

# Gain Modulation as a Mechanism for Coding Depth from Motion Parallax in Macaque Area MT

HyungGoo R. Kim,<sup>1</sup> Dora E. Angelaki,<sup>2</sup> and Gregory C. DeAngelis<sup>1</sup>

<sup>1</sup>Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York 14627, and <sup>2</sup>Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030

Observer translation produces differential image motion between objects that are located at different distances from the observer's point of fixation [motion parallax (MP)]. However, MP can be ambiguous with respect to depth sign (near vs far), and this ambiguity can be resolved by combining retinal image motion with signals regarding eye movement relative to the scene. We have previously demonstrated that both extra-retinal and visual signals related to smooth eye movements can modulate the responses of neurons in area MT of macaque monkeys, and that these modulations generate neural selectivity for depth sign. However, the neural mechanisms that govern this selectivity have remained unclear. In this study, we analyze responses of MT neurons as a function of both retinal velocity and direction of eye movement, and we show that smooth eye movements modulate MT responses in a systematic, temporally precise, and directionally specific manner to generate depth-sign selectivity. We demonstrate that depth-sign selectivity is primarily generated by multiplicative modulations of the response gain of MT neurons. Through simulations, we further demonstrate that depth can be estimated reasonably well by a linear decoding of a population of MT neurons with response gains that depend on eye velocity. Together, our findings provide the first mechanistic description of how visual cortical neurons signal depth from MP.

**Key words:** depth; extrastriate cortex; motion parallax; neural coding

## Significance Statement

Motion parallax is a monocular cue to depth that commonly arises during observer translation. To compute from motion parallax whether an object appears nearer or farther than the point of fixation requires combining retinal image motion with signals related to eye rotation, but the neurobiological mechanisms have remained unclear. This study provides the first mechanistic account of how this interaction takes place in the responses of cortical neurons. Specifically, we show that smooth eye movements modulate the gain of responses of neurons in area MT in a directionally specific manner to generate selectivity for depth sign from motion parallax. We also show, through simulations, that depth could be estimated from a population of such gain-modulated neurons.

## Introduction

Although we live in a three-dimensional world, visual information acquired through the eyes is a sequence of two-dimensional images that are projected onto the retinae. The brain uses a variety of cues to construct a 3D representation of the scene from the retinal images; among these, binocular disparity and motion par-

allax (MP) have been studied extensively (Howard and Rogers, 1995).

The mechanisms of depth perception from binocular disparity have become relatively well understood over time since it was demonstrated, using random-dot stereograms, that binocular disparity is a sufficient cue to depth (Julesz, 1964). For decades, theorists have proposed mathematical principles for computing depth from binocular disparity (Marr and Poggio, 1976; Longuet-Higgins, 1981), and experimental studies using animals have revealed how neurons in visual cortex encode depth from binocular disparity (Barlow et al., 1967; Ohzawa et al., 1990; Cumming and DeAngelis, 2001; Parker, 2007; Henriksen et al., 2016). Computational models (Lehky et al., 1990; Tsai and Victor, 2003) have demonstrated that depth can be decoded from biologically plausible neural representations of binocular disparity.

MP refers to the relative image motion among objects at different depths that results from observer translation. MP can also be a sufficient cue for depth perception (Rogers and Graham,

Received Feb. 9, 2017; revised June 30, 2017; accepted July 20, 2017.

Author contributions: H.R.K., D.E.A., and G.C.D. designed research; H.R.K. performed research; H.R.K. analyzed data; H.R.K., D.E.A., and G.C.D. wrote the paper.

This work was supported by NIH Grant EY013644 and by a CORE Grant (EY001319) from the National Eye Institute, and D.E.A. was supported by EY022538. We thank Dina-Jo Knoedl and Swati Shimpi for assistance with animal surgery and training, and Akiyuki Anzai for helpful comments on the paper.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Gregory C. DeAngelis, Department of Brain and Cognitive Sciences, Center for Visual Science, 310 Meliora Hall, University of Rochester, Rochester, NY 14627-0268. E-mail: gdeangelis@cvs.rochester.edu.

H. Kim's present address: Molecular and Cell Biology, Harvard University, Cambridge, MA 02138.

DOI:10.1523/JNEUROSCI.0393-17.2017

Copyright © 2017 the authors 0270-6474/17/378180-18\$15.00/0

1979), provided that information about observer movement is available. In the absence of signals related to observer movement, the sign of depth (near vs far) can be ambiguous (Farber and McCoskie, 1979; Hayashibe, 1991; Rogers and Rogers, 1992; Kim et al., 2015a). More specifically, theoretical work has shown that computation of depth from MP requires both retinal image motion and signals related to eye rotation relative to the scene (Nawrot and Stroyan, 2009). Consistent with predictions of this theory, human psychophysical studies have shown that command signals related to smooth pursuit eye movements provide a critical extra-retinal input for disambiguating depth sign (Nawrot, 2003; Naji and Freeman, 2004).

More recently, the neural basis for depth perception based on MP has begun to be revealed. Previous neurophysiological studies have demonstrated that neurons in area MT are selective for depth-sign from MP (Nadler et al., 2008), that MT responses are modulated by smooth eye movement command signals (Nadler et al., 2009), and that MT responses are correlated with perceptual judgments of depth based on MP (Kim et al., 2015a). Interestingly, MT neurons also show depth-sign selectivity when smooth eye rotations are visually simulated by large-field background motion (Kim et al., 2015b). Together, these previous studies have established area MT as a candidate neural substrate for depth perception based on MP.

However, compared with our understanding of how depth is computed from binocular disparity, there are unanswered fundamental questions about the neural processing of depth based on MP. First, although it is clear that MT neurons combine retinal image motion with both visual and extra-retinal signals regarding eye rotation to signal depth sign from MP (Nadler et al., 2009; Kim et al., 2015b), the nature of these interactions are unknown. Do eye movement signals interact additively or multiplicatively with responses to image motion? How does this interaction depend on the direction of eye movements? Is this interaction similar for visual and extra-retinal signals regarding eye rotation? This study addresses these important mechanistic questions for the first time. Our findings show that the depth-sign selectivity of MT neurons is generated primarily by a multiplicative gain modulation of responses by eye velocity, which occurs in a highly direction-dependent and temporally precise manner.

Second, it is not known how depth magnitude is computed from MP. This is not a trivial question because retinal image speed is a function of both the depth of an object and the speed of head translation; therefore, multiple combinations of depth and translation speed can generate the same retinal image speed. It is not clear how the brain resolves this ambiguity to estimate depth magnitude based on responses of neurons like those in area MT. We addressed this issue by simulating and decoding a population of model MT neurons with properties that are constrained by our experimental data. We show that if the response gain changes systematically with eye velocity, then the sign and magnitude of depth from MP can be decoded from population responses.

## Materials and Methods

The experimental results presented here are a combination of analyses from three different experimental protocols, two of which (Depth tuning measurement and Depth discrimination task, described below) were published previously (Kim et al., 2015a,b). We briefly describe each protocol below, and specify what parts of the data were used for the current analyses.

### Subjects and surgery

Two male monkeys (*Macaca mulatta*, 8–12 kg) participated in these experiments. Standard aseptic surgical procedures under gas anesthesia

were performed to implant a head restraint device. A Delrin (DuPont) ring was attached to the skull using a combination of dental acrylic cement, bone screws, and titanium inverted T-bolts (Gu et al., 2006). To monitor eye movements using the magnetic search coil technique, a scleral coil was implanted under the conjunctiva of the right eye in each animal.

A recording grid made of Delrin was affixed inside the ring using dental acrylic. The grid ( $2 \times 4 \times 0.5$  cm) contains a dense array of holes spaced 0.8 mm apart. Under anesthesia and using sterile technique, small burr holes ( $\sim 0.5$  mm diameter) were drilled vertically through the recording grid to allow the penetration of microelectrodes into the brain via a transdural guide tube. All surgical procedures and experimental protocols were approved by the University Committee on Animal Resources at the University of Rochester.

### Experimental apparatus

In each experimental session, animals were seated in a custom-built primate chair that was secured to a six degree-of-freedom motion platform (MOOG 6DOF2000E). The motion platform was used to generate passive body translation along an axis in the frontoparallel plane and the trajectory of the platform was controlled in real time at 60 Hz over a dedicated Ethernet link (Gu et al., 2006). A field-coil frame (CNC Engineering) was mounted on top of the motion platform to measure eye movements.

Visual stimuli were rear-projected onto a  $60 \times 60$  cm tangent screen using a stereoscopic projector (Christie Digital Mirage S + 3K), which was also mounted on the motion platform (Gu et al., 2006). The display screen was attached to the front side of the field-coil frame. To restrict the animal's field-of-view to visual stimuli displayed on the tangent screen, the sides and top of the field-coil frame were covered with matte black enclosures.

To generate accurate visual simulations of the animal's movement through a virtual environment, an OpenGL camera was placed at the location of one eye and the camera moved precisely according to the movement trajectory of the platform. Because the motion platform has its own dynamics, we characterized the transfer function of the motion platform, as described previously (Gu et al., 2006), and we generated visual stimuli according to the predicted motion of the platform. To account for a delay between the command signal and the actual movement of the platform, we adjusted a delay parameter to synchronize visual motion with platform movement. Synchronization was confirmed by presenting a world-fixed target in the virtual environment and superimposing a small spot by a room-mounted laser pointer while the platform was in motion (Gu et al., 2006).

### Electrophysiological recordings

We recorded extracellular single-unit activity using tungsten microelectrodes (FHC) having a typical impedance of 1–3 M $\Omega$ . The electrode was loaded into a transdural guide tube and was advanced with a hydraulic micromanipulator (Narishige). The voltage signal was amplified and filtered (1–6 kHz) using conventional hardware (BAK Electronics). Single-unit spikes were detected using a window discriminator (BAK Electronics), and the output was time-stamped with 1 ms resolution using TEMPO (Reflective Computing), which also sampled eye position signals at 200 Hz. The raw voltage signal from the microelectrode was digitized and recorded to disk at 25 kHz using a Power1401 data acquisition system (Cambridge Electronic Design). If necessary, single units were resorted off-line using a template-based method (Spike2, Cambridge Electronic Design).

The location of area MT was initially identified in each animal through analysis of structural MRI scans, which were segmented, flattened, and registered with a standard macaque atlas using CARET software (Van Essen et al., 2001). The position of area MT in the posterior bank of the superior temporal sulcus (STS) was then projected onto the horizontal plane, and grid holes around the projection area were explored systematically in mapping experiments, as described previously (Kim et al., 2015a).

### Visual stimuli

Visual stimuli were generated by a custom-written C++ program using the OpenGL 3D graphics library, and were displayed using a hardware-accelerated OpenGL graphics card (NVIDIA Quadro FX 1700). The location of the OpenGL camera was matched to the location of one of the animal's eyes, and images were generated using perspective projection. We calibrated the display such that the virtual environment had the same spatial scale as the physical space through which the platform moved the animal. Animals wore anaglyphic glasses with red and green filters (Kodak Wratten 2, nos. 29 and 61, respectively). The crosstalk between the eyes was measured using a photometer and found to be very small (0.3% for the green filter and 0.1% for the red filter). All visual stimuli were presented monocularly, except for the fixation target, which was presented dichoptically.

We used the following procedure to generate random-dot stimuli that produce a percept of depth from MP (for details, see Nadler et al., 2008 and their associated supplemental materials). A circular aperture, having a diameter slightly greater (~10%) than the optimal stimulus size for the neuron under study, was centered over the neuron's receptive field on the visual display. The initial position of each dot was generated by independently choosing random horizontal and vertical locations within the aperture. If dots lie on the theoretical horizontal horopter, the Vieth-Müller circle (VM), and the animal's eye moves along the VM circle while maintaining fixation on the target, then the retinal image motion of the dots will be zero. More generally, if dots are placed along a 3D surface that corresponds to a particular binocular disparity (referred to as an "equivalent disparity"), then the motion of dots will be homogeneous within the aperture. A set of such dots having a constant equivalent disparity forms a vertical cylinder. Figure 1A shows cross-sections through vertical cylinders corresponding to zero disparity (dotted circle), as well as near and far disparities (solid circles). Note that, in our experiments, the animal was translated along a straight path tangent to the VM circle; thus, there is slight relative motion in our stimulus even when the equivalent disparity is zero.

To present stimuli at a specific equivalent disparity, the set of random dots within the circular aperture was ray-traced onto a cylinder corresponding to the desired equivalent disparity, as described previously (Nadler et al., 2008). This ray-tracing procedure ensured that the size, location, and density of the random-dot patch were constant across simulated depths. Size and occlusion cues were eliminated by rendering transparent dots with a constant retinal size (0.39°). Critically, this procedure removed pictorial depth cues and rendered the visual stimulus depth-sign ambiguous on its own, thus requiring either extra-retinal (Nadler et al., 2009) or global visual motion (Kim et al., 2015b) signals regarding eye rotation to perceive unambiguous depth.

The above description assumes lateral translation of the observer in the horizontal plane. However, in our experiments, animals were translated along an axis in the frontoparallel plane that was aligned with the preferred-null axis of the neuron under study (to elicit robust neural responses). In this case, we rotated the virtual stimulus cylinder about the naso-occipital axis such that the axis of translation of the observer was always orthogonal to the long axis of the cylinder. This ensures that dots having the same equivalent disparities produce the same retinal speeds regardless of the axis of observer translation.

In some stimulus conditions, we introduced "depth coherence" to manipulate the amount of depth noise in the random-dot display. Depth coherence determines the proportion of dots located at a designated depth, with the remaining dots uniformly distributed in a range from  $-2^\circ$  to  $+2^\circ$  of equivalent disparity. As a result, all dots in a 100% coherence stimulus had the same direction and speed of motion at each point in time, whereas dots in a 0% coherence stimulus had a range of speeds defined by their individual depths. Because half of the dots were located near and the other half were located far at 0% coherence, half of the dots moved in each of the two opposite directions of motion at all times.

