MT neurons combine visual motion with a smooth eye movement signal to code depth sign from motion parallax

Jacob W. Nadler[1], **Mark Nawrot**[2], **Dora E. Angelaki**[1], and **Gregory C. DeAngelis**[3]

[1]Dept. of Anatomy & Neurobiology, Washington University School of Medicine, St. Louis, MO 63110

[2]Center for Visual Neuroscience, Department of Psychology, North Dakota State University, Fargo, ND 58105

[3]Dept. of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, NY 14627

Abstract

The capacity to perceive depth is critical for an observer to interact with their surroundings. During observer movement, information about depth can be extracted from the resulting patterns of image motion on the retina (motion parallax). Without extra-retinal signals related to observer movement, however, depth sign (near vs. far) from motion parallax can be ambiguous. We previously demonstrated that MT neurons combine visual motion with extra-retinal signals to code depth sign from motion parallax in the absence of other depth cues. In that study, head translations were always accompanied by compensatory tracking eye movements, allowing for at least two potential sources of extra-retinal input. We now show that smooth eye movement signals provide the critical extra-retinal input to MT neurons for computing depth sign from motion parallax. Our findings demonstrate a powerful modulation of MT activity by eye movements, as predicted by human studies of depth perception from motion parallax.

INTRODUCTION

The brain makes use of a variety of cues to compute the 3D structure of the visual scene from retinal images, with binocular disparity and motion parallax providing two of the most potent cues (Howard and Rogers, 1995). When an observer moves through the environment, information about depth can be extracted from the resulting patterns of object motion on the retina (Nakayama and Loomis, 1974; Richards, 1985; Rogers and Graham, 1979). In this context, motion parallax refers to the relative image motion (between objects at different depths) that results from observer movement. However, without extra-retinal signals related to self-movement, depth-sign (near or far) from visual motion alone can be ambiguous when pictorial depth cues such as relative size or occlusion are absent (Farber and McConkie, 1979; Hayashibe, 1991; Nawrot, 2003b; Rogers and Rogers, 1992).

Address for correspondence: Gregory C. DeAngelis, Professor of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, 245 Meliora Hall, Rochester, NY 14627, Office Phone: (585)275-8677, FAX: (585)271-3043, gdeangelis@cvs.rochester.edu.

What is the nature of the extra-retinal signal that is used to compute depth-sign from motion parallax? It could be a vestibular signal related to head movements, as suggested by some studies (Rogers and Rogers, 1992). Alternatively, it could be a smooth eye movement signal because translation of the head must be accompanied by a compensatory smooth eye movement to maintain visual fixation on a target of interest. Recent psychophysical work in humans suggests that a smooth eye movement command signal, not a vestibular signal, provides the extra-retinal signal necessary for perception of depth from motion parallax (Naji and Freeman, 2004; Nawrot, 2003a, b; Nawrot and Joyce, 2006; Nawrot et al., 2004). Thus, these human studies predict that smooth eye movements, not head movements, should modulate the responses of neurons that are involved in coding depth from motion parallax.

We have recently reported that neurons in visual area MT are selective for depth sign from motion parallax, and that this selectivity depends on the action of extra-retinal inputs to MT (Nadler et al., 2008). However, the findings of that study did not address the origin of the extra-retinal signal. Previous physiological studies have indicated that smooth pursuit eye movements do not strongly modulate responses of MT neurons (Newsome et al., 1988). If MT provides critical sensory signals for depth perception based on motion parallax, the lack of pursuit responses shown by MT neurons appears to be at odds with the human studies (described above) that implicate smooth eye movements in depth from motion parallax. A possible resolution to this quandary is that MT neurons are more strongly modulated by smooth eye movements when these movements interact with a visual stimulus in the receptive field. This notion is encouraged by the finding that extra-retinal signals interact nonlinearly with visual motion signals in MT (Nadler et al., 2008).

To examine this issue more directly, we have tested MT neurons under conditions that isolate extra-retinal signals related to head movements or eye movements, and we compare the results of these tests to data collected during the same motion parallax protocols used previously (Nadler et al., 2008). We find robust depth-sign selectivity in MT neurons when the animals perform a smooth tracking eye movement without any head movement. In contrast, we generally find no depth-sign selectivity when animals are translated without making any smooth eye movements. Thus, under the conditions of our experiment, responses of MT neurons to visual motion are powerfully modulated by a smooth eye movement command signal and not by a vestibular signal. These findings, which were predicted by human psychophysical studies (Naji and Freeman, 2004; Nawrot, 2003a, b; Nawrot and Joyce, 2006; Nawrot et al., 2004), resolve the apparent inconsistency with previous physiological studies regarding effects of smooth eye movements on MT responses. Our results support the hypothesis that MT provides sensory signals used for judging depth from motion parallax, and elucidate a neural mechanism by which signals related to self-motion modify incoming sensory signals to aid estimation of the three-dimensional structure of the environment.

RESULTS

We collected extracellular recordings from 79 well-isolated single units in area MT. No cells were discarded, even though some responded poorly at all simulated depths due to the range of speeds that occurred in the receptive field. Each MT neuron in this sample was tested under four stimulus conditions (see Methods for additional details). The first two conditions, the Retinal Motion (RM) and Motion Parallax (MP) conditions, were identical those described in a previous study (Nadler et al., 2008). In the RM condition (Fig. 1A), the motion platform was stationary and the monkey simply viewed visual motion stimuli containing motion parallax consistent with surfaces lying at several different depths relative to the plane of fixation. There was neither head movement nor eye movement in the RM condition. In the MP condition (Fig. 1B), the monkey was translated back and forth by a motion platform while maintaining fixation on a world-fixed target. Thus, this condition involved both head movements and compensatory

eye movements, and generated the same retinal image motion as experienced in the RM condition.

