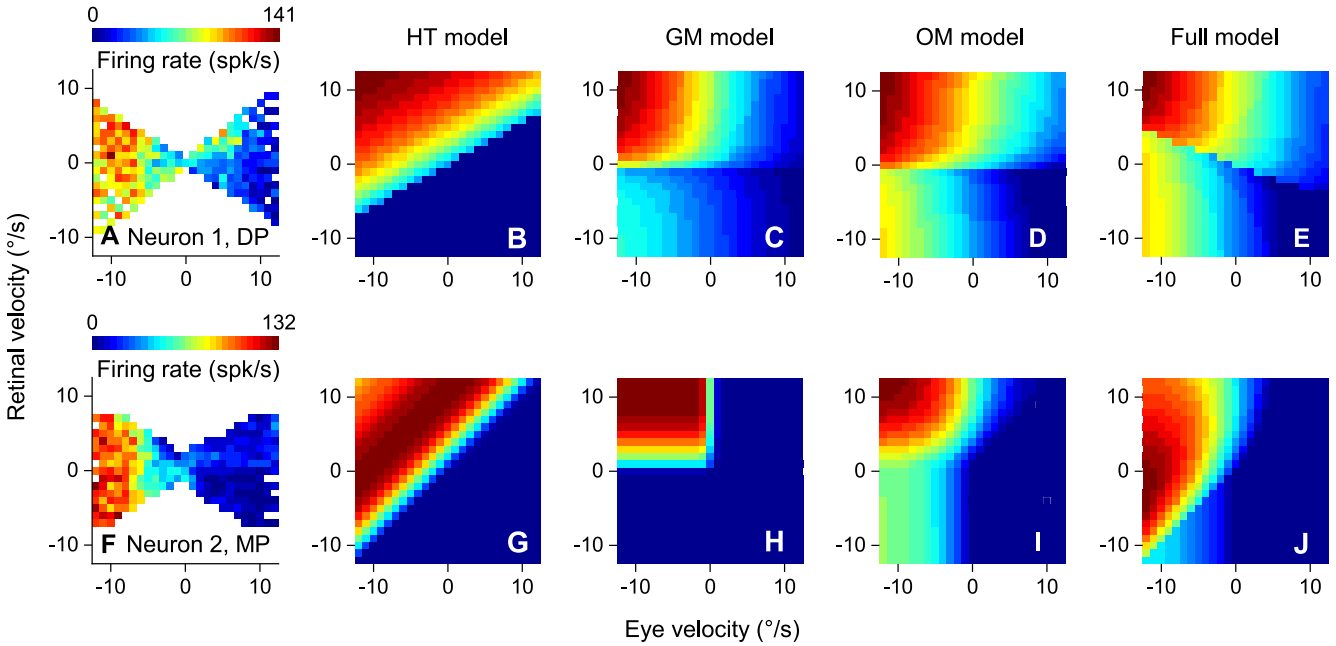

Figure 14. Joint velocity tuning profiles and model fits for two neurons with responses that are dominated by the eye velocity signal. ***A***, Joint velocity tuning profile of one example neuron (DP condition) that shows responses that are heavily dependent on eye velocity (*x* axis), with little dependence on retinal velocity (*y* axis). ***B–E***, Model predictions of joint tuning for this neuron generally fail to capture the pattern in the data (***B***, HT model; ***C***, GM model; ***D***, OM model; ***E***, Full model). ***F–J***, Data (***F***) and model fits (***G–J***) for another example neuron that shows a dominant effect of eye velocity, this time in the MP condition.

coordinates, we have not directly examined the link between head-centered velocity tuning and the CT context that explicitly requires this kind of representation. It would be interesting to see whether the diagonal structure in the joint velocity tuning profile of MT neurons is observed more frequently if an animal is trained to judge the head-centered motion of an object during pursuit eye movements. Comparing the neural interaction of retinal and eye velocities in different task contexts will also provide insights into how much the computations performed by MT neurons are shaped by top-down task-dependent signals.

Implications for the role of MT in inferring object motion and depth during self-motion