### Stimulus conditions

**Motion parallax condition.** At stimulus onset, animals experienced passive whole-body translation which followed a modified sinusoidal trajectory in the frontoparallel plane (Nadler et al., 2008, 2009). Each

movement involved one cycle of a 0.5 Hz sinusoid that was windowed by a Gaussian function that was exponentiated to a high even power (Nadler et al., 2013), to prevent rapid accelerations at stimulus onset and offset. The resulting retinal velocity profiles for stimuli at two different depths are shown in Figure 1B. On half of the trials, platform movement started toward the neuron's preferred direction of motion ( $0^\circ$  phase). On the remaining half, the motion started toward the neuron's null direction ( $180^\circ$  phase). The animal was required to move his eyes to maintain visual fixation on a world-fixed target. Along with the physical translation of the head, we moved the OpenGL camera such that the camera followed the trajectory of the animals' actual eye position. This ensures that the animals experience optical stimulation consistent with self-motion through a stationary 3D virtual environment. In this condition, smooth pursuit eye movement command signals are available to disambiguate depth sign, as demonstrated previously (Nadler et al., 2008, 2009).

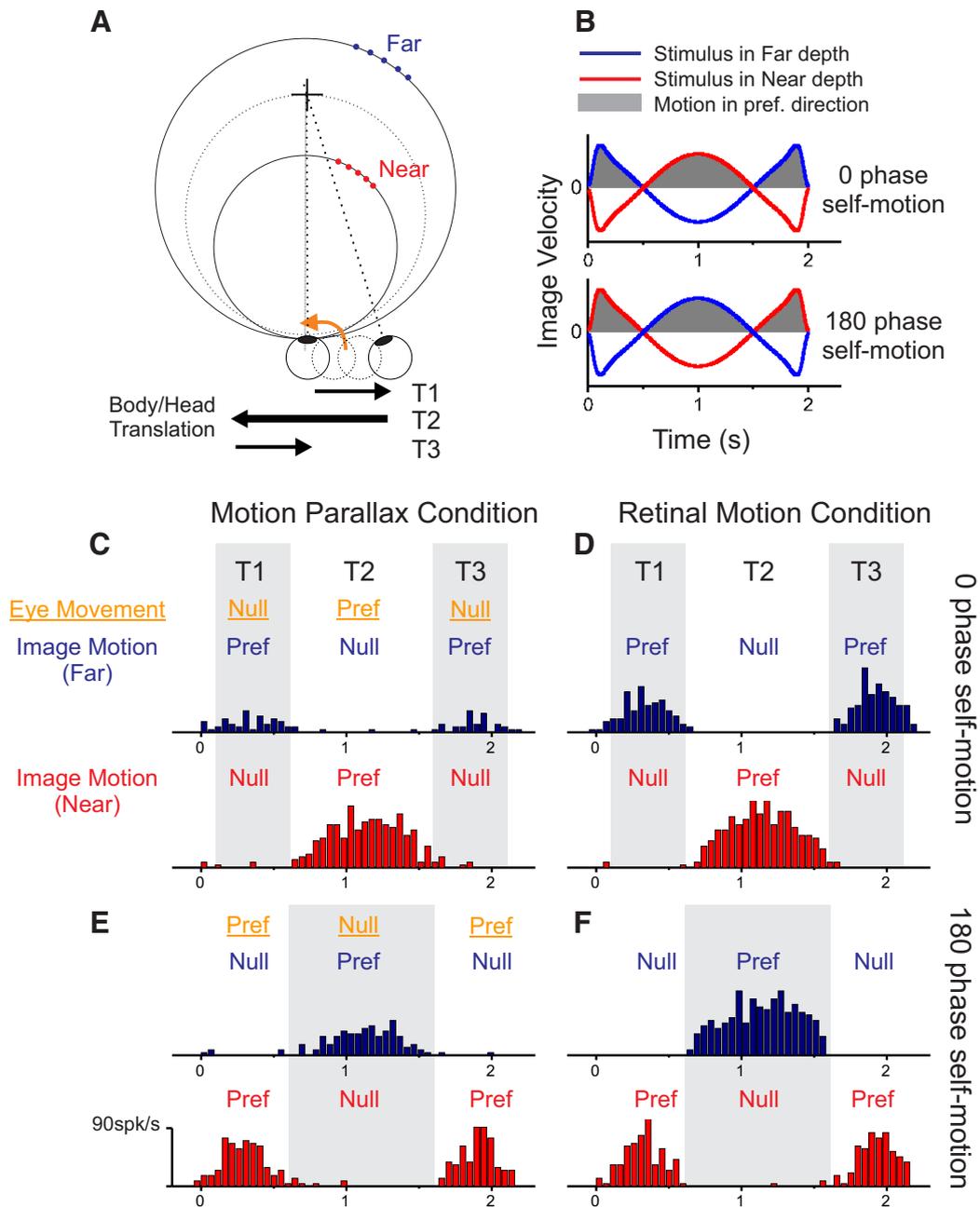
**Retinal motion condition.** The retinal image motion of the random-dot patch was the same as in the MP condition, but this condition lacks physical head translation and the corresponding counteractive eye movements. In this condition, the OpenGL camera was translated and counter-rotated such that the camera was always aimed at the fixation target, thus effectively simulating the image motion that results from the combination of head translation and pursuit eye movement that occurs in the MP condition. Thus, the retinal motion (RM) condition reproduces the visual stimulus that would be experienced in the MP condition (assuming that animals pursued the fixation target accurately in the MP condition), but eliminates the extra-retinal signals related to head and eye movements.

**Dynamic perspective condition.** The motion of the random-dot patch over the receptive field was identical to the RM condition, but the scene also contained additional elements ( $0.22 \times 0.22$  cm triangles) that formed a 3D background. The motion of these background dots provided robust dynamic perspective cues regarding changes in eye orientation relative to the scene (Kim et al., 2015b). Background dots were randomly positioned in a volume that spanned a range of equivalent depths of  $\pm 20$  cm around the fixation target, and the dot density was 0.01 dots/cm<sup>3</sup>. Background dots were masked within a circular region that was centered on the receptive field, and the masked region was typically two to three times larger than the diameter of the receptive field of each neuron. The annular mask area included the fixation target in most cases (85/103). The mask ensured that the movement of background dots did not encroach upon the classical receptive field of the neuron under study.

### Experimental protocols

**Depth tuning measurement.** Depth tuning of MT neurons based on MP was measured in the three stimulus conditions described above. All neurons were tested in the RM and MP conditions, and a subset of 103 neurons was also tested in the dynamic perspective (DP) condition. At the beginning of each trial, a fixation target appeared at the center of the display. After the animal established fixation for 0.2 s, a patch of dots was presented (monocularly) in the receptive field of the MT neuron and its motion simulated one of nine different depths corresponding to equivalent disparities of  $-2^\circ$ ,  $-1.5^\circ$ ,  $-1^\circ$ ,  $-0.5^\circ$ ,  $0^\circ$ ,  $0.5^\circ$ ,  $1^\circ$ ,  $1.5^\circ$ , and  $2^\circ$ . During stimulus presentation, the monkey experienced real (MP condition) or simulated (RM and DP conditions) translation. In the MP condition, the monkey was required to generate smooth eye movements to track the world-fixed fixation target; in the RM and DP conditions, no eye movements were required to maintain fixation and eye rotation was simulated by rotation of the OpenGL camera. A small electronic window around the fixation target was used to monitor the accuracy of pursuit or fixation of the eye to which all monocular visual stimuli were presented. In the MP condition, the initial size of the target window was  $3-4^\circ$ , and it shrunk to  $2.1-2.8^\circ$  after 250 ms of platform movement. This allowed the animal a brief period of time to initiate pursuit and to execute a catch-up saccade to arrive on target. At the end of visual stimulation, both the fixation target and the visual stimulus disappeared, and the animal obtained a liquid reward (0.2–0.4 ml) for keeping his eye within the target window.

**Modulation test with random-depth stimuli.** For a subset of neurons, we measured responses to 0% depth coherence stimuli in a separate block of



**Figure 1.** Temporal decomposition of MT responses according to eye velocity and retinal velocity. **A**, In the MP condition, an animal was translated side to side (black arrows, T1, T2, T3) while counter-rotating its eyes to maintain fixation on a world-fixed target (cross). In each trial, random-dot motion simulated a surface at one of nine depths ranging from near (red) to far (blue). Random dots were presented monocularly, and the fixation target was presented dichoptically. **B**, Velocity profile of image motion. Translation started toward the preferred direction of a neuron (0° phase), or toward the anti-preferred direction (180° phase). The retinal image velocity depends on both the depth of the stimulus and the phase of self-motion. **C**, PSTHs of responses from an example near-preferring neuron for 0° phase in the MP condition. Data are shown for near and far depths of  $-2^\circ$  and  $+2^\circ$  (red and blue PSTHs, respectively). Firing rates and average retinal velocities were computed in three time windows (T1, T2, T3; see Materials and Methods). **D**, Responses to 0° phase stimuli in the RM condition. **E, F**, Responses to 180° phase stimuli in the same format as **C** and **D**.

trials. For a 0% depth coherence stimulus, dots within the receptive field were located at random depths between  $-2$  and  $+2$  degrees of equivalent disparity. The MP and RM conditions were randomly interleaved in this protocol. For 37 of 95 neurons that were tested with this protocol, the animals were simply required to maintain fixation on the world-fixed target during stimulus presentation; no behavioral report was required, and reward was given at the end of stimulus presentation. The remainder of the 0% coherence data (58/95 neurons) were collected while the animals were performing a depth discrimination task in which they reported the depth-sign (near vs far) of the random-dot stimulus based on MP (Kim et al., 2015a; 0% coherence condition). We combined 0% coherence data from the fixation and discrimination tasks because we did not find a noticeable difference in our main results.

*Data analysis*

*Decomposition of responses.* Animals were translated back and forth laterally to generate MP (Fig. 1A). On one-half of trials, translation of the motion platform started toward the preferred direction of visual motion of the neuron under study (e.g., rightward initial platform motion for a neuron that prefers rightward visual image motion; Fig. 1B; 0° phase). On the other half of trials, translation of the platform started toward the neuron’s anti-preferred direction of motion (Fig. 1B; 180° phase). Our stimulus geometry (Fig. 1A) assured that image motion for a near depth and 0° phase (Fig. 1B, top, red) is the same as image motion for a far depth and 180° phase (Fig. 1B, bottom, blue). Because the fixation target was stationary in the world, the direction of eye movement was always opposite to the direction of platform motion (Fig. 1C,E, orange). Note that the

direction of image motion depends on both the depth sign of the stimulus (Fig. 1C–F, red and blue) and the direction of self-motion.

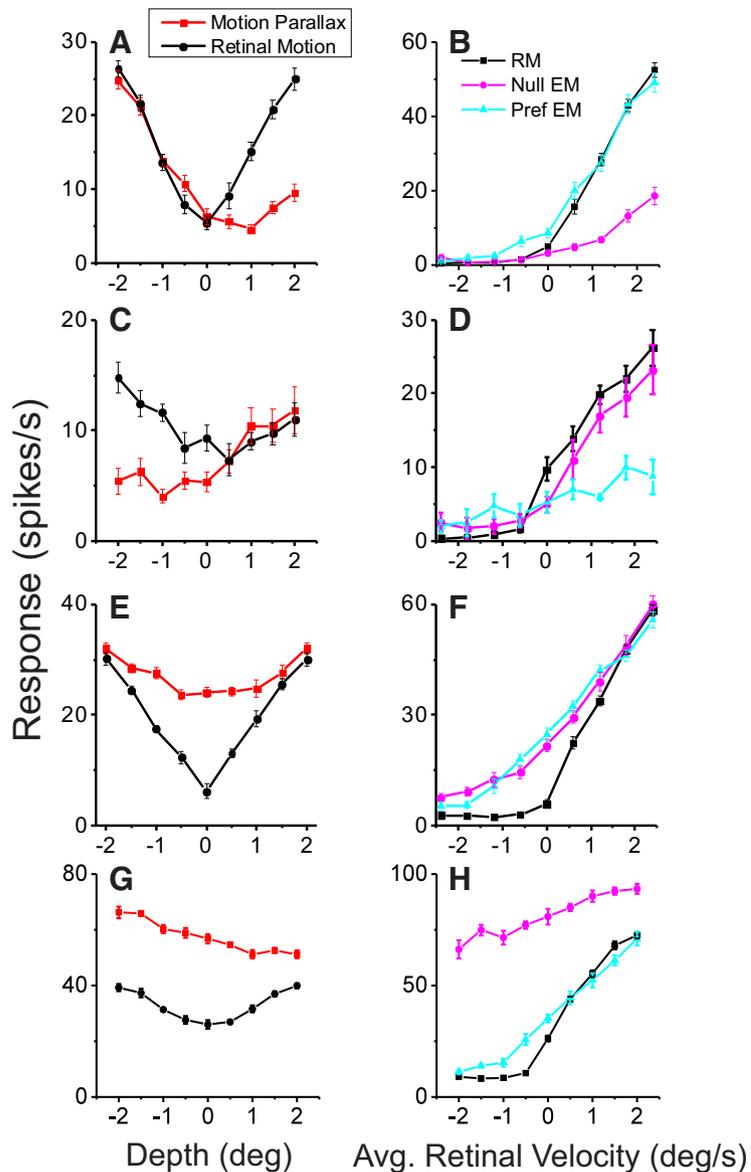
To analyze the effect of eye movements on neural responses, we defined three time windows (Fig. 1C; T1, T2, and T3) based on the direction of self-motion. T1, T2, and T3 were defined as 0–500, 500–1500, and 1500–2000 ms, respectively, relative to movement onset of the motion platform. Analysis windows were shifted 100 ms to account for the response latency of most neurons to visual motion. Because the visual stimulus appeared on average 100 ms before movement onset, phasic responses to stimulus onset were excluded. To compute firing rates, spike counts were summed across time windows T1 and T3 and then divided by the cumulative duration of those two windows (1 s). Similarly, spike counts in time window T2 were divided by its duration (also 1 s). These two measurements yielded average firing rates for each specific combination of eye movement direction (preferred vs null) and retinal image velocity. Firing rates from the 0° and 180° phases of self-motion were combined for each specific combination of eye direction and image motion direction [e.g., data from windows T1 + T3 (Fig. 1D) would be combined with data from T2 (Fig. 1F)].

Repeating this decomposition procedure for the four pairs of depths that are symmetric around the plane of fixation ( $\pm 0.5^\circ$ ,  $\pm 1.0^\circ$ ,  $\pm 1.5^\circ$ ,  $\pm 2.0^\circ$ ) yielded four sets of responses with different average retinal speeds. To quantify the average retinal speed in each time window (assuming accurate tracking of the fixation target), we used the retinal velocity trajectories to compute the average retinal speed for each time window and depth, combined across the two self-motion phases. This allowed us to transform the data into retinal velocity tuning curves for each direction of eye movement (Fig. 2B, D, F, H, magenta and cyan). We also computed a retinal velocity tuning curve under conditions in which there was no eye movement by using responses from the RM condition.