The last two stimulus conditions were devised to isolate head or eye movements as the source of the extra-retinal inputs to area MT. In the Head Only (HO) condition (Fig. 1C), the animal was translated by the platform as in the MP condition, but the fixation target moved with the animal such that no eye movement was required. In the Eye Only (EO) condition (Fig. 1D), the animal remained stationary while the entire virtual environment translated in front of the animal, thus requiring the same eye movement to track the fixation target as in the MP condition. Thus, if vestibular signals provide the main extra-retinal input to MT neurons, then responses in the HO condition should exhibit depth-sign selectivity similar to that seen in the MP condition, whereas the RM and EO conditions should show no depth-sign selectivity. If smooth eye movements provide the key extra-retinal input, then the EO and MP conditions should show similar depth-sign selectivity, whereas the HO and RM conditions should not. Note that all four stimulus conditions produce the same retinal image motion (assuming accurate pursuit, Nadler et al., 2008), such that differences between stimulus conditions should reflect the composition of extra-retinal signals that interact with visual motion responses in area MT.

Single-unit Examples

Figure 2 illustrates results obtained from an example MT neuron. Peri-stimulus time histograms (PSTHs) are shown for each stimulus condition (Fig. 2a–d), with responses grouped into columns according to the phase of the real or simulated movement of the observer (see Methods). Near and far simulated depths of the same magnitude have opposite retinal velocity profiles (gray curves in Fig. 2a). Retinal image velocities range from ~0°/s for a simulated depth of 0° up to ±5°/s for the largest simulated depths of ±2°. In the RM condition (Fig. 2a), neural responses follow the velocity of the retinal motion stimulus such that near and far simulated depths of the same magnitude (e.g., −2° vs. +2°) elicit nearly identical responses. Because this neuron responded best to speeds greater than 5°/s, response modulations generally became stronger as the magnitude of the simulated depth increased. When average firing rates are plotted against simulated depth, this neuron shows a depth-tuning curve that is symmetric around 0° (Fig. 2e, blue curve). Because pictorial depth cues were removed from the visual stimulus, responses under the RM condition are ambiguous with respect to depth sign (near vs. far). This was quantified using a Depth Sign Discrimination Index (DSDI, see Methods), which takes on values approaching +1 for far-preferring neurons and values approaching −1 for near-preferring neurons. In the RM condition, the DSDI for this example neuron was 0.01, not significantly different from zero (permutation test).

In the MP condition, as reported previously (Nadler et al., 2008), responses of the same MT neuron show robust depth-sign selectivity (Fig. 2b). Responses to near stimuli (e.g., −2°) were enhanced relative to the RM condition whereas responses to far stimuli (e.g., +2°) were suppressed. This is most clearly seen in the depth-tuning curve (Fig. 2e, black curve), which shows firing rates that decline monotonically from near to far. This selectivity for depth sign must arise from extra-retinal inputs to MT because the retinal image motion is the same in the MP and RM conditions (Nadler et al., 2008). The near preference of this neuron is highly significant, with a DSDI value of −0.65 in the MP condition, which is significantly different from zero (p<0.001) by permutation test.

Does the depth-sign selectivity of this MT neuron in the MP condition arise from extra-retinal signals related to head/body movement or to the corresponding smooth eye movement? In the HO condition (Fig. 2c), responses of this neuron are similar to those obtained in the RM condition, and the DSDI value is not significantly different from zero. In contrast, responses of this neuron under the EO condition (Fig. 2d) strongly resemble those measured under the

MP condition, and the DSDI value (−0.68) is highly significant (p<0.001). Thus, depth-sign selectivity of this example neuron seems to be driven by eye movements but not head movements. This can be seen clearly in the depth tuning curves (Fig. 2e) The MP (black) and EO (orange) conditions produce almost identical tuning functions, whereas data from the HO (green) condition exhibit the same even-symmetry as those from the RM condition (blue).

Data from six additional example neurons are shown in Figure 3. In all cases, tuning curves from the MP and EO conditions show robust depth-sign selectivity, whereas tuning curves for the RM and HO conditions are roughly symmetric around zero depth (which corresponds to the fixation distance). Examples in the left column have significant near preferences for depth in the MP and EO conditions (permutation test, p<0.01 in all cases), whereas examples in the right column have significant far preferences. Note that tuning curves were generally a bit shallower in the HO condition versus the RM condition. This may be due to the fact that monkeys made more small eye movements within the fixation window during the HO condition than during the RM condition. In fact, the average standard deviation of eye position (across time) was significantly larger in the HO condition (0.20±0.003 deg) than in the RM condition (0.11±0.002 deg). These additional small eye movements in the HO condition may, in part, reflect a weak translational vestibulo-ocular reflex that is not fully suppressed (see Discussion).

For 48/79 neurons, including those in Fig. 3B, D, E, and F, the four stimulus conditions were randomly interleaved within a single block of trials. For a minority of neurons (31/79), including those in Fig. 3A, C, EO and HO data were collected in a separate block of trials from the MP and RM data, thus precluding a direct comparison of firing rates between conditions for these neurons. We sometimes observed gradual changes in neural responsiveness over time, which might be related to slow changes in attentional state or arousal. Note, however, that such gradual changes in responsivity would not be expected to induce depth-sign selectivity, even for the subset of cases in which all stimulus conditions were not interleaved.