Theories of sensorimotor transformation have shown that transformation from retinal to head coordinates can be achieved at the population level by a group of neurons that are gain modulated by eye position (Zipser and Andersen, 1988; Salinas and Abbott, 1995; Pouget and Sejnowski, 1997; Salinas and Thier, 2000). It has been shown that depth from MP can be expressed as the difference between the logarithms of two velocity signals, and that this computation might also be achieved at the population level by gain modulation (Nawrot and Stroyan, 2009; Kim et al., 2017). Therefore, neurons that are gain-modulated by eye velocity could be used for encoding both head-centered velocity and depth from MP. Our study provides additional evidence that transformation from retinal to head-centered coordinates not only happens at the population level, but can also occur in individual MT neurons to some extent. The diagonal structure of joint velocity tuning that we found in some neurons could be a result of feedback signals from higher-level sensory/motor areas that represent velocity in head coordinates (Deneve et al., 2001). It could also emerge from lateral recurrent connections between neurons in MT, or through the feedback of eye velocity signals that are combined with image motion within MT. Both gain modulation by eye velocity and partial shifts in tuning toward head-centered velocity might serve as basic functions that support computations of multiple sensory variables (Salinas and Abbott, 1995; Pouget and Sejnowski, 1997; Deneve et al., 2001; Pouget et al., 2002). Further studies will be needed to understand the mechanisms that give rise to velocity selectivity in head coordinates, as well as how it shapes the functional contributions of area MT.

Our analyses suggest a shared representation between depth from MP and object motion in the world, indicating that MT is a candidate brain area for studying perceptual inferences that depend on these two variables. It would be interesting to see whether MT responses differ when an animal is trained to perform depth- or motion-related tasks using a common set of stimuli. Neurons might show diverse patterns of joint tuning for the same stimuli when the task context changes. If MT neurons indeed convey information about both depth from MP and object motion in the world, another important question is how downstream brain areas selectively read out these variables from the MT population based on task demands.

A major limitation of the motion-pursuit law is that it assumes that an object is stationary in the world, such that its retinal image motion is solely produced by self-motion (Nawrot and Stroyan, 2009). More generally, objects and the observer often move at the same time, such that retinal image motion reflects a mixture of components related to the observer's self-motion and independent object motion. In such cases, image motion needs to be parsed into these components and the motion-pursuit law should only apply to the component of image motion that is associated with self-motion. Whether and how depth perception from MP accounts for independent object motion is a topic of current investigation in our laboratory. This problem is an instance of causal inference (Kording et al., 2007), and the neural

basis of these computations is an important direction for additional research (French and DeAngelis, 2020).

In conclusion, this study set out to gain a better understanding of the interactions between retinal velocity and eye velocity that give rise to depth-sign selectivity from MP in MT. By developing computational models that capture the modulatory effects of eye velocity, we have demonstrated that MP-based depth-sign selectivity could also emerge from velocity tuning that is at least partially shifted toward head coordinates. These findings highlight the potential role of MT in representing higher-level sensory variables, including depth from MP and object motion in the world.

References

Acerbi L, Ma WJ (2017) Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. Adv Neural Information Process Syst 2017:1837–1847.

Albright TD (1984) Direction and orientation selectivity of neurons in visual area MT of the macaque. J Neurophysiol 52:1106–1130.

Arandia-Romero I, Tanabe S, Drugowitsch J, Kohn A, Moreno-Bote R (2016) Multiplicative and additive modulation of neuronal tuning with population activity affects encoded information. Neuron 89:1305–1316.

Born RT, Bradley DC (2005) Structure and function of visual area MT. Annu Rev Neurosci 28:157–189.

Bradshaw MF, Rogers BJ (1996) The interaction of binocular disparity and motion parallax in the computation of depth. Vision Res 36:3457–3468.

Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA (1996) A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis Neurosci 13:87–100.

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. J Neurosci 12:4745–4765.

Brostek L, Buttner U, Mustari MJ, Glasauer S (2015) Eye velocity gain fields in MSTd during optokinetic stimulation. Cereb Cortex 25:2181–2190.

Busse L, Wade AR, Carandini M (2009) Representation of concurrent stimuli by population activity in visual cortex. Neuron 64:931–942.

Chowdhury SA, Takahashi K, DeAngelis GC, Angelaki DE (2009) Does the middle temporal area carry vestibular signals related to self-motion? J Neurosci 29:12020–12030.

Chukoskie L, Movshon JA (2009) Modulation of visual signals in macaque MT and MST neurons during pursuit eye movement. J Neurophysiol 102:3225–3233.

DeAngelis GC, Newsome WT (1999) Organization of disparity-selective neurons in macaque area MT. J Neurosci 19:1398–1415.

DeAngelis GC, Newsome WT (2004) Perceptual 'read-out' of conjoined direction and disparity maps in extrastriate area MT. PLoS Biol 2:E77.

DeAngelis GC, Cumming BG, Newsome WT (1998) Cortical area MT and the perception of stereoscopic depth. Nature 394:677–680.

Deneve S, Latham PE, Pouget A (2001) Efficient computation and cue integration with noisy population codes. Nat Neurosci 4:826–831.

Felleman DJ, Kaas JH (1984) Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. J Neurophysiol 52:488–513.

French RL, DeAngelis GC (2020) Multisensory neural processing: from cue integration to causal inference. Curr Opin Physiol 16:8–13.