**Depth-sign discrimination index.** We quantified selectivity for depth-sign (near vs far) using a depth-sign discrimination index (DSDI; Nadler et al., 2008):

$$DSDI = \frac{1}{4} \sum_{i=1}^4 \frac{R_{far(i)} - R_{near(i)}}{|R_{far(i)} - R_{near(i)}| + \sigma_{avg(i)}} \quad (1)$$

For each pair of depths (index by  $i$ ) that is symmetrical around zero (for example,  $\pm 2^\circ$ ), the difference in mean response between far and near depths ( $R_{far(i)} - R_{near(i)}$ ) was computed relative to response variability ( $\sigma_{avg(i)}$ , the average SD of responses to that pair of depths). This quantity was then averaged across the four pairs of depth magnitudes to obtain the DSDI ( $-1 < DSDI < +1$ ). Near-preferring neurons have negative DSDI values, whereas far-preferring neurons have positive DSDI values. Statistical significance of DSDI values was evaluated using a permutation test in which DSDI values were computed 1000 times after shuffling responses across depths. If the measured DSDI value is negative, the  $p$  value is the proportion of shuffled DSDIs less than the measured DSDI value. If the measured DSDI is positive, the  $p$  value is the proportion of DSDIs greater than the measured DSDI value.



**Figure 2.** Transformation of depth tuning curves into retinal velocity tuning curves. **A**, Depth tuning curves for the example neuron of Figure 1. Responses in the RM condition are symmetrical around zero depth (black), whereas responses to far stimuli in the MP condition are suppressed (red; DSDI =  $-0.67$ ). **B**, Data from the same neuron are plotted as retinal velocity tuning curves, conditioned on eye movement direction (magenta: eye movement toward the neuron's null direction; cyan: eye movement toward preferred direction; black: no eye movement). Responses are suppressed for null direction eye movements, yielding MI =  $-0.56$ . **C**, Depth tuning curves for an example far-preferring neuron (DSDI =  $0.62$ ). **D**, Retinal velocity tuning curves for the cell in **C** show that responses are suppressed when the eye moves toward the preferred direction (MI =  $0.39$ ). **E**, An example neuron that is not selective for depth sign (DSDI =  $-0.09$ ). **F**, Retinal velocity tuning curves for the neuron of **E** show similar effects of preferred and null direction eye movements (MI =  $-0.004$ ). **G**, Another example of a near-preferring neuron (DSDI =  $-0.7$ ). **H**, The neuron shows increased baseline response when the eye moves in the null direction, resulting in a positive MI value (MI =  $0.23$ ).

**Eye movement modulation index.** Velocity tuning curves corresponding to preferred and null direction eye movements (Fig. 2B, D, F, H, cyan and magenta curves, respectively) characterize the effect of eye movements on the retinal velocity selectivity of MT neurons. We quantified the amount of response modulation between the preferred and null directions of eye movements using a modulation index (MI):

$$MI = \frac{1}{4} \sum_{i=1}^4 \frac{R_{null(i)} - R_{pref(i)}}{R_{null(i)} + R_{pref(i)}} \quad (2)$$

For each retinal speed in the preferred direction of motion, we computed the difference in response between the null and preferred directions of eye movements ( $R_{null(i)} - R_{pref(i)}$ ) relative to the sum of the two. The

results were averaged across the four retinal speeds corresponding to motion in the neuron's preferred direction. Note that we define a "preferred" eye movement as a movement in the direction of the cell's visual motion preference; for a neuron that prefers rightward visual motion on the display screen, a rightward eye movement is a preferred eye movement and a leftward eye movement is "null". Preferred and null eye movement directions are not defined based on the neural responses they elicit in the MP condition. Negative MI values indicate that responses are greater when the eye moves toward the preferred direction relative to the null direction. Positive MI values indicate that responses are greater during null direction eye movements.

**Comparison of linear and nonlinear models.** In one of our main analyses (Figs. 4, 5), we tested whether a simple linear model can predict the responses of MT neurons during eye movements (MP condition) from responses during fixation (RM condition). The linear model is a first-order polynomial such that the effect of eye movements has both additive and multiplicative effects on neural response:

$$R_{MP} = a_0 + a_1 R_{RM}, \quad (3)$$

where  $R_{MP}$  and  $R_{RM}$  denote responses in the MP and RM conditions, respectively, and  $a_0$  (offset) and  $a_1$  (gain) are free parameters. To evaluate whether a nonlinear model (i.e., a higher-order polynomial) could fit the data substantially better, we examined whether the goodness of fit was significantly improved by adding an additional term to the linear model:

$$R_{MP} = a_0 + a_1 R_{RM} + a_2 R_{RM}^n, \quad (4)$$

where  $a_2$  is an additional free parameter. In our main analysis, the exponent ( $n$ ) was set to 2; however, we also fit the nonlinear model separately with exponents ranging from 3 to 10, in integer steps, to determine whether a greater exponent would substantially improve the fits. Parameters were fitted by a least-squares method, and goodness of fit ( $R^2$ ) was defined as follows:

$$R^2 = 1 - \frac{SSE}{SST}, \quad (5)$$

where SSE is the sum of squared errors between model predictions and data, and SST is the sum of squared differences between the overall mean response and the data. Because both models are linear in their parameters (with  $n$  set to a specific value), a sequential  $F$  test was used to test whether the nonlinear model provided a significantly better fit than the linear model.

### Simulation and decoding of model responses

**Generation of model responses.** We simulated responses of a population of 2000 model MT neurons to test whether gain-modulated responses can reliably encode depth from MP. We simulated responses for leftward and rightward retinal image motions, and we modeled the speed tuning of two populations of neurons, one tuned to each direction of motion, with responses that are modulated by eye velocity. By convention, negative speeds correspond to leftward motion, and positive speeds correspond to rightward motion. Speed tuning curves are modeled as log-Gaussian functions, and the speed preferences of neurons are assumed to be equally spaced on a logarithmic speed axis for both leftward (negative speeds) and rightward (positive speeds) directions (see Fig. 8A). These assumptions regarding the shape and distribution of speed tuning of model neurons are well justified by empirical work (Nover et al., 2005).

More specifically, the speed tuning of the  $i$ th model neuron was modeled as follows:

$$h_i(s) = A_i \times \exp\left(-\frac{(\log(s + s_{\text{offset}}) - \log(s_i + s_{\text{offset}}))^2}{2\sigma_i^2}\right), \text{ for } s_i \geq 0, \quad (6)$$

$$h_i(s) = A_i \times \exp\left(-\frac{(\log(-s + s_{\text{offset}}) - \log(-s_i + s_{\text{offset}}))^2}{2\sigma_i^2}\right), \text{ for } s_i < 0,$$

where  $A_i$  denotes the amplitude of the tuning curve (uniformly distributed from 60 to 90 spikes/s),  $s_i$  denotes the speed preference of each neuron (uniformly distributed in log speed from 0.31–20.0 deg/s for rightward motion, and with the same range of negative values for leftward motion), and  $\sigma_i$  indicates the width of the Gaussian in log speed units (uniform from 0.5 to 1.5).  $s_{\text{offset}}$  (fixed at 0.1) is a constant to keep the function from becoming undefined when stimulus speed approaches zero. Note that we set  $h_i(s) = 0$  if the value inside the first log-term is  $\leq 0$ .

To model the effect of eye movements on MT responses, the response gain of each model neuron was assumed to be linearly related to the logarithm of eye velocity:

$$g_i(e) = b_i \log(e + 1) \times \text{sign}(s_i) + c_i \quad e \geq 0 \\ g_i(e) = -b_i \log(-e + 1) \times \text{sign}(s_i) + c_i \quad e < 0, \quad (7)$$

where  $e$  corresponds to scene-relative eye velocity, and  $b_i$  characterizes the slope of the dependency of response gain on log eye velocity.  $c_i$  determines baseline gain for each neuron, which is randomly drawn from a normal distribution (mean = 0.75, SD = 0.12) that is chosen to roughly mimic the observed data.  $\text{Sign}(s_i)$  is 1 if the neuron prefers rightward motion and  $-1$  if the neuron prefers leftward motion. We will explain how to assign  $b_i$  to each neuron below. Positive values of  $e$  indicate rightward eye movements, and negative values denote leftward eye movements. Because the response gains (slopes) of real neurons in our experiments are mostly within a range from 0.2 to 1.4 (see Fig. 5A), we set upper and lower bounds to constrain  $g_i$  values to a physiologically plausible range. The upper bound on  $g_i$  was randomly drawn from a uniform distribution between 1.2 and 1.4, and the lower bound was randomly drawn from a uniform distribution between 0.2 and 0.4. With these parameter choices, model neurons show a diverse range of gain functions (see Fig. 8B).

In addition to gain changes, we observed differences in response offsets between opposite directions of eye movements (Figs. 2H, 5D). To capture this, we modeled a response offset as follows:

$$o_i(e) = \text{rect}(\text{sign}(e) \times \text{sign}(s_i) \times \Delta\text{intercept}_i) + \text{const}_i, \quad (8)$$

where  $\text{sign}(e)$  is  $+1$  if the eye movement is rightward and  $-1$  if the eye movement is leftward. Thus,  $\text{sign}(e) \times \text{sign}(s_i)$  is positive when the eye moves toward the preferred direction of neuron, and the result is negative when the eye moves toward the null direction of neuron. The  $\text{rect}()$  function performs rectification,  $\Delta\text{intercept}_i$  is a response offset that depends on eye movement direction for each neuron, and  $\text{const}_i$  is a constant baseline (randomly drawn from a Poisson distribution with mean = 25). We found a highly significant negative correlation between slope difference and intercept difference (see Fig. 8D, black dots;  $N = 82$ ,  $r = -0.66$ ,  $p < 10^{-10}$ , Spearman correlation). Thus, we attempted to model this empirical relationship by drawing  $b_i$  (Eq. 7) and  $\Delta\text{intercept}_i$  (Eq. 8) from a multivariate normal distribution with means = [0.068, 13.2] and covariance matrix = [0.019,  $-1.72$ ,  $-1.72$  750] (see Fig. 8D, gray dots). These parameters were determined manually to approximate the empirical data and avoid biases caused by a few outliers.

Having specified the functional forms of retinal speed tuning and response modulation by eye velocity, mean responses of model neurons were then generated by multiplying the response gain associated with a particular eye velocity by the output of the speed tuning function (see Fig. 8A, dashed curves) and by adding the corresponding offset:

$$f_i(s, e) = g_i(e) \times h_i(s) + o_i(e). \quad (9)$$

We compared performance among four variants of this generative model (see Fig. 8E, F). In the full model (see Fig. 8E, F, Gain +  $\Delta\text{Offset}$ ), we used  $b_i$  and  $\Delta\text{intercept}_i$  values that were randomly chosen from the multivariate normal distribution. In the gain-only model (see Fig. 8E, F, Gain),  $\Delta\text{intercept}_i$  of each individual neuron was set to zero. In the offset-only model (see Fig. 8E, F,  $\Delta\text{Offset}$ ),  $b_i$  of each individual neuron was set to zero, such that  $g_i = 1$ . In the retinal motion model (see Fig. 8E, F, RM), we used  $b_i = 0$  and  $\Delta\text{intercept}_i = 0$  for each neuron. Responses of model neurons for individual simulated trials were then generated from independent Poisson random variables, each having the mean response of a

neuron,  $f_i(s)$ . In this way, we generated model population responses that could be decoded.

**Depth and retinal motion parameters for decoding simulation.** To quantify how well depth could be estimated from the population responses of our model neurons, we examined the performance of the model using stimulus parameters similar to those used in a previous human psychophysics study (Nawrot et al., 2014). The depth ratio (depth/viewing distance) of stimuli in the simulation ranged from  $\pm 0.05$  to  $\pm 0.25$  in steps of 0.05. Positive depth ratios indicate stimuli that are far relative to the point of fixation, and negative depth ratios indicate near stimuli (see Fig. 8C). With a fixation distance of 50 cm, this produces a range of depths from  $-12.5$  to  $+12.5$  cm. We added zero depth to the range of parameters used by Nawrot et al. (2014).

The speeds of image motion in the simulation ranged from 0.14 deg/s to 1.65 deg/s, equally spaced on a log scale for each direction of motion. For each combination of depth ratio and image speed, eye velocity was computed using the motion/pursuit law (Nawrot and Stroyan, 2009):

$$\frac{d}{f} \approx \frac{d\alpha/dt}{d\theta/dt} \quad (10)$$

$$d\alpha/dt \approx \frac{d}{f} d\theta/dt,$$

where  $d$  is the depth of an object relative to the fixation point,  $f$  is the fixation distance,  $d\alpha/dt$  indicates eye velocity, and  $d\theta/dt$  indicates the retinal image velocity of the object. Eye movement speeds faster than 12 deg/s or slower than 1.1 deg/s were eliminated (Nawrot et al., 2014). At zero depth, retinal image velocity is set to zero, and eye velocity takes on values that are equally spaced on a linear axis between  $-11$  deg/s and  $+11$  deg/s.

Note that, as shown in Figure 8C, the same retinal image speed can be generated by inverting both eye movement direction and stimulus depth (depth-sign ambiguity). For example, in Figure 8C, the blue square denotes a far stimulus with a leftward eye rotation that generates a retinal speed of 1.65 deg/s. The red circle represents a corresponding near stimulus with a rightward eye rotation that generates the same retinal image speed. Because there are multiple depth ratios that correspond to the same speed of image motion, it is not a viable strategy to estimate depth based solely on retinal image speed. Thus, the question that we address with these simulations is whether the gain and/or offset modulations of responses to retinal velocity provide a neural substrate from which depth can be estimated across a variety of combinations of eye velocity and retinal velocity.

**Training the network and decoding population responses.** Responses of model neurons were decoded by a simple linear decoder. Depth of the stimulus was estimated as a weighted linear sum of responses of model MT neurons. We generated 1000 simulated trials of population responses for each distinct combination of stimulus depth and eye velocity. Decoding weights were fitted by linear regression of depth onto responses, using half of the total trials.

$$depth = a + \sum_i b_i r_i, \quad (11)$$

where  $a$  and  $b$  represent free parameters to be estimated. The remaining half of simulated trials was used to test the performance of the decoder. Estimated depths were combined across eye velocities to compute the mean and SD of each depth estimate (see Fig. 8F).