Population Results

Among 79 neurons examined in two monkeys (26 from Monkey 1, 53 from Monkey 2), 60 (76%) exhibited significant depth-sign selectivity from motion parallax in the MP condition. Filled symbols in Fig. 4A compare DSDIs for the EO condition to DSDIs obtained in the MP condition. There is a very strong correlation between DSDIs from the MP and EO conditions (r=0.94, p<0.0001; Spearman rank correlation). Of the 60 neurons with significant selectivity in the MP condition, 53 also show significant selectivity for depth sign in the EO condition, and the sign of the depth preference (near or far) is preserved in all cases. Very different results were obtained in the HO condition (Fig. 4A, open symbols). DSDI values for the HO condition tend to scatter around the horizontal axis and there is no significant correlation between DSDIs from the HO and MP conditions (r=0.00, p=0.99; Spearman rank correlation). Among 60 neurons with significant selectivity in the MP condition, only 6 show significant selectivity for depth sign in the HO condition. Moreover, 4 of these 6 neurons have an opposite depth-sign preference in the HO condition. Even when only considering the 60 selective neurons, the fraction showing significant selectivity for depth sign in the HO condition (6/60) is not significantly greater than the 5% expected by chance (p = 0.076; chi-square test). Among 19 neurons with poor depth-sign tuning in the MP condition (p > 0.05, permutation test), only 5 showed significant tuning in the EO condition and none showed significant tuning in the HO condition.

Figure 4B shows DSDI values from the EO and HO conditions plotted against the corresponding values from the RM condition. Neither data set shows a significant correlation (p > 0.05, Spearman rank correlation). The mean DSDI in the HO condition is 0.054, which is slightly but significantly greater than zero (p=0.004; t-test), while the mean DSDI in the RM

condition (0.016) is not (p=0.431; t-test). This positive bias in the HO condition only exists for one monkey (M1: p=0.212, M2: p=0.009, t-test).

The results of Fig. 4 suggest that smooth eye movement command signals, not vestibular or proprioceptive signals related to head movement, contribute to the computation of depth sign from motion parallax in MT. However, it is possible that stronger translational motion stimuli would have revealed a greater vestibular contribution (see Discussion). To examine this possibility, we tested a subset of neurons in the Head Only condition with 1.0 Hz movements. Nine cells (all from Monkey 2) were chosen for this additional study (Fig. 4C). All but one were significantly selective for depth-sign from motion parallax at 0.5Hz in the MP condition, and none showed significant selectivity in the HO condition at 0.5Hz (open symbols). At 1.0 Hz (filled symbols), only two of nine cells had significant DSDIs in the HO condition. Among these two, one showed the same depth-sign preference as in the MP condition (diamond symbols), and the other showed the opposite depth-sign preference (triangle symbols). It was not possible to test neurons at frequencies appreciably above 1 Hz (see Discussion).

Relationship to Receptive Field Parameters

How is depth-sign selectivity related to the tuning preferences and receptive field properties of MT neurons? We previously reported (Nadler et al., 2008) a significant negative correlation between DSDI in the MP condition and the speed preference of MT neurons, such that neurons preferring very slow speeds tended to have far depth-sign preferences. We suggested that this trend may reflect an ecological adaptation to the fact that near objects will generate a larger range of retinal image speeds than far objects under the viewing conditions of our experiment (and other natural viewing conditions as well). We found a very similar negative correlation with the speed preference of MT neurons in the EO condition (r=−0.54, p<0.0001; Spearman rank correlation), as expected given the tight correlation between DSDI values in the MP and EO conditions (Fig. 4A). DSDIs in the HO condition did not show a significant correlation with speed preference (r=0.02, p=0.85; Spearman rank correlation).

In contrast to the dependence of depth-sign selectivity on speed preference, we found no significant dependence of DSDI on direction preference, for either the MP or EO conditions (circular-linear correlation, p > 0.23 for both conditions). Thus, MT neurons tuned to all directions of motion showed depth-sign selectivity. This is, perhaps, not surprising given that the axis of the translational motion stimuli was tailored to the preference of each MT neuron studied, with a corresponding rotation of virtual curved surfaces that defined stimuli at different depths (as detailed in Nadler et al., 2008).

We further examined whether depth-sign selectivity was related to the location and eccentricity of MT receptive fields. If a smooth eye movement signal provides the critical extra-retinal signal, one could imagine that it would selectively target MT neurons with receptive fields located near the fovea, as these cells may signal retinal slip of the target. Figure 5 provides a graphical summary of the location and size of the receptive field for each of the MT neurons recorded in the two monkeys. As expected, receptive field diameter was robustly correlated with eccentricity (r=0.46, p<0.0001). The color of each circle in Fig. 5 represents the DSDI value in the MP condition (for which the dataset is larger, N=144). No systematic relationship is visible between the magnitude of DSDI and receptive field location or eccentricity. This was confirmed by statistical analyses showing no significant correlation between the polar angle of the receptive field center and the magnitude of DSDI in both the MP and EO conditions (circular-linear correlation, p > 0.45 for both conditions). Similarly, we found no significant correlation between DSDI magnitude and receptive field eccentricity (linear regression, p > 0.52 for both conditions).

Thus, with the exception of speed preference, depth-sign selectivity was generally not correlated with the receptive field properties of MT neurons. Rather, smooth eye movement signals appear to interact with visual motion to sculpt depth-sign selectivity throughout the visual field representation in MT. This finding is consistent with the notion that eye movement signals modulate MT responses for the purpose of depth perception, as this would be expected to take place in parallel throughout the visual field.

Eye Movements

All of the stimulus conditions used in these experiments produce the same retinal image motion as a function of simulated depth, provided that the monkey accurately maintains gaze on the fixation target during the course of the trial. In the MP and EO conditions, this requires accurate smooth pursuit eye movements. As discussed previously (Nadler et al., 2008), pursuit with a low gain could confound our results, hence it is important to examine the relationship between depth-sign selectivity and pursuit gain for the EO and MP conditions.