Gu Y, Watkins PV, Angelaki DE, DeAngelis GC (2006) Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. J Neurosci 26:73–85.

Howard IP, Rogers BJ (2002) Seeing in depth, Vol. 2: Depth perception. Toronto: University of Toronto.

Ilg UJ, Thier P (2003) Visual tracking neurons in primate area MST are activated by smooth-pursuit eye movements of an 'imaginary' target. J Neurophysiol 90:1489–1502.

Inaba N, Shinomoto S, Yamane S, Takemura A, Kawano K (2007) MST neurons code for visual motion in space independent of pursuit eye movements. J Neurophysiol 97:3473–3483.

Inaba N, Miura K, Kawano K (2011) Direction and speed tuning to visual motion in cortical areas MT and MSTd during smooth pursuit eye movements. J Neurophysiol 105:1531–1545.

Kim HR, Angelaki DE, DeAngelis GC (2015) A novel role for visual perspective cues in the neural computation of depth. Nat Neurosci 18:129–137.

Kim HR, Angelaki DE, DeAngelis GC (2017) Gain modulation as a mechanism for coding depth from motion parallax in macaque area MT. J Neurosci 37:8180–8197.

Komatsu H, Wurtz RH (1988) Relation of cortical areas MT and MST to pursuit eye movements: I. Localization and visual properties of neurons. J Neurophysiol 60:580–603.

Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. PLoS One 2:e943.

Lagae L, Raiguel S, Orban GA (1993) Speed and direction selectivity of macaque middle temporal neurons. J Neurophysiol 69:19–39.

Maunsell JH, Van Essen DC (1983a) 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.

Maunsell JH, Van Essen DC (1983b) Functional properties of neurons in middle temporal visual area of the macaque monkey: II. Binocular interactions and sensitivity to binocular disparity. J Neurophysiol 49:1148–1167.

Nadler JW, Angelaki DE, DeAngelis GC (2008) A neural representation of depth from motion parallax in macaque visual cortex. Nature 452:642–645.

Nadler JW, Nawrot M, Angelaki DE, DeAngelis GC (2009) MT neurons combine visual motion with a smooth eye movement signal to code depth-sign from motion parallax. Neuron 63:523–532.

Nawrot M (2003) Eye movements provide the extraretinal signal required for the perception of depth from motion parallax. Vision Res 43:1553–1562.

Nawrot M, Stroyan K (2009) The motion/pursuit law for visual depth perception from motion parallax. Vision Res 49:1969–1978.

Newsome WT, Wurtz RH, Komatsu H (1988) Relation of cortical areas MT and MST to pursuit eye movements: II. Differentiation of retinal from extraretinal inputs. J Neurophysiol 60:604–620.

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.

Ono ME, Rivest J, Ono H (1986) Depth perception as a function of motion parallax and absolute-distance information. J Exp Psychol Hum Percept Perform 12:331–337.

Ono H, Rogers BJ, Ohmi M, Ono ME (1988) Dynamic occlusion and motion parallax in depth perception. Perception 17:255–266.

Palanca BJ, DeAngelis GC (2003) Macaque middle temporal neurons signal depth in the absence of motion. J Neurosci 23:7647–7658.

Pouget A, Sejnowski TJ (1997) Spatial transformations in the parietal cortex using basis functions. J Cogn Neurosci 9:222–237.

Pouget A, Deneve S, Duhamel JR (2002) A computational perspective on the neural basis of multisensory spatial representations. Nat Rev Neurosci 3:741–747.

Rodman HR, Albright TD (1987) Coding of visual stimulus velocity in area MT of the macaque. Vision Res 27:2035–2048.

Rogers BJ (1993) Motion parallax and other dynamic cues for depth in humans. Rev Oculomot Res 5:119–137.

Rogers B, Graham M (1979) Motion parallax as an independent cue for depth perception. Perception 8:125–134.

Rogers B, Graham M (1982) Similarities between motion parallax and stereopsis in human depth perception. Vision Res 22:261–270.

Salinas E, Abbott LF (1995) Transfer of coded information from sensory to motor networks. J Neurosci 15:6461–6474.

Salinas E, Thier P (2000) Gain modulation: a major computational principle of the central nervous system. Neuron 27:15–21.

Schwarz G (1978) Estimating the dimension of a model. Ann Statist 6:461–464.

Smolyanskaya A, Ruff DA, Born RT (2013) Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable. J Neurophysiol 110:2806–2816.

Uka T, DeAngelis GC (2004) Contribution of area MT to stereoscopic depth perception: choice-related response modulations reflect task strategy. Neuron 42:297–310.

Zeki SM (1974) Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. J Physiol 236:549–573.

Zipser D, Andersen RA (1988) A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature 331:679–684.