### Experimental design and statistical analysis

For each animal and experimental protocol, we recorded from a sufficient number of neurons (typically 50–100 per animal) to allow us to make robust inferences, as per the standards of the field. Data were generally pooled across the two animals, although key results are also presented separately for each animal in the Results section and all main effects are consistent across animals. Because both animals were males, no information is available with regard to possible sex differences. For each experimental protocol, as described in detail above, all relevant stimulus conditions were randomly interleaved to avoid possible con-

foundations of nonstationary neural responses. All collection of neural data was accomplished by computer.

All statistical analyses were performed with MATLAB software (MathWorks). Given that most of the distributions of outcome metrics showed some deviation from normality, nonparametric statistics were generally used. For comparison of the central tendencies (e.g., medians) of two distributions, we used the Wilcoxon signed rank test for paired data and the Wilcoxon rank sum test for unpaired data. To assess correlations between variables, we used Spearman rank correlations. Full reporting of the statistics and sample sizes for each analysis can be found in Results. To assess whether custom-designed indices, such as the DSDI (Eq. 1) or MI (Eq. 2), were significantly different from zero for individual neurons, we performed permutation tests by scrambling the order of stimulus conditions. Bootstrap analysis was used to compute 95% confidence intervals for slopes and intercepts of model fits, and the sequential  $F$  test was used to compare polynomial models with different numbers of terms.

## Results

Our overall goal is to understand how signals related to eye movements modulate the responses of MT neurons to create a neural representation of depth sign from MP. We begin by examining how responses to image motion are modulated by the direction of smooth-pursuit eye movements. Next, we explore whether the nature of response modulation by eye movements is additive or multiplicative. Finally, we use simulations to show how depth sign and magnitude might be decoded from the population responses of model neurons that are based on our data.

We collected data from 231 neurons in area MT of two macaque monkeys (76 from Monkey 1, 155 from Monkey 2). We recorded from any isolatable single neuron, with the exception of  $\sim 5\%$  of neurons that did not respond to the range of speeds of motion used in our task (which is  $\sim 0$ – $7$  deg/s). Among this total sample of 231 neurons, 103 were included in a previous study of dynamic perspective cues (Kim et al., 2015b), 64 were included in a previous study of neuronal and behavioral depth discrimination (Kim et al., 2015a), and 69 neurons were new to this study.

### Temporal decomposition of responses

To measure depth-sign tuning from MP, animals were translated by a motion platform along each neuron's preferred-null axis of motion (Nadler et al., 2008, see their Materials and Methods and Fig. S1 for details). Figure 1A illustrates the stimulus geometry for a case in which the neuron prefers rightward motion, such that body translation occurs along the horizontal (i.e., interaural) axis. During translation by the motion platform, the animal was required to maintain fixation on a world-fixed target by counter-rotating its eyes. A patch of random dots was placed over the receptive field of the neuron under study, and the motion of these dots simulated a stationary surface that was located at one of nine depths, ranging from near to far relative to the fixation point (Fig. 1A; see Materials and Methods).

In each trial, the animal was translated through one cycle of a modified 0.5 Hz sinusoid (see Materials and Methods). As a result, there were three distinct time periods with opposite directions of body motion (Fig. 1A, T1, T2, T3). On half of the trials, translation started in the preferred direction of the neuron ( $0^\circ$  phase trials); on the other half, translation started in the null direction ( $180^\circ$  phase trials). Because the eyes needed to counter-rotate to maintain fixation on the world-fixed target, the direction of eye movement was always opposite to that of body movement. For  $0^\circ$  phase self-motion, as the body moves in the same direction as the neuron's preferred direction, the eye moves in the neuron's null direction (Fig. 1C, orange), and vice versa for the  $180^\circ$  phase of self-motion (Fig. 1E, orange). Note that we simply define a preferred eye movement as a

movement in the direction of the cell's visual motion preference. Preferred and null eye movement directions are not defined based on the neural responses they elicit in the MP condition.

The image motion of dots within a neuron's receptive field depended on both the simulated depth of the dots and the phase of self-motion. The retinal motion direction of far stimuli is always in the same direction as self-motion (Fig. 1*B,C,E*, blue), whereas the image motion of near stimuli is always in the direction opposite to self-motion (Fig. 1*B,C,E*, red). Thus, by decomposing responses into time windows linked to the direction of self-motion (T1, T2, T3), we can quantify neural responses for all four combinations of image motion directions and eye movement directions for each depth. In addition, because retinal speeds vary across stimulus depths, this decomposition allows us to measure retinal velocity tuning for each direction of eye movement.

To evaluate the effects of eye movements on MT responses, we also analyzed data collected in the RM control condition. In this condition, animals were stationary and did not make any systematic eye movements, but the visual motion of the dots simulated the image motion produced by translation and rotation of the eye during the MP condition. This condition produces the same alternating patterns of image motion in the receptive field (Fig. 1*D,F*), but in the absence of extra-retinal signals related to eye or body motion. Note, in particular, that the image motion of far dots during 0° phase self-motion is the same as the image motion of near dots during 180° phase self-motion (Fig. 1*B*).

### Selective modulation of MT responses by eye movement direction

Peristimulus time histograms (PSTHs) of the responses of a near-preferring MT neuron, to simulated depths of  $-2^\circ$  and  $+2^\circ$ , are illustrated in Figure 1*C–F*. Comparison of responses in the MP and RM conditions (panels *C* vs *D* and *E* vs *F*) allows the reader to observe the effect of eye movements on MT responses to the visual motion sequence. During time periods in which the eye moved in the direction matching the neuron's null direction and image motion was in the neuron's preferred direction (Fig. 1*C*, top row, T1 and T3; *E*, top row, T2), the responses of this neuron were clearly suppressed relative to the corresponding time periods in the RM condition (Fig. 1*D,F*). In contrast, during time periods in which the eye moved toward the neuron's preferred direction and image motion was also in the preferred direction (Fig. 1*C*, bottom row, T2; *E*, bottom row, T1 and T3) responses were very similar to those in the corresponding time periods of the RM condition. Hence, for this near-preferring neuron, responses were suppressed when the eye moved toward the null direction, whereas responses were largely unaffected when the eye moved toward the preferred direction.

Figure 2*A* shows depth tuning curves for the same example neuron, in the format described in previous studies (Nadler et al., 2008, 2009, 2013). Note that the depth tuning of this neuron in the RM condition is symmetrical around zero depth (Fig. 2*A*, black), reflecting the fact that visual image motion is depth-sign ambiguous. In striking contrast, responses in the MP condition show much greater activity for near than far depths (Fig. 2*A*, red), which results from a modulatory influence of eye movements (Nadler et al., 2009). To quantify depth-sign selectivity, we computed a DSDI (see Materials and Methods), which is negative for neurons that prefer near depths and positive for neurons that prefer far depths. The example neuron of Figure 2*A* has a DSDI value of  $-0.67$ , which is significantly different from zero ( $p < 0.001$ , permutation test; see Materials and Methods).

As described above, the temporal decomposition of responses allows us to transform the depth tuning curves of Figure 2*A* into retinal velocity tuning curves for each direction of eye movement in the MP condition (and fixation in the RM condition). Figure 2*B* shows the resulting retinal velocity tuning curves for the same example neuron. Because this neuron preferred fast speeds (preferred speed  $\sim 10$  deg/s), responses increase monotonically with positive retinal velocities (image motion in the preferred direction) and are close to zero for negative retinal velocities (image motion in the null direction). Responses were clearly suppressed when the eye moved toward the null direction (Fig. 2*B*, magenta), whereas eye movements in the preferred direction (Fig. 2*B*, cyan) did not substantially alter responses relative to the RM condition (Fig. 2*B*, black). Notice that null direction eye movements (Fig. 2*B*, magenta) appear to scale down the responses to retinal motion (Fig. 2*B*, black), which will be discussed in detail below.

The effect of eye movement direction on the responses to image motion was quantified using a MI (see Materials and Methods). To compute MI, responses to retinal motion during preferred eye movements were subtracted from responses during null eye movements; this difference was then divided by the sum of the two quantities and the result was averaged across all retinal velocities in neuron's preferred direction (see Materials and Methods). An MI value of zero indicates no difference in response between eye movements toward the preferred and null directions. A negative MI means that responses are weaker when the eye moves toward the null direction, and a positive MI means that responses are weaker when the eye moves toward the preferred direction. The MI value for the near-preferring example neuron of Figure 2*A,B* was  $-0.56$  ( $p < 0.001$ , permutation test).

Analogous data for another example neuron are shown in Figure 2, *C* and *D*. This neuron shows a clear preference for far depths (Fig. 2*C*; DSDI = 0.62,  $p < 0.001$ , permutation test). Interestingly, the effect of eye movement direction is opposite to the previous example neuron: responses are suppressed when the eye moves toward the preferred direction of the neuron (Fig. 2*D*, cyan), but not when the eye moves toward the null direction (Fig. 2*D*; magenta, MI = 0.39;  $p < 0.001$ , permutation test).

For neurons that did not show significant depth-sign selectivity, such as the example cell shown in Figure 2*E* (DSDI =  $-0.09$ ,  $p = 0.28$ , permutation test), retinal velocity tuning curves tended not to differ between preferred and null direction eye movements (Fig. 2*F*; MI =  $-0.004$ ;  $p = 0.41$ , permutation test). For this cell, responses are somewhat enhanced around retinal velocities of zero during both directions of eye movements (Fig. 2*F*, magenta and cyan). This results in enhanced responses around zero depth but does not produce an asymmetry in the depth-sign tuning curve for the MP condition (Fig. 2*E*, red).

We also observed neurons that showed substantial changes in baseline activity between opposite directions of eye movements. The example neuron shown in Figure 2*G* shows a preference for near depths in the MP condition (DSDI =  $-0.7$ ,  $p < 0.001$ , permutation test), and the depth tuning curve is shifted toward higher responses, as compared with the RM condition. This results from increased responses during eye movements in the null direction (Fig. 2*H*, magenta). This baseline shift is dependent on eye movement direction, resulting in a significant positive MI value (MI = 0.23,  $p < 0.001$ , permutation test). This case lies in contrast to the (more typical) near-preferring neuron in Figure 2*B*, which has a negative MI value.

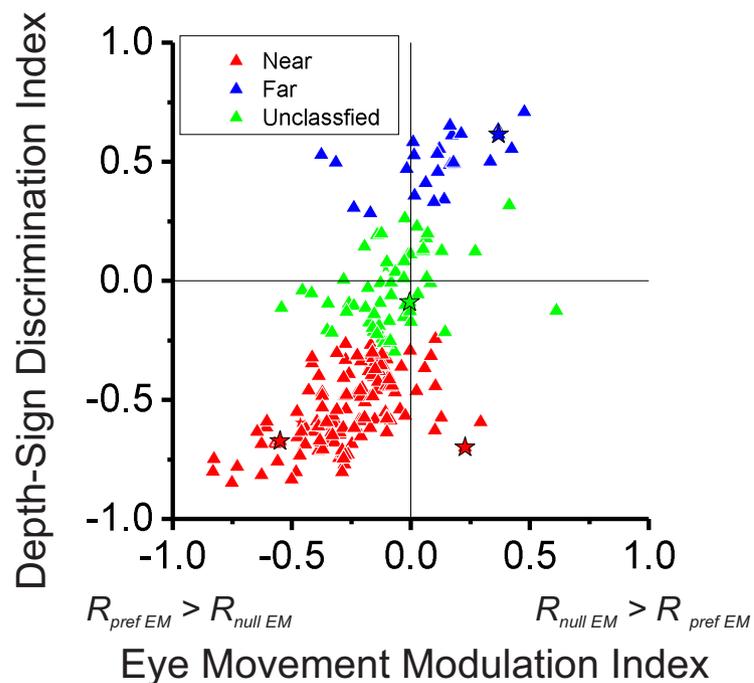
Across the population of MT neurons studied, the effect of eye movements on depth-sign tuning was generally consistent with the example neurons of Figure 2 (which are indicated by stars in

Fig. 3). Near-preferring neurons (Fig. 3, red) tend to show weaker responses when the eye moves toward the null direction than the preferred direction. The median MI value for near cells is  $-0.25$ , which is significantly less than zero ( $p = 9.5 \times 10^{-22}$ , Wilcoxon signed rank test,  $n = 138$ ). In contrast, most far-preferring neurons (Fig. 3, blue) show weaker responses when the eye moves toward the preferred direction than the null direction, and the median MI value (0.11) is significant greater than zero ( $p = 0.011$ , Wilcoxon signed rank test,  $n = 29$ ). For neurons with nonsignificant depth-sign tuning (Fig. 3, green), responses during null direction eye movements are on average slightly weaker than responses during preferred eye movements. The median MI for these unclassified neurons ( $-0.08$ ) is small but significantly different from zero ( $p = 0.001$ , Wilcoxon signed rank test,  $n = 64$ ). Across all neurons, DSDI values are strongly correlated with MI values (Spearman rank correlation,  $r = 0.67$ ,  $p < 1 \times 10^{-31}$ ,  $n = 231$ ), and this effect was consistent across the two animals (Monkey 1:  $r = 0.71$ ,  $p < 10^{-31}$ ,  $n = 155$ ; Monkey 2:  $r = 0.59$ ,  $p = 1.9 \times 10^{-8}$ ,  $n = 76$ , Spearman rank correlations). These results suggest that direction-dependent eye movement modulation could drive the depth-sign tuning that previous studies have found (Nadler et al., 2008, 2009).

Retinal image motion is the same between the MP and RM conditions, assuming that the animals pursue the fixation target accurately in the MP condition. Although the animals were very highly trained to pursue these targets and made few catch-up saccades, some retinal slip is inevitable. Could the direction-specific effects of eye movements on MT responses be caused by retinal slip? To address this question, we measured the average retinal slip (across trials) for each distinct stimulus condition, and then computed the predicted neural response modulation caused by retinal slip from the retinal velocity tuning in the RM condition. For a subset of neurons (219/231) whose speed tuning curves were obtained, a modulation index was computed from these predicted responses to retinal slip, and these values were compared with the MI values of Figure 3. The median absolute value of the predicted modulation index based on retinal slip (0.03) was much less than the median absolute MI (0.17) computed by Equation 2, and this difference was highly significant ( $p < 10^{-30}$ , Wilcoxon signed rank test,  $n = 219$ ). Moreover, predicted modulation indices based on retinal slip were not significantly correlated with the MI values of Figure 3 ( $r = 0.02$ ,  $p = 0.82$ , Spearman rank correlation,  $n = 219$ ); thus, the eye movement modulations that we have described above cannot be attributed to retinal slip.