Pursuit velocity gain was quantified as the amplitude of the modulation in eye velocity at 0.5 Hz divided by the amplitude of the eye velocity required to maintain fixation accurately (Nadler et al., 2008), such that unity gain corresponds to perfect pursuit. As shown in Fig. 6, pursuit velocity gain was slightly less than one in both the EO and MP conditions. The average pursuit gain in the MP condition (0.967 ± 0.004 SE) was just slightly larger than the average pursuit gain in the EO condition (0.956 ± 0.004 SE), and this difference was significant ($p < 0.01$, paired t-test). Importantly, however, there is no significant correlation between the magnitude of DSDI and pursuit gain for either the EO or MP conditions (Spearman rank correlation, $p > 0.4$ for both conditions). Indeed many neurons are found to have strong depth-sign selectivity when pursuit gain is very close to unity (Fig. 6). Thus, there is no evidence that modest departures from perfect pursuit gain affected our findings in the EO condition, as also established previously for the MP condition (Nadler et al., 2008).

DISCUSSION

We recently demonstrated that neurons in area MT combine visual image motion with extra-retinal signals to code depth sign from motion parallax (Nadler et al., 2008), but the nature of the extra-retinal signals was not clear. In that study (and the Motion Parallax condition used here), monkeys made a compensatory eye movement during whole-body translation to maintain gaze on a world-fixed target, such that extra-retinal signals could originate from either head or eye movements. By employing additional experimental conditions that dissociate head and eye movements, we show that depth-sign selectivity is driven by signals related to eye movements, not head movements. Depth-sign selectivity in the Eye Only condition was nearly identical to that in the Motion Parallax condition, whereas no depth-sign selectivity was generally seen in the Head Only condition. Thus, smooth eye movement commands appear to provide the critical extra-retinal input for computing depth sign from motion parallax in MT, at least under the conditions of our experiment. This finding was predicted by human psychophysical studies (Naji and Freeman, 2004; Nawrot, 2003a, b; Nawrot and Joyce, 2006; Nawrot et al., 2004), which showed that head translations are not required to achieve unambiguous perception of depth from motion parallax. Although it remains to be demonstrated that area MT is causally involved in perception of depth based on motion parallax, these findings support the idea that MT provides robust sensory signals for this behavior.

Role of Vestibular Signals

The lack of depth-sign selectivity in the Head Only condition suggests that otolithic vestibular signals related to head translation were not used to disambiguate visual motion signals in area

MT. Were the vestibular stimuli used in our standard Motion Parallax and Head Only conditions (0.5Hz, peak velocity = 5.9 cm/s; peak acceleration = 14.1 cm/s^2) simply too weak to elicit effects in MT? On the one hand, stimuli with our frequency content (0.5 Hz) generally elicit very little translational vestibulo-ocular reflex (TVOR) in monkeys (Angelaki, 1998; Telford et al., 1997). Rather, these studies show that TVOR gain becomes substantially higher at frequencies above 1 Hz. Thus, it is possible that vestibular signals might contribute to depth-sign coding if the TVOR were engaged, as discussed further below. On the other hand, it is clear that our stimuli elicit robust vestibular signals that can reliably drive behavior. For example, our 0.5 Hz stimuli are well above the threshold linear accelerations needed to detect inertial motion (Benson et al., 1986), and our 1 Hz stimuli are even stronger. Moreover, preliminary results from our laboratory (using an identical motion platform) reveal that human subjects can readily discriminate between two opposite directions of whole-body translation (e.g, left vs right) in darkness at peak accelerations substantially lower than those used in this study (MacNeilage and Angelaki, unpublished). In addition, we previously reported that macaques can discriminate heading around straight forward with thresholds ranging from 1 −4° (Gu et al., 2007). This corresponds to discriminating an interaural (left vs. right) component of translation having a peak acceleration of 1.8−7.0 cm/s^2, well below the 14 cm/s^2 peak acceleration used here. Thus, there are reliable vestibular signals in the brain that could have interacted with visual motion signals in area MT during our experiment, but these signals do not appear to be used by MT to compute depth sign.

Nawrot and colleagues (Nawrot, 2003b; Nawrot and Joyce, 2006) have reported unambiguous perception of depth sign during stimulus conditions that are analogous to our Head Only condition. In this case, however, the depth percept has the opposite polarity. In other words, the same retinal image motion that generated a far percept under conditions analogous to our Motion Parallax condition instead produced a near percept, despite the fact that subjects were not required to make any eye movement. The explanation for this result likely lies in the interaction between the TVOR and pursuit eye movements. If a robust TVOR were generated in the Head Only condition, then a smooth eye movement command signal would have to be generated (in the opposite direction) to suppress the TVOR and maintain gaze on the head-fixed target. This covert smooth eye movement signal is thought to produce the reversed depth-sign tuning observed by Nawrot et al. (Nawrot, 2003b; Nawrot and Joyce, 2006), whose head movement stimuli were strong enough to produce a robust TVOR in humans. Indeed, Nawrot (2003a) has shown that, as viewing distance varies, the magnitude of perceived depth from motion parallax changes in parallel with the gain of pursuit.