#### Gain modulation as a mechanism for generating depth-sign selectivity

We further investigated the nature of response modulation by eye movements, to examine whether the observed modulation acts mainly as a multiplicative gain change or an additive offset change (or some combination of the two). Multiplicative gain

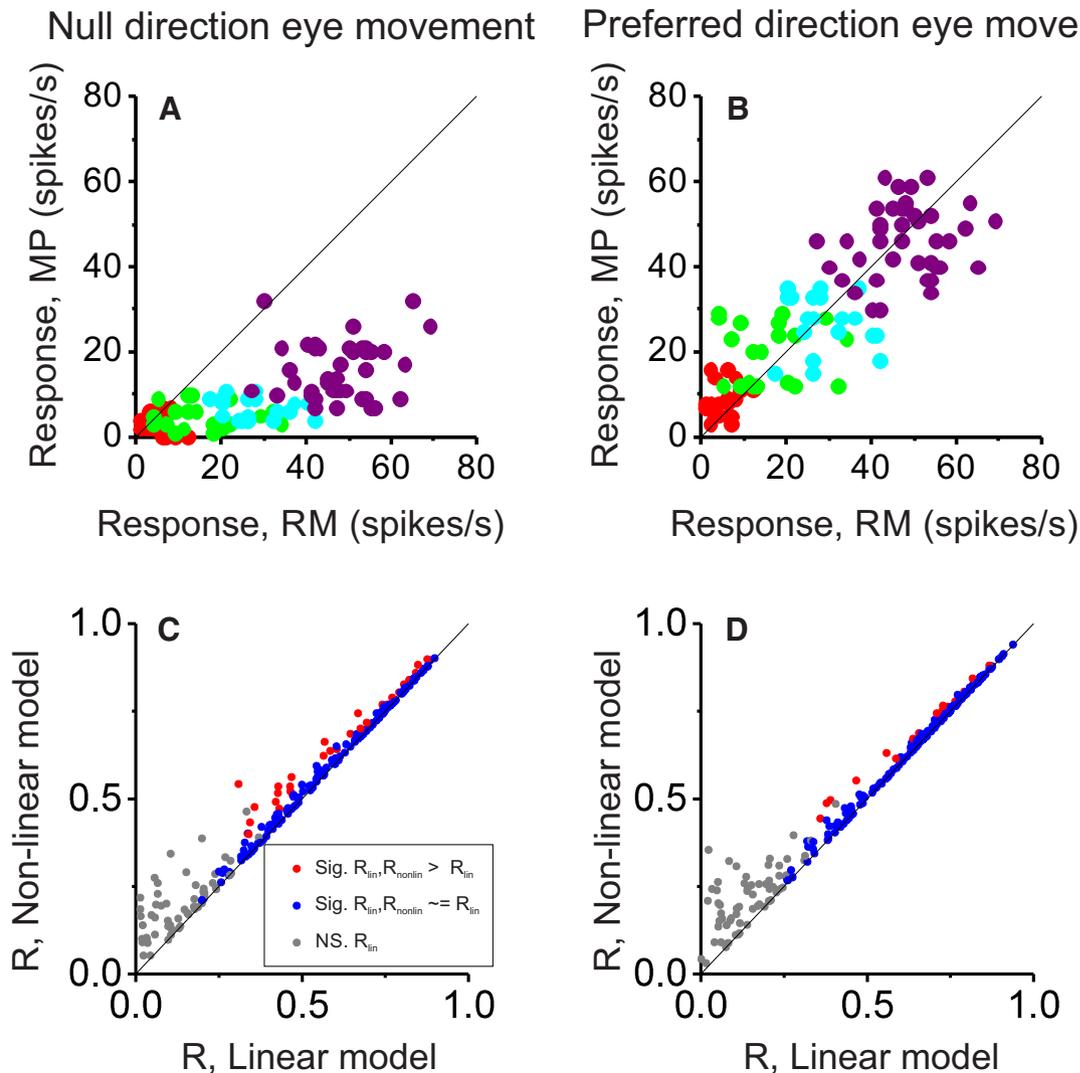


**Figure 3.** Relationship between depth-sign selectivity and eye movement modulation. For each neuron, DSDI was computed from the depth tuning curve in the MP condition, and MI was computed from the retinal velocity tuning curves for preferred and null direction eye movements. The two metrics are strongly correlated ( $r = 0.67$ ,  $p < 1 \times 10^{-31}$ , Spearman rank correlation,  $n = 231$ ). Responses of near-preferring neurons tend to be greater for eye movements toward the neuron's preferred direction, whereas responses of far-preferring neurons tend to be greater for eye movements toward the neuron's null direction.

modulation (for review, see Salinas and Thier, 2000) has been proposed as a mechanism of coordinate transformation (Andersen et al., 1985; Zipser and Andersen, 1988) and attentional modulation (McAdams and Maunsell, 1999), whereas additive interactions have also been demonstrated in cortical (Morgan et al., 2008) and subcortical (Eshel et al., 2015) neurons. Thus, we asked whether the depth-sign selectivity generated by eye movement signals depends on a multiplicative (gain) interaction, an additive (offset) effect, or both.

To quantify the nature of response modulations by eye movements, single-trial responses during eye movements (MP condition) were plotted against responses without eye movements (RM condition), and the data were fit with a linear model using type II regression. The linear model is a first-order polynomial (see Materials and Methods; Eq. 3), such that eye movements can have both additive and multiplicative effects on firing rate. Figure 4A shows data from an example neuron for eye movements in the null direction. The slope (0.28) of the linear fit is significantly less than unity [95% CI = (0.23, 0.34), bootstrap method], and the intercept (1.1 spikes/s) is not significantly different from zero [95% CI = (-0.41, 2.52), bootstrap]. Thus, the effect of null direction eye movements for this neuron is well described by a reduction in response gain. In contrast, for eye movements toward the neuron's preferred direction, data cluster around the unity slope diagonal (Fig. 4B). Correspondingly the slope of the best-fitting linear model (0.84) is close to unity [95% CI = (0.74, 0.95)], and there is also a modest but significant intercept [5.3 spikes/s, 95% CI = (2.30, 8.13)]. Thus, the depth-sign selectivity of this neuron arises mainly from a selective response gain reduction during null direction eye movements.

In general, the linear model provided adequate fits to the responses of MT neurons. To evaluate whether a more complex model would fit the data substantially better, we also fit a nonlinear model, a second-order polynomial that includes a quadratic



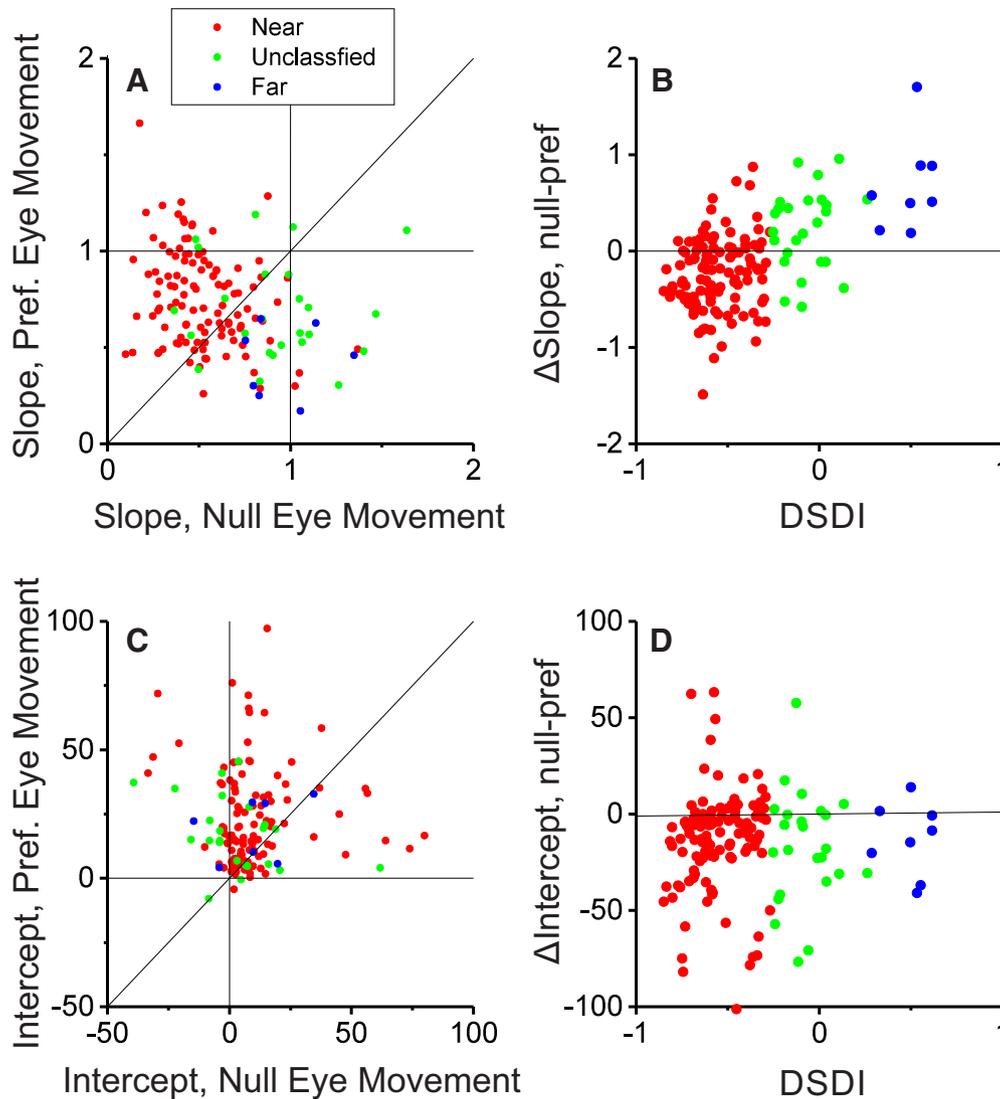
**Figure 4.** Modulation of responses by eye movements are well described by a linear model. **A**, Firing rates from individual trials of the MP condition with null direction eye movements are plotted against responses from the RM condition for an example neuron. Slope (gain) and intercept (offset) of a linear fit were obtained by type 2 regression. Colors indicate different depth magnitudes. **B**, Data from the same example neuron when the eye moves toward the preferred direction in the MP condition. **C**, Variance accounted for by a nonlinear model with a quadratic term (see Materials and Methods) is plotted against variance accounted for by a linear model fit. Data are shown for conditions in which the eye moved in each neuron's null direction. Gray symbols denote neurons that did not show significant linear fits ( $n = 55/231$ ). Among the remainder, blue data points indicate neurons for which the nonlinear model did not provide a significantly better fit than the linear model (145/176), whereas red data points indicate significantly superior fits of the nonlinear model (31/176). **D**, Same as **C** but for conditions in which the eye moves toward the preferred direction (150/166 neurons were not fit significantly better by the nonlinear model, blue).

response term (Eq. 4;  $n = 2$ ), and we compared the goodness of fit of the linear and nonlinear models. For null direction eye movements (Fig. 4C), little additional variance was accounted for by the quadratic nonlinear model (median correlation coefficient,  $R$ , for the linear model = 0.48, median  $R$  for the nonlinear model = 0.52). Among neurons that were significantly fit by the linear model (blue and red symbols), most neurons (145/176) were not fit significantly better by the more complex model (sequential  $F$  test,  $p > 0.05$ ). Results are similar for preferred direction eye movements (Fig. 4D), with a median  $R$  of 0.54 for the linear model and 0.55 for the nonlinear model. Adding the nonlinear term did not significantly improve the fits for 150/166 neurons (sequential  $F$  test,  $p > 0.05$ ).

To examine whether this finding was sensitive to the specifics of the nonlinear model used, we also fit nonlinear models in which the exponent ( $n$ ; Eq. 4) took on values ranging from 3 to 10 (in integer steps). Choosing the nonlinear model with the best fit

for each neuron had very little effect on the outcome. The median  $R$  values obtained by using the best exponent for each neuron (0.53 and 0.56 for null and preferred direction eye movements, respectively) were very similar to those obtained with the quadratic model (0.52 and 0.55, for null and preferred directions, respectively). Overall, first-order polynomial fits provide a reasonable description of the effect of eye movements on MT responses, allowing us to summarize the multiplicative and additive effects of eye movements using the slope and intercept parameters of the linear model.

To quantify the effects of eye movements across the population of MT neurons, slopes and intercepts were examined for neurons that showed significant linear fits for both directions of eye movements (Fig. 4C,D, blue and red symbols). Slopes of the linear fits are systematically related to the depth-sign tuning of MT neurons (Fig. 5A). For near-preferring neurons (Fig. 5A, red), most data points are located above the diagonal, indicating that response gain is smaller when the eye moves toward the



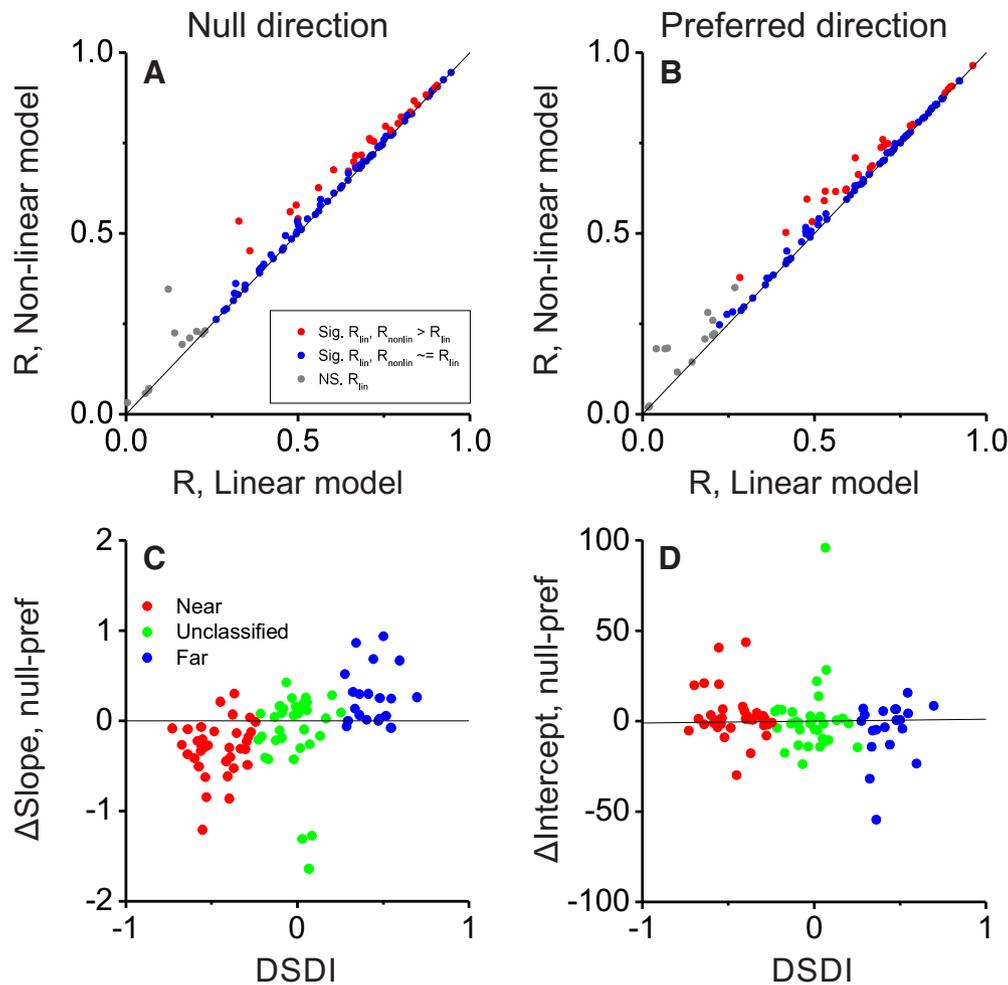
**Figure 5.** Gain modulation predicts depth-sign tuning in the MP condition. **A**, Slopes of the relationship between responses in the MP and RM conditions (Fig. 4*A, B*) are compared for preferred direction (ordinate) and null direction (abscissa) eye movements. Red, Near-preferring neurons; blue, far-preferring neurons; green, nonselective neurons. Data from three outlier neurons are not shown for display purposes, but are included in quantitative analyses. This analysis includes 149 neurons for which the linear fit was significant for both preferred and null direction eye movements. **B**, The difference in slope between null and preferred eye movement directions is significantly correlated with DSDI ( $r = 0.44$ ,  $p = 3.2 \times 10^{-8}$ , Spearman rank correlation,  $n = 149$ ). One neuron with large slope differences is not shown for display purposes, but is included in statistics. **C**, Offsets are compared between preferred and null direction eye movements. Format as in **A**. Data from three neurons are not shown for display purposes. **D**, The difference in offset between preferred and null direction eye movements is not significantly correlated with DSDI ( $r = 0.09$ ,  $p = 0.28$ , Spearman rank correlation,  $n = 149$ ). Two neurons are not shown for display purposes, but are included in statistics.

neuron's null direction. The median difference in slope between null and preferred directions is significantly less than zero for near cells (Fig. 5*B*, red; median =  $-0.22$ ,  $p = 1.8 \times 10^{-9}$ , Wilcoxon signed rank test,  $n = 117$ ).

In contrast, data for far-preferring neurons lie below the diagonal in Figure 5*A* (blue), indicating that response gain is higher when the eye moves toward the neuron's null direction. The median slope difference between eye movements in the null and preferred directions is significantly greater than zero for far cells (Fig. 5*B*, blue; median =  $0.55$ ,  $p = 0.008$ , Wilcoxon signed rank test,  $n = 8$ ). Neurons without significant depth-sign tuning show intermediate effects: the median slope difference is positive (Fig. 5*B*, green; median =  $0.34$ ,  $p = 0.02$ , Wilcoxon signed rank test,  $n = 24$ ), but significantly less than that of far-preferring neurons ( $p = 0.03$ , Wilcoxon rank sum test). Across all types of depth-sign tuning, slope difference was highly significantly correlated with DSDI (Fig. 5*B*;  $r = 0.44$ ,  $p = 3.2 \times 10^{-8}$ , Spearman rank

correlation,  $n = 149$ ), and the effect was significant for each animal separately (Monkey 1:  $r = 0.49$ ,  $p = 6.0 \times 10^{-7}$ ,  $n = 97$ ; Monkey 2:  $r = 0.33$ ,  $p = 0.02$ ,  $n = 52$ , Spearman rank correlations). The weaker correlation for Monkey 2 is likely explained by the fact that no neurons were significantly categorized as far cells for this animal. Overall, these findings suggest that direction-dependent changes in response gain may be the source of depth-sign selectivity.

The same analysis was done for the intercepts of the fits, to examine whether additive changes in responses caused by eye movements also correlate with depth-sign selectivity. For all three groups of neurons (near, far, and unclassified), the intercept tends to be greater when the eye moves toward the neuron's preferred direction (Fig. 5*C*), and some neurons have substantial additive components of response that depend on eye movement direction. The median difference between intercepts for null and preferred direction eye movements is significantly less than zero



**Figure 6.** Gain modulation also predicts depth-sign selectivity based on dynamic perspective cues. **A, B**, Goodness of fit of the linear and nonlinear models are compared for simulated eye movements in the null (**A**) and preferred (**B**) directions. Format as in Figure 4C,D. The linear model did not provide a significant fit for 11/103 neurons for the null direction and 13/103 neurons for the preferred direction (gray). Among the remainder, the nonlinear model did not provide a significantly better fit than the linear model for most neurons (68/92 for the null direction and 66/90 for the preferred direction, blue). **C**, The difference in slope between preferred and null directions of simulated eye rotations is significantly correlated with DSDI ( $r = 0.61$ ,  $p < 10^{-10}$ , Spearman rank correlation,  $n = 89$  neurons with significant linear fits for both preferred and null direction simulated eye movements). Format as in Figure 5B. Data from one outlier neuron is not shown for display purposes, but is included in statistics. **D**, The difference in offset between preferred and null eye movements is not significantly correlated with DSDI ( $r = -0.13$ ,  $p = 0.23$ ,  $n = 89$ ). Format as in Figure 5D.

for near cells (median =  $-7.5$  spikes/s,  $p = 1.9 \times 10^{-9}$ , Wilcoxon signed rank test) and unclassified cells (median =  $-19.2$  spikes/s,  $p = 0.003$ ), but not for far cells (median =  $-11.5$  spikes/s,  $p = 0.11$ ; Fig. 5D). There is no significant correlation between intercept difference and DSDI across the population of all neurons (Fig. 5D;  $r = 0.09$ ,  $p = 0.28$ , Spearman rank correlation,  $n = 149$ ), nor for either animal separately (Monkey 1:  $r = 0.13$ ,  $p = 0.22$ ,  $n = 97$ ; Monkey 2:  $r = 0.00$ ,  $p = 0.99$ ,  $n = 52$ ). Thus, although eye movements cause both multiplicative and additive changes in MT responses, we conclude that the multiplicative changes mainly contribute to depth-sign selectivity.

#### Gain modulation also underlies depth-sign tuning based on dynamic perspective cues

Thus far, we have examined the effect of smooth eye movements on the responses of MT neurons by comparing responses in the MP and RM conditions. Recently, however, we also showed that many MT neurons exhibit depth-sign selectivity when a large visual background stimulus simulates eye rotation, in the absence of real eye movements (Kim et al., 2015b). In this case, we argued

that the brain can infer eye rotation based on perspective distortions in the global pattern of image motion, which we call DP, and can use those signals to generate depth-sign selectivity (Kim et al., 2015b). Thus, it is of considerable interest to know whether the effects of DP cues on the responses of MT neurons are similar to the effects of actual smooth eye movements. We therefore analyzed responses of MT neurons in the DP condition using the same linear model fits described above.

Again, we found that the linear model (first-order polynomial) provided a good fit to responses for the DP condition, and that the quadratic nonlinear model accounted for little additional variance (median  $R$  for null direction eye movement: 0.63 for the linear fit and 0.65 for the quadratic nonlinear fit; preferred direction eye movement: 0.62 for the linear fit and 0.63 for the nonlinear fit; Fig. 6A, B). Using the second order polynomial fit (Eq. 4) with the best exponent in the range from 2 to 10 did not change the  $R$  values appreciably (0.65 for null direction and 0.64 for preferred direction eye movements, respectively). Thus, we again used the slopes and intercepts from the linear model to characterize the multiplicative and additive effects of simulated eye rotations on MT responses.

For neurons with significant fits of the linear model, we found that the difference in slope between simulated eye movements in the null and preferred directions was highly predictive of depth sign preferences (Fig. 6C;  $r = 0.61$ ,  $p < 10^{-10}$ , Spearman rank correlation,  $n = 89$ ; Monkey 1:  $r = 0.43$ ,  $p = 0.006$ ,  $n = 40$ ; Monkey 2:  $r = 0.67$ ,  $p = 3.2 \times 10^{-7}$ ,  $n = 49$ ). In contrast, we found that differences in intercepts were not correlated with depth-sign selectivity in the DP condition (Fig. 6D;  $r = -0.13$ ,  $p = 0.23$ , Spearman rank correlation,  $n = 89$ ; Monkey 1:  $r = -0.25$ ,  $p = 0.13$ ,  $n = 40$ ; Monkey 2:  $r = -0.08$ ,  $p = 0.60$ ,  $n = 49$ ).

These results show that dynamic perspective cues modulate the responses of MT neurons to generate depth-sign selectivity in a very similar manner to actual pursuit eye movements. This finding suggests that both visual and extra-retinal signals regarding eye rotations may be represented in a similar format, such that they can generate depth-sign tuning using a common mechanism. This result also reinforces our conclusion that the results from the MP condition are not substantially influenced by retinal slip during pursuit eye movements, as we see a very similar pattern of results in the DP condition, for which there is no pursuit.

### Responses to noise dots reveal rapid time course of modulation

It is clear that smooth eye movements modulate the responses of MT neurons in a directionally specific manner. However, the effects of eye velocity can generally only be seen during time periods when there is visual motion in the neuron's preferred direction, as otherwise responses tend to be strongly suppressed during visual motion in the null direction. As a result, it is difficult to appreciate the time course over which eye movements modulate neural responses.

To more clearly reveal the dynamics of eye movement modulation, we also analyzed data from experiments in which the random-dot stimulus within the receptive field was a 3D cloud of dots (Fig. 7A), with simulated depths ranging from  $-2^\circ$  to  $+2^\circ$  of equivalent disparity. When this stimulus is viewed during sinusoidal translation of the observer, it contains bidirectional retinal image motion throughout the course of a trial. At each point in time, half of the dots moved toward the neuron's preferred direction, while the remainder moved toward the null direction (Fig. 7B). Thus, there were some dots moving in the preferred direction to activate the neuron throughout each trial (Fig. 7B, shaded area), allowing us to better visualize the modulatory effects of eye movements.

Figure 7C illustrates results of 3D cloud stimulation for a near-preferring MT neuron. When there was no eye movement (RM condition), responses of this neuron showed three clear peaks (Fig. 7C, left column) that resembled the pattern of retinal image motion in the preferred direction (Fig. 7B). Correspondingly, the eye movement modulation index was close to zero ( $MI = -0.1$ ). Strikingly, however, when the animal moved their eyes toward the neuron's null direction of motion in the MP condition (Fig. 7C, right column, magenta shading), responses were powerfully suppressed ( $MI = -0.73$ ). In contrast, when the eye moved toward the preferred direction (Fig. 7C, blue shading), responses were nearly identical to those in the RM condition. Thus, this directionally specific suppression by eye movement has a rapid time course, such that it can effectively abolish a neuron's response during a brief period of time.

The strength of eye movement modulation in response to the noise stimulus was quantified using a form of the modulation index (with no averaging across depths). MI values for the MP condition vary widely between  $-1$  and  $1$  (Fig. 7D, red) and most

of the MI values are significantly different from zero (Fig. 7D, filled red bars). MI values for the DP condition are also distributed widely (Fig. 7D, blue). In contrast, MI values for the RM condition are largely clustered around zero (Fig. 7D, black). MI values from noise stimuli are highly correlated with those from 100% coherence depth stimuli ( $n = 95$ ,  $r = 0.87$ ,  $p < 10^{-30}$  for the MP condition,  $n = 37$ ,  $r = 0.68$ ,  $p < 10^{-5}$  for the DP condition, Spearman correlation, data not shown).

To further examine the time course of eye movement modulation, we computed average PSTHs. Results for the RM condition show responses with three clear peaks (Fig. 7E, F, light gray). We then computed differential PSTHs between the two phases of platform motion to reveal the time course of response modulation associated with opposite directions of (real or simulated) eye movement. For the RM condition, this differential PSTH hovers around zero, as expected (Fig. 7E, dark gray). In contrast, the differential PSTH for the MP condition demonstrates rapid response modulations (Fig. 7E, red). Note that the zero crossings of the differential PSTH for the MP condition (red) are well aligned with the troughs in the RM response (light gray), indicating that the eye-direction-dependent response modulations are closely in phase with the visual responses. Results for the DP condition (Fig. 7F) show a very similar pattern of temporal modulations. These findings demonstrate that response modulations induced by real and simulated eye movements are temporally coincident with the responses to visual motion.