By this logic, we might expect MT neurons to exhibit reversed depth-sign selectivity in our Head Only condition when the translational motion stimulus is strong enough to elicit a robust TVOR. If this were the case for the HO condition at higher temporal frequencies (Fig. 4C), we would expect neurons with positive DSDIs in the MP condition to have negative DSDIs in the HO condition, and neurons with negative DSDIs in the MP condition to have positive DSDIs in the HO condition. We failed to observe this effect at 1Hz (Fig. 4C), and we were unable to perform this test at higher frequencies. Even at 1Hz, previous studies in monkeys have shown that the TVOR is still several-fold undercompensatory (Angelaki, 1998; Telford et al., 1997), and this may explain the lack of reversed depth-sign selectivity. Attempts to perform this experiment at higher frequencies were unsuccessful for a few reasons. To get a reasonably robust TVOR, frequencies above 2Hz are required (Angelaki, 1998; Telford et al., 1997), but the gain of pursuit falls off in this range (Rashbass, 1961; Robinson, 1965). As a consequence, the ability to suppress eye movements of vestibular origin is relatively poor for head movements that exceed 1–2 Hz (Barnes et al., 1978; Lau et al., 1978). As a result of these factors, we found that fixation stability was not good above 1Hz in the Head Only condition, and this compromises the depth-sign ambiguity of our visual stimuli. Pursuit gain in the Motion Parallax condition was also low above 1 Hz. In addition, recording stability becomes poor for

frequencies above 1 Hz, thus making single-unit isolation difficult. For these reasons, we have been unable to conduct a decisive neurophysiology experiment analogous to the depth-sign reversal experiment performed with humans (Nawrot, 2003b; Nawrot and Joyce, 2006).

We therefore cannot exclude the possibility that vestibular signals might contribute to computations of depth sign from motion parallax in MT at high frequencies where a robust TVOR is elicited. Under the conditions of our experiment, however, it is clear that robust vestibular signals are generated, yet they do not contribute to depth-sign selectivity in MT. As predicted by the findings of Nawrot et al. (Nawrot, 2003b; Nawrot and Joyce, 2006), our results suggest strongly that it is the smooth eye movement—whether generated volitionally or reflexively to suppress the TVOR—that is responsible for generating depth from motion parallax. When vestibular signals contribute to this process, they may do so only indirectly via altering the requirement for smooth eye movements. This notion is consistent with a preliminary report that MT neurons do not respond to translational and rotational movements in darkness, whereas MSTd neurons show robust spatial tuning for the same stimuli (Chowdhury et al., 2008).

Strength and Origin of Eye Movement Signals in MT

Lesions of area MT can disrupt both the initiation and maintenance of smooth pursuit eye movements (Dursteler and Wurtz, 1988; Dursteler et al., 1987). However, previous studies suggest that MT neurons mainly carry visual, not extra-retinal, signals related to pursuit. For example, briefly extinguishing a pursuit target or stabilizing the retinal image during maintained pursuit largely eliminates the responses of MT neurons (Newsome et al., 1988). Moreover, robust responses in MT during pursuit maintenance were found mainly among neurons with receptive fields that overlap the fovea, such that retinal slip of the pursuit target drives responses (Komatsu and Wurtz, 1988; Newsome et al., 1988).

In this light, a surprising aspect of our results is that smooth eye movements strongly modulate the responses of MT neurons to moving visual stimuli. These interactions are strongly dependent on the phase of visual motion relative to eye movement, and result in depth-sign selectivity for motion parallax. As reported previously (Nadler et al., 2008), the modulation of visual responses by smooth eye movement is often large compared to the effect of eye movements on MT responses when there is no visual stimulus within the receptive field. Moreover, both the strength and the polarity of interactions between smooth eye movements and visually-driven responses are complex and appear to be nonlinear (see Supplement to Nadler et al., 2008). This nonlinear, phase-dependent interaction may explain why we see strong effects of smooth eye movements on MT responses, whereas extra-retinal signals related to pursuit were not clearly apparent in previous studies (Newsome et al., 1988).

What then is the origin of the smooth eye movement signals that we observe in MT? Proprioceptive signals from extra-ocular muscles (Wang et al., 2007) are an unlikely source because they could not explain the unambiguous, but reversed, percepts of depth seen by Nawrot (2003b) when human subjects are tested under conditions similar to our HO condition (in which the eyes do not move). It seems more likely that area MT receives an efference copy of smooth eye movement command signals. Extra-retinal signals related to smooth pursuit have been seen in a wide number of brain areas, including the brainstem (Krauzlis et al., 2000; Lisberger et al., 1994; Zhang et al., 1995), cerebellum (Lisberger and Fuchs, 1978) and cerebral cortex (Bremmer et al., 1997; Heinen, 1995; Sakata et al., 1983). However, it remains unclear exactly how these pursuit signals arrive in area MT. Even for neighboring area MSTd, where strong extra-retinal signals related to pursuit have been long apparent, the pathways that provide pursuit inputs to MSTd are unclear (Ono and Mustari, 2006). Understanding the source and exact nature of the extra-retinal eye movement inputs to area MT will thus require further studies.

In closing, our findings show that a smooth eye movement signal is both necessary and sufficient to produce selectivity for depth sign from motion parallax in area MT. This result is fully consistent with the predictions of human psychophysical studies (Naji and Freeman, 2004; Nawrot, 2003a, b; Nawrot and Joyce, 2006; Nawrot et al., 2004). Although head movements often generate retinal image sequences containing motion parallax, the visual system appears to rely on the accompanying smooth eye movement signals to compute depth.

EXPERIMENTAL PROCEDURES

Subjects and Preparation

Two alert, trained male monkeys (*Macaca mulatta*) were used for neurophysiological experiments. Surgeries were performed under general anesthesia to implant a head holding device and a pair of scleral coils for measuring eye movements. The head holding device consisted of a large Delrin ring attached to the skull with T-bolts and acrylic cement. This ring was locked into place during recording experiments and provided good recording stability for up to 2 hours even in the presence of full body motion. An eye coil was implanted under the conjunctiva in each eye, and terminated in a plug that was embedded in acrylic within the head-restraint ring. Animals were allowed to fully recover from surgery before electrophysiological recordings in MT were begun. All animal procedures have been approved by the Animal Studies Committee at Washington University School of Medicine and by the University Committee on Animal Resources at the University of Rochester. Additional details regarding preparation of animal subjects can be found elsewhere (Gu et al., 2006).