### Decoding depth sign and magnitude from a population of gain-modulated neurons

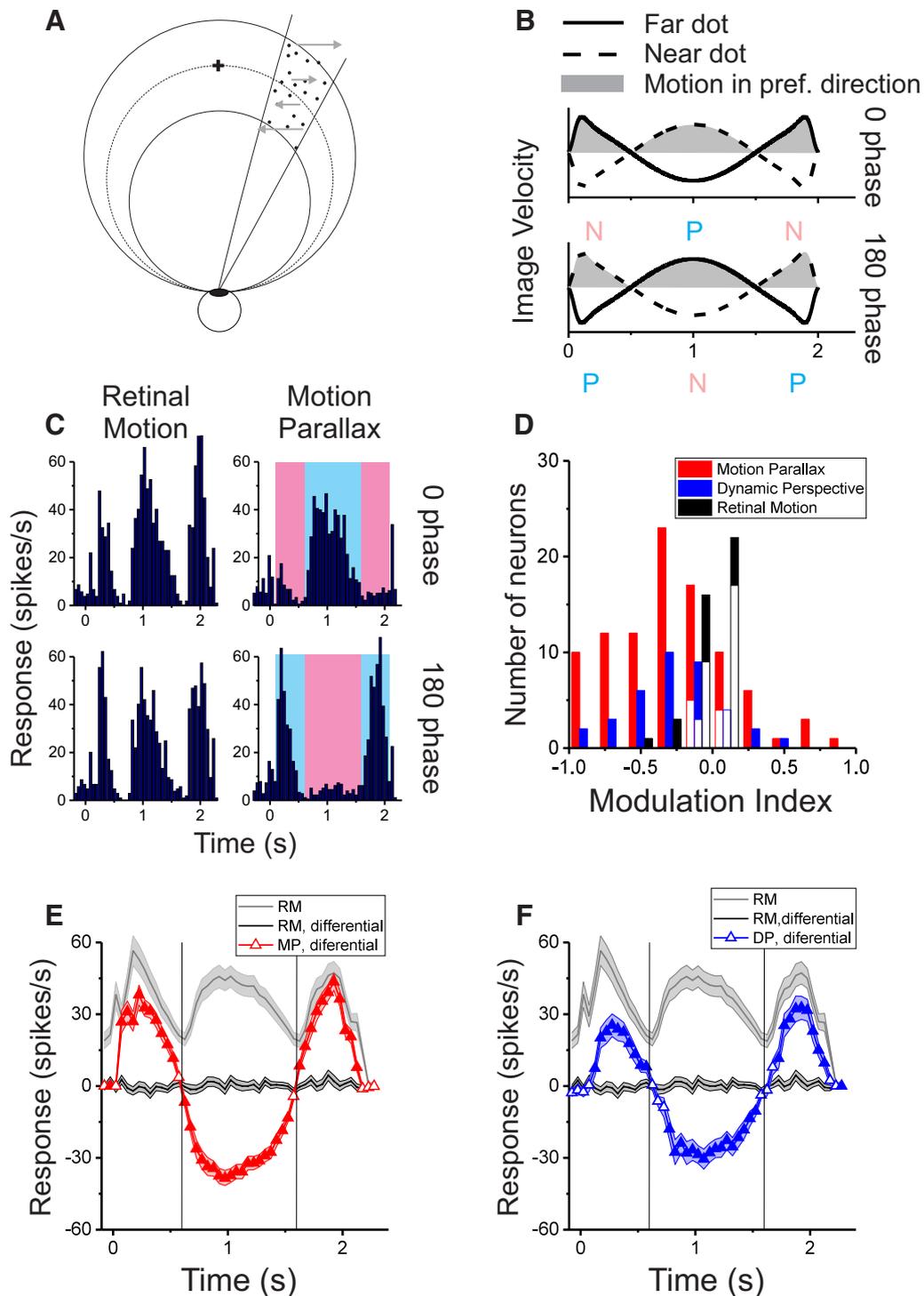
Our previous physiological studies have focused on the ability of neurons and monkeys to discriminate the *sign* of depth, that is near versus far relative to the fixation distance (Nadler et al., 2008, 2009, 2013; Kim et al., 2015a,b). However, psychophysical studies have shown that humans can judge both depth sign and magnitude based on MP (Rogers and Graham, 1979; Ono et al., 1986), and that perceived depth magnitude depends on the ratio of retinal velocity to eye velocity (Nawrot and Stroyan, 2009). Thus, a gap in our understanding of the neural basis of depth perception involves the question of how the brain computes depth *magnitude* from MP.

Gain modulation is well established as a mechanism for implementing coordinate transformations, which often involve representing the sum or difference of two variables (Andersen et al., 1985; Zipser and Andersen, 1988). For example, a population of neurons with Gaussian, eye-centered receptive fields can represent the head-centered position of a target if the neural responses are gain modulated by eye position (Salinas and Thier, 2000). This suggests that gain-modulated neurons can generally represent a linear combination of two variables: one that is represented by bell-shaped tuning curves and another that gain modulates responses in a monotonic fashion. Now consider that a simple logarithmic transformation of the motion/pursuit law (Nawrot and Stroyan, 2009) reveals that it can be expressed as a difference of variables:

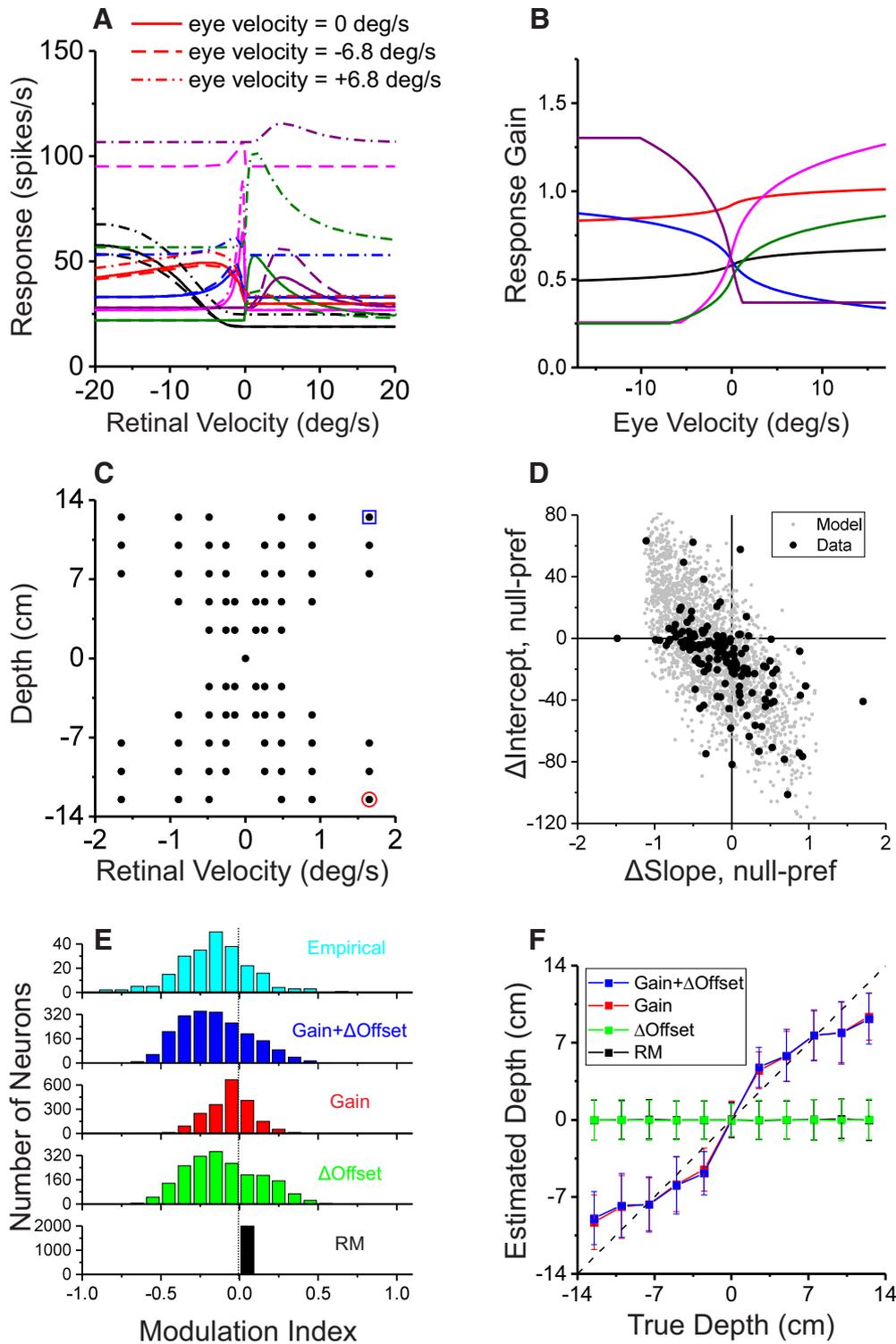
$$\frac{d}{f} \cong \frac{d\theta/dt}{d\alpha/dt} \quad (12)$$

$$\log\left(\frac{d}{f}\right) \cong \log\left(\frac{d\theta/dt}{d\alpha/dt}\right) = \log(d\theta/dt) - \log(d\alpha/dt),$$

where  $d$  is depth relative to fixation,  $f$  is the fixation distance,  $d\theta/dt$  is retinal image velocity, and  $d\alpha/dt$  represents eye velocity



**Figure 7.** Time course of eye movement modulation revealed by random-depth stimuli. **A**, A subset of neurons were tested with stimuli in which dots were uniformly distributed over a range of depths. **B**, Near and far dots moved in opposite directions, such that preferred direction motion was present throughout the course of a trial. **C**, In the RM condition, responses of an example neuron show three clear peaks (left column;  $MI = -0.1$ ). Strikingly, when the eye moved toward the null direction, responses were strongly and rapidly suppressed (bottom row;  $MI = -0.73$ ), whereas there was little or no effect when the eye moved in the preferred direction. **D**, Histogram of MI values for noise stimuli in the MP, DP, and RM conditions. Filled (open) bars denote MI values that are significantly (not significantly) different from zero. **E**, Time course of average PSTHs across neurons that have MI values significantly different from zero. Responses from the two phases of motion are combined for the RM condition (gray curve). Differential PSTHs between the two phases of motion in the RM condition were computed by subtracting responses to 180 phase from responses to 0 phase (black curve). For the MP condition, PSTHs for 180 phase were subtracted from PSTHs for 0 phase when MI values were negative, and vice versa when MI values were positive (red curve). Filled triangles denote data points that are significantly different from the differential response in the RM condition ( $n = 86$  for the MP condition;  $n = 42$  for the RM condition; permutation test using ROC analysis). **F**, Differential PSTH for the DP condition (blue curve). Format as in **E**. Nonsignificant data points around the inversions are more common than in the MP condition due to the smaller number of neurons involved ( $n = 30$  for the DP condition;  $n = 42$  for the RM condition).



**Figure 8.** Depth can be decoded from a simulated population of gain-modulated neurons. **A**, Speed tuning curves of six example model neurons. Dashed and dotted curves show modulated versions of the speed tuning curves for eye movements of a particular speed. **B**, The response gain of each model neuron is a function of the logarithm of eye velocity, with lower and upper bounds that vary across neurons (see Materials and Methods). Colors correspond to the identities of the model neurons in **A**. **C**, The set of depths and image speeds that comprised the stimulus set used to train a linear decoder of model neuron responses (see Materials and Methods). **D**, Intercept change is plotted against slope change for each neuron, revealing a negative correlation; two outlier data points were not shown for display purposes but are included in statistics. **E**, Histograms of MIs for the empirical dataset (cyan), as well as for model neurons based on four variants of the model: gain and offset model (blue), gain only model (red), offset only model (green), and the RM condition (black). **F**, Performance of the linear decoder. Estimated depths are plotted against true depths for each model. Color coding as in **E**. Error bars represent SD.

relative to the scene. By analogy to the coordinate transformation problem, a population of MT neurons should be able to represent  $\log\left(\frac{d}{f}\right)$  if the neurons have log-Gaussian speed tuning, which has been shown to be approximately true (Nover et al., 2005), and if their responses are gain-modulated by the log of eye velocity.

To explore whether a population of gain-modulated MT neurons can represent depth sign and magnitude, we modeled population responses (see Materials and Methods). Each model neuron had log-Gaussian speed tuning (Fig. 8A) and speed preferences of the population were distributed uniformly on a log speed axis, similar to what has been described previously for MT neurons (Nover et al., 2005). One-half of the model neurons prefer rightward motion directions (neurons with positive speed preferences, Fig. 8A), whereas the other half of neurons prefer leftward directions (negative speed preferences; Fig. 8A). Response gain was assumed to be proportional to log eye velocity with lower and upper bounds, and different model neurons had different proportionality constants (Fig. 8B; see Materials and Methods). The bounds on the gain functions (Fig. 8B) were chosen to approximately reflect the range of gains observed in our dataset (Fig. 5A); however, considerable variation of these bounds had modest effects on the results. Responses of model neurons on individual simulated trials were generated from independent Poisson distributions (see Materials and Methods).

In the simulation, visual stimuli were located at various depths ( $d/f$  ratios ranging from  $-0.25$  to  $0.25$ ), and had a range of retinal image velocities ( $-1.65$  to  $1.65$  deg/s). For each combination of depth and retinal image velocity (Fig. 8C), eye velocity was determined by the motion/pursuit law, and each stimulus condition was presented 1000 times. Importantly, many combinations of depth and eye velocity produce the same retinal image velocity (Fig. 8C), and stimuli at any given depth can have either positive (rightward) or negative (leftward) retinal velocities. As a result, the network cannot simply estimate depth magnitude based on retinal image motion.

We observed a highly significant negative correlation between the slope change and intercept change (Fig. 8D, black) in our dataset ( $r = -0.65$ ,  $p < 10^{-10}$ , Spearman rank correlation,  $n = 149$ ). Thus, we generated model neuron responses according to a bivariate Gaussian, to mimic this relationship (Fig. 8D, gray; see Materials and Methods). As a result, model neurons exhibit eye-velocity-dependent gain modulations and eye-direction-dependent offset changes similar to real MT neurons (Gain +  $\Delta$ Offset model). The resulting distribution of MI values computed from model neurons (Fig. 8E, blue) resembles the distribution of empirical MI values (Fig. 8E, cyan), even though model parameters were not chosen to try to match the MI distributions. Note also that performance of the model was very similar if the negative correlation between slope and intercept changes was removed; thus, this correlation is not essential to the functioning of the model.

To examine the contributions of gain and offset variations to population performance, we also tested two variants of the main model. In the Gain model, the eye-direction-dependent offset was set to zero (Fig. 8E, red). In the  $\Delta$ Offset model, gain modulation was disabled by setting the gain to one for all eye velocities (Fig. 8E, green). Note that the Gain model showed a narrower distribution of MI values, whereas the  $\Delta$ Offset model produced an MI distribution similar to the Gain +  $\Delta$ Offset model. We also tested a model without any eye movement-related response modulation (Fig. 8E, black).