`A bilateral recording grid was attached to the skull using acrylic. Prior to recordings, small burr holes were drilled through the skull at known stereotaxic locations aligned with the grid. At the beginning of each recording session, a hydraulic microdrive was mounted above the recording grid, and a tungsten microelectrode (FHC) was inserted into the brain via a transdural guide tube. Area MT was localized based on stereotaxic coordinates, structural MRI images, grey-white matter transitions during electrode penetrations, and characteristic response properties (Albright, 1984; DeAngelis and Newsome, 1999; DeAngelis and Uka, 2003; Maunsell and Van Essen, 1983). Extracellular recordings were made from isolated single neurons. Raw signals were amplified and band-pass filtered between 400Hz and 5kHz using an 8-pole filter. Spikes were isolated with a dual voltage-time window discriminator and spike times were recorded with 1ms resolution.

Apparatus

During training and recording sessions, monkeys were head restrained and seated in front of a 60×60cm tangent screen at a viewing distance of 32cm. The screen subtended roughly 90° ×90° of visual angle, and images were refreshed at 60Hz. The stereoscopic projector (Christie Digital Mirage 2000) and display screen were mounted to a six degree-of-freedom motion platform (MOOG 6DOF2000E) that allowed us to translate the animal along any direction within the frontoparallel plane (no fore-aft movements were used in this study). Platform movements and visual stimuli were controlled by computer at 60Hz, and the measured transfer function of the system (verified by accelerometer measurements) allowed us to accurately record the platform's position at all times (Gu et al., 2007; Gu et al., 2006). Using the transfer function to predict the upcoming movement of the platform, we were able to adjust a variable delay between movement command signals and video signals in order to eliminate delays between platform motion and visual image motion. A room-mounted laser was used to ensure that platform motion was synchronized with the video display (to within ~1ms). A more detailed description of the motion platform and projection system has been published previously (Gu et al., 2006).

Visual Stimuli

The visual stimuli used in these experiments have been previously described in detail (Nadler et al., 2008). For each isolated neuron, we first obtained quantitative measurements of direction tuning, speed tuning, and receptive field location. Direction tuning was assessed by presenting random-dot patterns (within a circular aperture) that moved at a constant speed in one of eight directions separated by 45 degrees. A direction tuning curve was obtained by collecting three to five measurements of response for each of the eight directions. To measure speed tuning, each neuron was tested with random-dot patterns that had speeds of 0, 0.5, 1, 2, 4, 8, 16 and 32 degrees/s and that moved in the neuron's preferred direction. The preferred speed was obtained from the peak of a fitted Gamma function (Nover et al., 2005). The receptive field was mapped quantitatively by presenting small circular patches of moving dots at sixteen locations (on a 4×4 grid) that covered the area of the receptive field as estimated by hand mapping. A two-dimensional Gaussian was fit to this quantitative map to determine the center location and size of the receptive field. Results of these preliminary tests were used to specify the location, size, and direction axis of the random-dot stimulus used in our main experiment.

For the main experimental conditions, the OpenGL graphics library was used to render a fixation target and a random-dot surface in a world-fixed virtual environment. The random-dot surface was a portion of a cylinder (analogous to the geometric horopter) that was oriented perpendicular to the axis of translation (see Nadler et al., 2008 for details). After the monkey achieved fixation for 500 ms, the random dot stimulus was presented in the neuron's receptive field. In each trial, one of nine simulated depths was chosen pseudo-randomly from the following set of equivalent disparities: $0.0°$, $±0.5°$, $±1.0°$, $±1.5°$, and $±2.0°$. 'Null' trials, in which no stimulus was presented over the receptive field, were also interleaved. The random-dot patch was positioned and scaled, using a ray tracing procedure, so as to eliminate all cues to depth other than motion parallax. The stimulus was also transparent such that no occlusion cues were present when dots overlapped the fixation target. Stimuli were viewed monocularly through ferro-electric liquid crystal shutters that were synchronized to the monitor refresh. For almost all cells, visual stimuli were presented to the eye contralateral to the recording hemisphere. This was achieved by keeping the shutter for one eye closed while the other remained open.

Stimulus Conditions

For this study, we have collected data under four stimulus conditions: the Motion Parallax and Retinal Motion conditions (as described in detail previously, Nadler et al., 2008), and two new conditions described below (Fig. 1). The monkey performed 6–10 repetitions of each stimulus condition at each simulated depth, randomly interleaved.

In the Motion Parallax (MP) condition (Fig. 1b) the animal was translated through one cycle of a 0.5 Hz sinusoid having a total displacement of 4 cm, which is slightly more than one interocular separation. The 2 s trajectory was windowed with a high powered Gaussian to smooth out the beginning and end of the movement. All movements were generated along the axis in the fronto-parallel plane that corresponded to the visual preferred-null axis of the recorded neuron, as determined by a preliminary measurement of direction tuning. As a result, random-dot motion oscillated back and forth along the neuron's preferred-null axis during translation of the animal. For example, if the neuron preferred horizontal visual motion, the platform moved 2 cm to the right, 4 cm to the left (passing through the starting point), and then 2 cm to the right, over the period of 2 s. The opposite starting direction was used on the other half of trials. Thus, platform movement produced visual image motion that began in the preferred direction for half of the trials, and began in the null direction for the other half of the trials. Throughout the movement, the monkey's only task was to fixate a small world-centered cross, and successful completion of the trial required that the monkey's gaze remained within

an electronic fixation window. To keep gaze on the world-centered target, the monkey was required to make a compensatory smooth pursuit eye movement (e.g., platform motion to the right required a leftward eye movement) To allow the monkey an opportunity to make an initial catch-up saccade (if necessary) at the onset of pursuit, the fixation window was initially 3.0–4.0 degrees square, and shrunk to 1.5–2.0 degrees after 250 ms. Horizontal and vertical eye position was monitored with a scleral search coil and captured at a sampling rate of 250 Hz.