We used a simple linear decoder to estimate depth from population responses (see Materials and Methods). One-half of the trials were used to fit the readout weights of the decoder, and the remaining trials were used to measure performance. Decoder weights were trained separately for each model variant described above. Figure 8F shows that the decoder can estimate depth with reasonable accuracy when responses have both gain and offset modulations (blue curve). When the model neurons have gain modulations without offset changes, performance of the decoder is essentially unchanged (Fig. 8F, red), indicating that offsets that depend on eye direction are not necessary for decoding. Furthermore, when model neurons have only offset changes without gain modulations, the decoder completely fails to estimate depth (Fig. 8F, green), similar to the case in which model responses lack both gain and offset variations (Fig. 8F, black squares). Thus, gain modulations with eye direction, but not offset changes, are sufficient to estimate depth magnitude and sign.

These results show that when responses are modulated by eye movements in a manner similar to what we observe in area MT, population activity can represent both depth sign and magnitude based on MP cues. This format of interaction between retinal motion and eye velocity allows depth estimates to be extracted by a simple linear decoder. Note that we also tested a model in which responses to retinal motion are additively modulated by the log of eye velocity. In this case, the decoder fails to estimate depth, similar to the  $\Delta$ Offset model (data not shown). This suggests that gain modulations are a critical component for estimating depth from a population of MT responses.

## Discussion

We investigated how real and simulated eye movements modulate the responses of neurons in area MT to generate selectivity for depth from MP. Responses to image motion were modulated by eye movements in a temporally precise, directionally specific manner. The sign and magnitude of these modulations were highly predictive of the depth-sign preferences of MT neurons. Although response modulations by eye movements had both multiplicative and additive components, multiplicative gain changes predicted the depth-sign preferences of MT neurons, whereas the additive component did not. Finally, our biologically constrained simulations suggest that gain modulations can provide a simple mechanism to compute both depth sign and magnitude from the responses of a population of MT neurons.

A limitation of the present study is that our single-neuron analyses characterize the effects of eye movement direction on MT responses, but not the full effect of eye velocity (direction and speed). It will be valuable to extend this work to model MT responses as a joint function of both retinal velocity and eye velocity. Experiments that include additional speeds of self-motion may help to constrain such a model of the joint velocity tuning of MT neurons.

### Relationship to previous studies of pursuit eye movements in MT

Several previous studies have investigated the effects of smooth-pursuit eye movements on responses of MT neurons. MT neurons with foveal receptive fields, like MST neurons, responded while monkeys pursued a small moving target across a blank background (Newsome et al., 1988). MT responses were greatly reduced or abolished when retinal image motion was eliminated by blinking off or stabilizing the target during pursuit (Newsome et al., 1988), or when trained animals tracked an imaginary target (Ilg and Thier, 2003). In contrast, MST activity was not substan-

tially reduced by these manipulations. These results suggested that some MT neurons respond to the retinal slip produced during pursuit, but were not directly driven by extra-retinal signals related to pursuit (e.g., efference copy). Thus, these previous results might appear to be in conflict with our findings. However, these findings can be largely reconciled by our observation that MT responses to retinal motion are predominantly gain modulated by eye movements. The multiplicative nature of this interaction means that pursuit eye movements alone, without retinal image motion to elicit robust responses, are generally not sufficient to modulate the responses of MT neurons.

Is there previous evidence for gain modulation of MT responses by pursuit eye movements? Two previous studies have examined the effects of pursuit on the responses of MT neurons to retinal motion (Chukoskie and Movshon, 2009; Inaba et al., 2011). Chukoskie and Movshon (2009) used a step-ramp paradigm to measure the speed tuning of MT neurons while animals pursued targets that moved along the cardinal axes. Their pursuit speed was faster (mainly 20 deg/s) than ours (average pursuit speed of 6.9 deg/s; peak speed of 13 deg/s), and was constant over the course of a trial. Thus, direct comparison of results is somewhat difficult. However, a representative example neuron from Chukoskie and Movshon (2009, their Fig. 3, MT neuron 1) clearly shows changes in response gain that are dependent on the direction of pursuit. Responses of this example neuron to the preferred direction of motion appear to have a response gain for null direction eye movements that is approximately one-half of that for pursuit in the preferred direction. Overall, Chukoskie and Movshon (2009) found no consistent effect of pursuit on the responses of a population of MT neurons: responses could be either enhanced or suppressed by pursuit in the preferred or null directions. However, it is difficult for us to compare their results with ours because we do not know the depth-sign preferences of the individual neurons in their study. Ignoring depth-sign preference, our data also show a range of effects in both directions (Fig. 5B); thus, pursuit direction would also have less systematic effects on response gain in our study if all neurons were combined together regardless of depth-sign tuning. Thus, we suspect that the findings of Chukoskie and Movshon (2009) are largely consistent with ours.

Inaba et al. (2007, 2011) used full-field random-dot stimuli to probe the speed and direction tuning of MT and MST neurons. Although their representative cells (Inaba et al., 2007, their Fig. 3F, 2011, their Fig. 5D) do not show much response amplitude modulation, their population data (Inaba et al., 2011, their Fig. 9J) indicate that a large fraction of MT neurons show either enhancement or suppression of responses, with suppression being generally stronger when the eye moves toward the null direction of the neuron. This finding is consistent with our overall pattern of results (Fig. 5B), given that we find substantially more near-preferring neurons than far-preferring cells.

Although these previous studies provided clear evidence that the responses of MT neurons are modulated by smooth pursuit eye movements, the functional role of these pursuit signals in MT has remained somewhat unclear. Our study supports one clear computational purpose of pursuit modulations in MT: extra-retinal signals regarding pursuit provide one important source of eye rotation relative to the scene that is necessary to compute depth from MP (Nawrot and Stroyan, 2009). The major advance of the present study is to reveal the specific rules by which eye movement signals modulate neural responses to carry out this computation.

### Multiple sources of eye rotation information for disambiguating depth sign

The motion/pursuit law demonstrates that eye velocity relative to the scene is the critical variable for computing depth sign from MP (Nawrot and Stroyan, 2009). When the head does not rotate relative to the scene, eye rotation relative to the scene is the same as eye rotation relative to the head. Thus, in the absence of scene-relative head rotation, it is sensible that extra-retinal signals related to pursuit eye movements should disambiguate depth sign. However, pursuit signals on their own should be insufficient when there is substantial head rotation relative to the scene. Under these more general conditions, there are two possible ways that an estimate of eye velocity relative to the scene can be derived. First, because optic flow is determined by translation and rotation of the eye relative to the scene, the rotational component of optic flow can be used to disambiguate depth sign (Rogers, 2016). Indeed, we recently reported that dynamic perspective cues in optic flow can generate depth-sign selectivity in MT neurons, in the absence of actual eye movements (Kim et al., 2015b). Here, we demonstrate that dynamic perspective cues modulate the responses of MT neurons in a similar manner to extra-retinal pursuit signals. Thus, it is possible that visual and extra-retinal signals regarding eye rotation are integrated to provide a generalized rotation signal for computing depth from MP.

A second possible way to compute eye rotation relative to the scene when head rotations are present is to combine multiple extra-retinal signals to compute eye velocity relative to the scene. This would potentially require vestibular signals regarding head rotation as well as efference copy of commands for active head rotations. Thus, when reliable optic flow is not available, a testable prediction is that pursuit command signals would be combined with other extra-retinal signals regarding head rotation to estimate eye velocity relative to the scene and compute depth. Depth perception based on MP may, therefore, provide an excellent model system for exploring how multiple visual and extra-retinal signals are combined to perform computations that require an estimate of eye velocity relative to the scene.

### Comparison with human psychophysics

Nawrot et al. (2014) showed that human perception of depth magnitude based on MP follows a modified power-law version of the motion-pursuit law that has much smaller exponents on the eye velocity term ( $da/dt$ ) and the retinal velocity term ( $d\theta/dt$ ) than the theory predicts. As a result, depth magnitude estimates were substantially compressed toward the fixation target. Although our simple model also predicts some depth compression (Fig. 8F), it is substantially less than the compression observed psychophysically. How might we reconcile the underestimation of depth magnitude in humans with results from our simple decoding model?

Suboptimal decoding might contribute to perceptual underestimation of depth magnitude. Under natural conditions in which the scene is viewed binocularly, both binocular disparity and MP cues are available when an observer moves, whereas only binocular disparity cues are available when the observer is stationary. We have shown that depth tuning curves of MT neurons in response to combined disparity and MP cues can be quite different from depth tuning based on MP alone (Nadler et al., 2013). In particular, depth preferences of many MT neurons based on MP are opposite to their depth preferences based on binocular disparity (Nadler et al., 2013). Therefore, neural population responses to combinations of disparity and MP cues will

be different from the responses to the pure MP condition that was used in the psychophysical task of Nawrot et al. (2014). As a result, a decoder that is trained to estimate depth optimally in the natural environment may not be optimal for estimating depth from MP cues alone. The brain needs to perform marginalization to estimate depth accurately regardless of the source of depth cues (Beck et al., 2011), but such marginalization may produce biases in estimation in heterogeneous neural populations for which preferences of neurons do not match between cues (Kim et al., 2016). Thus, the brain is likely to do suboptimal decoding of responses to monocular MP stimuli.

In closing, our findings establish a basic mechanism, directionally specific modulations of response gain, by which real or simulated eye movements modulate the responses of macaque MT neurons to generate selectivity for depth sign. These findings establish a specific computational function for eye velocity signals in area MT and suggest that both visual and extra-retinal signals related to eye rotation may interact with incoming retinal image motion signals in a consistent manner.

## References

- Andersen RA, Essick GK, Siegel RM (1985) Encoding of spatial location by posterior parietal neurons. *Science* 230:456–458. [CrossRef Medline](#)
- Barlow HB, Blakemore C, Pettigrew JD (1967) The neural mechanism of binocular depth discrimination. *J Physiol* 193:327–342. [CrossRef Medline](#)
- Beck JM, Latham PE, Pouget A (2011) Marginalization in neural circuits with divisive normalization. *J Neurosci* 31:15310–15319. [CrossRef Medline](#)
- Chukoskie L, Movshon JA (2009) Modulation of visual signals in macaque MT and MST neurons during pursuit eye movement. *J Neurophysiol* 102:3225–3233. [CrossRef Medline](#)
- Cumming BG, DeAngelis GC (2001) The physiology of stereopsis. *Annu Rev Neurosci* 24:203–238. [CrossRef Medline](#)
- Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, Uchida N (2015) Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* 525:243–246. [CrossRef Medline](#)
- Farber JM, McConkie AB (1979) Optical motions as information for unsigned depth. *J Exp Psychol Hum Percept Perform* 5:494–500. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Hayashibe K (1991) Reversals of visual depth caused by motion parallax. *Perception* 20:17–28. [CrossRef Medline](#)
- Henriksen S, Tanabe S, Cumming B (2016) Disparity processing in primary visual cortex. *Philos Trans R Soc Lond B Biol Sci* 371:20150255. [CrossRef Medline](#)
- Howard IP, Rogers BJ (1995) *Binocular vision and stereopsis*. New York: Oxford UP.
- 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. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Julesz B (1964) Binocular depth perception without familiarity cues. *Science* 145:356–362. [CrossRef Medline](#)
- Kim HR, Angelaki DE, DeAngelis GC (2015a) A functional link between MT neurons and depth perception based on motion parallax. *J Neurosci* 35:2766–2777. [CrossRef Medline](#)
- Kim HR, Angelaki DE, DeAngelis GC (2015b) A novel role for visual perspective cues in the neural computation of depth. *Nat Neurosci* 18:129–137. [CrossRef Medline](#)
- Kim HR, Pitkow X, Angelaki DE, DeAngelis GC (2016) A simple approach to ignoring irrelevant variables by population decoding based on multisensory neurons. *J Neurophysiol* 116:1449–1467. [CrossRef Medline](#)
- Lehky SR, Pouget A, Sejnowski TJ (1990) Neural models of binocular depth perception. *Cold Spring Harb Symp Quant Biol* 55:765–777. [CrossRef Medline](#)
- Longuet-Higgins HC (1981) A computer algorithm for reconstructing a scene from two projections. *Nature* 293:133–135. [CrossRef](#)
- Marr D, Poggio T (1976) Cooperative computation of stereo disparity. *Science* 194:283–287. [CrossRef Medline](#)
- McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 19:431–441. [Medline](#)
- Morgan ML, DeAngelis GC, Angelaki DE (2008) Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* 59:662–673. [CrossRef Medline](#)
- Nadler JW, Angelaki DE, DeAngelis GC (2008) A neural representation of depth from motion parallax in macaque visual cortex. *Nature* 452:642–645. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Nadler JW, Barbash D, Kim HR, Shimpi S, Angelaki DE, DeAngelis GC (2013) Joint representation of depth from motion parallax and binocular disparity cues in macaque area MT. *J Neurosci* 33:14061–14074. [CrossRef Medline](#)
- Naji JJ, Freeman TC (2004) Perceiving depth order during pursuit eye movement. *Vision Res* 44:3025–3034. [CrossRef Medline](#)
- Nawrot M (2003) Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. *Vision Res* 43:1553–1562. [CrossRef Medline](#)
- Nawrot M, Stroyan K (2009) The motion/pursuit law for visual depth perception from motion parallax. *Vision Res* 49:1969–1978. [CrossRef Medline](#)
- Nawrot M, Ratzlaff M, Leonard Z, Stroyan K (2014) Modeling depth from motion parallax with the motion/pursuit ratio. *Front Psychol* 5:1103. [CrossRef Medline](#)
- 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. [Medline](#)
- 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. [CrossRef Medline](#)
- Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249:1037–1041. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Parker AJ (2007) Binocular depth perception and the cerebral cortex. *Nat Rev Neurosci* 8:379–391. [CrossRef Medline](#)
- Rogers B (2016) The effectiveness of vertical perspective and pursuit eye movements for disambiguating motion parallax transformations. *Perception* 45:1279–1303. [CrossRef Medline](#)
- Rogers B, Graham M (1979) Motion parallax as an independent cue for depth perception. *Perception* 8:125–134. [CrossRef Medline](#)
- Rogers S, Rogers BJ (1992) Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Percept Psychophys* 52:446–452. [CrossRef Medline](#)
- Salinas E, Thier P (2000) Gain modulation: a major computational principle of the central nervous system. *Neuron* 27:15–21. [CrossRef Medline](#)
- Tsai JJ, Victor JD (2003) Reading a population code: a multi-scale neural model for representing binocular disparity. *Vision Res* 43:445–466. [CrossRef Medline](#)
- Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. *J Am Med Inform Assoc* 8:443–459. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)