In the Retinal Motion (RM) condition (Fig. 1a), the monkey remained stationary and we replicated the visual image seen during the MP condition by translating the OpenGL camera along the same trajectory that the monkey followed in the MP condition. As the OpenGL camera translated, it also rotated to maintain simulated gaze at the fixation target. This simulates the smooth tracking eye movements of the animal, and generates visual stimuli that match those in the MP condition. Thus, this condition serves as a control for retinal image motion.

In the MP condition, both head movements and eye movements occur, such that either vestibular signals or smooth eye movement command signals could provide extra-retinal inputs to area MT (Fig. 1b). To distinguish between head and eye movements as the origin of extra-retinal signals, we designed two additional stimulus conditions. In the *Head Only (HO)* condition (Fig. 1c), the animal is translated by the motion platform as in the MP condition, but the fixation point now moves with the animal (head-fixed) such that the animal is not required to execute any eye movement to maintain gaze on the target. Retinal image motion is generated as in the RM condition. Thus, the HO condition isolates vestibular and proprioceptive inputs related to movement of the head/body. In the *Eye Only (EO)* condition (Fig. 1d), the motion platform remains stationary and the entire virtual environment, including the fixation target, is translated back and forth in front of the animal. As a result, the animal is required to make the same smooth eye movement as in the MP condition but there are no vestibular/proprioceptive signals related to head or body movement. Thus the EO condition isolates the smooth eye movement signals that occur in the MP condition. The specific movement parameters for these additional conditions are the same as for the Motion Parallax condition. In the MP and HO conditions (in which the platform moves), this results in peak head movement speeds of 5.9 cm/s at the midpoint of the cycle (t = 1.0 s), peak accelerations of 14.1 cm/s^2 (0.014 G) at the start and stop of the cycle (t = 0.0, 2.0 s), and peak accelerations of 9.7 cm/s^2 (0.010 G) when the platform changes direction (t = 0.5, 1.5 s). Importantly, the visual image motion in the receptive field is the same under all four conditions assuming that the monkey accurately tracks the fixation target with his eyes. As shown previously, small deviations in the gain of pursuit do occur, but these cannot account for the depth-sign selectivity seen in the MP condition (Nadler et al., 2008).

Data Analysis

Data analysis and significance testing has been described in detail previously (Nadler et al., 2008). Single-unit data were analyzed using custom software written in Matlab (Mathworks Inc.). For generation of peri-stimulus time histograms (PSTHs, Fig. 2a–d), firing rate was computed in 50 ms bins. To quantify selectivity for depth sign, we combined data across the two phases of platform motion and computed a mean firing rate across the total duration of each trial. Spikes were counted within a temporal window that began 80ms after stimulus onset and ended 80ms after stimulus offset (to compensate for response latency). For each neuron and each of the four stimulus conditions, we computed a Depth-Sign Discrimination Index (DSDI):

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

For each pair of depths symmetric around zero (e.g. +/−1 degree), we calculated the difference in response between far (R_{far}) and near (R_{near}), relative to response variability (σ_{avg}, the average standard deviation of the two responses). We then averaged across the four matched pairs of depths to obtain the DSDI which ranges from −1 to 1. Neurons that respond more strongly to near stimuli will have negative values, while neurons that prefer far stimuli will have positive values. DSDIs were classified as significantly different from zero by permutation test (1000 permutations, $p < 0.05$). The DSDI metric has the advantage of taking into account trial-to-trial variations in response while quantifying the magnitude of response differences between near and far depths.

Acknowledgements

REFERENCES

Albright TD. Direction and orientation selectivity of neurons in visual area MT of the macaque. J Neurophysiol 1984;52:1106–1130. [PubMed: 6520628]

Angelaki DE. Three-dimensional organization of otolith-ocular reflexes in rhesus monkeys. III. Responses To translation. J Neurophysiol 1998;80:680–695. [PubMed: 9705461]

Barnes GR, Benson AJ, Prior AR. Visual-vestibular interaction in the control of eye movement. Aviat Space Environ Med 1978;49:557–564. [PubMed: 305777]

Benson AJ, Spencer MB, Slott JR. Thresholds for the detection of the direction of whole-body, linear movement in the horizontal plane. Aviat Space Environ Med 1986;57:1088–1096. [PubMed: 3790028]

Bremmer F, Distler C, Hoffmann KP. Eye position effects in monkey cortex. II. Pursuit- and fixation-related activity in posterior parietal areas LIP and 7A. J Neurophysiol 1997;77:962–977. [PubMed: 9065861]

Chowdhury SA, Takahashi K, DeAngelis GC, Angelaki DE. Does area MT carry vestibular signals related to self-motion? Soc Neurosci Abstr. 2008

DeAngelis GC, Newsome WT. Organization of disparity-selective neurons in macaque area MT. J Neurosci 1999;19:1398–1415. [PubMed: 9952417]

DeAngelis GC, Uka T. Coding of horizontal disparity and velocity by MT neurons in the alert macaque. J Neurophysiol 2003;89:1094–1111. [PubMed: 12574483]

Dursteler MR, Wurtz RH. Pursuit and optokinetic deficits following chemical lesions of cortical areas MT and MST. J Neurophysiol 1988;60:940–965. [PubMed: 3171667]

Dursteler MR, Wurtz RH, Newsome WT. Directional pursuit deficits following lesions of the foveal representation within the superior temporal sulcus of the macaque monkey. J Neurophysiol 1987;57:1262–1287. [PubMed: 3585468]

Farber JM, McConkie AB. Optical motions as information for unsigned depth. J Exp Psychol Hum Percept Perform 1979;5:494–500. [PubMed: 528954]

Gu Y, DeAngelis GC, Angelaki DE. A functional link between area MSTd and heading perception based on vestibular signals. Nat Neurosci 2007;10:1038–1047. [PubMed: 17618278]

Gu Y, Watkins PV, Angelaki DE, DeAngelis GC. Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. J Neurosci 2006;26:73–85. [PubMed: 16399674]

Hayashibe K. Reversals of visual depth caused by motion parallax. Perception 1991;20:17–28. [PubMed: 1945729]

Heinen SJ. Single neuron activity in the dorsomedial frontal cortex during smooth pursuit eye movements. Exp Brain Res 1995;104:357–361. [PubMed: 7672029]

Howard, IP.; Rogers, BJ. Binocular vision and stereopsis. New York: Oxford University Press; 1995.

Komatsu H, Wurtz RH. Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. J Neurophysiol 1988;60:580–603. [PubMed: 3171643]

Krauzlis RJ, Basso MA, Wurtz RH. Discharge properties of neurons in the rostral superior colliculus of the monkey during smooth-pursuit eye movements. J Neurophysiol 2000;84:876–891. [PubMed: 10938314]

Lau CG, Honrubia V, Jenkins HA, Baloh RW, Yee RD. Linear model for visual-vestibular interaction. Aviat Space Environ Med 1978;49:880–885. [PubMed: 307382]

Lisberger SG, Fuchs AF. Role of primate flocculus during rapid behavioral modification of vestibuloocular reflex. I. Purkinje cell activity during visually guided horizontal smooth-pursuit eye movements and passive head rotation. J Neurophysiol 1978;41:733–763. [PubMed: 96225]

Lisberger SG, Pavelko TA, Broussard DM. Responses during eye movements of brain stem neurons that receive monosynaptic inhibition from the flocculus and ventral paraflocculus in monkeys. J Neurophysiol 1994;72:909–927. [PubMed: 7983546]

Maunsell JH, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. J Neurophysiol 1983;49:1127–1147. [PubMed: 6864242]

Nadler JW, Angelaki DE, DeAngelis GC. A neural representation of depth from motion parallax in macaque visual cortex. Nature 2008;452:642–645. [PubMed: 18344979]

Naji JJ, Freeman TC. Perceiving depth order during pursuit eye movement. Vision Res 2004;44:3025–3034. [PubMed: 15474575]

Nakayama K, Loomis JM. Optical velocity patterns, velocity-sensitive neurons, and space perception: a hypothesis. Perception 1974;3:63–80. [PubMed: 4444922]

Nawrot M. Depth from motion parallax scales with eye movement gain. J Vis 2003a;3:841–851. [PubMed: 14765966]

Nawrot M. Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. Vision Res 2003b;43:1553–1562. [PubMed: 12782069]

Nawrot M, Joyce L. The pursuit theory of motion parallax. Vision Res 2006;46:4709–4725. [PubMed: 17083957]

Nawrot M, Nordenstrom B, Olson A. Disruption of eye movements by ethanol intoxication affects perception of depth from motion parallax. Psychol Sci 2004;15:858–865. [PubMed: 15563332]

Newsome WT, Wurtz RH, Komatsu H. Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. J Neurophysiol 1988;60:604–620. [PubMed: 3171644]

Nover H, Anderson CH, DeAngelis GC. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. J Neurosci 2005;25:10049–10060. [PubMed: 16251454]

Ono S, Mustari MJ. Extraretinal Signals in MSTd Neurons Related to Volitional Smooth Pursuit. J Neurophysiol 2006;96:2819–2825. [PubMed: 16790593]

Rashbass C. The relationship between saccadic and smooth tracking eye movements. J Physiol 1961;159:326–338. [PubMed: 14490422]

Richards W. Structure from stereo and motion. J Opt Soc Am A 1985;2:343–349. [PubMed: 3973765]

Robinson DA. The mechanics of human smooth pursuit eye movement. J Physiol 1965;180:569–591. [PubMed: 5846794]

Rogers B, Graham M. Motion parallax as an independent cue for depth perception. Perception 1979;8:125–134. [PubMed: 471676]

Rogers S, Rogers BJ. Visual and nonvisual information disambiguate surfaces specified by motion parallax. Percept Psychophys 1992;52:446–452. [PubMed: 1437477]

Sakata H, Shibutani H, Kawano K. Functional properties of visual tracking neurons in posterior parietal association cortex of the monkey. J Neurophysiol 1983;49:1364–1380. [PubMed: 6875628]

Telford L, Seidman SH, Paige GD. Dynamics of squirrel monkey linear vestibuloocular reflex and interactions with fixation distance. J Neurophysiol 1997;78:1775–1790. [PubMed: 9325347]

Wang X, Zhang M, Cohen IS, Goldberg ME. The proprioceptive representation of eye position in monkey primary somatosensory cortex. Nat Neurosci 2007;10:640–646. [PubMed: 17396123]

Zhang Y, Partsalis AM, Highstein SM. Properties of superior vestibular nucleus flocculus target neurons in the squirrel monkey. I. General properties in comparison with flocculus projecting neurons. J Neurophysiol 1995;73:2261–2278. [PubMed: 7666137]

A
Retinal Motion (RM)

B
Motion Parallax (MP)

C
Head Only (HO)

D
Eye Only (EO)
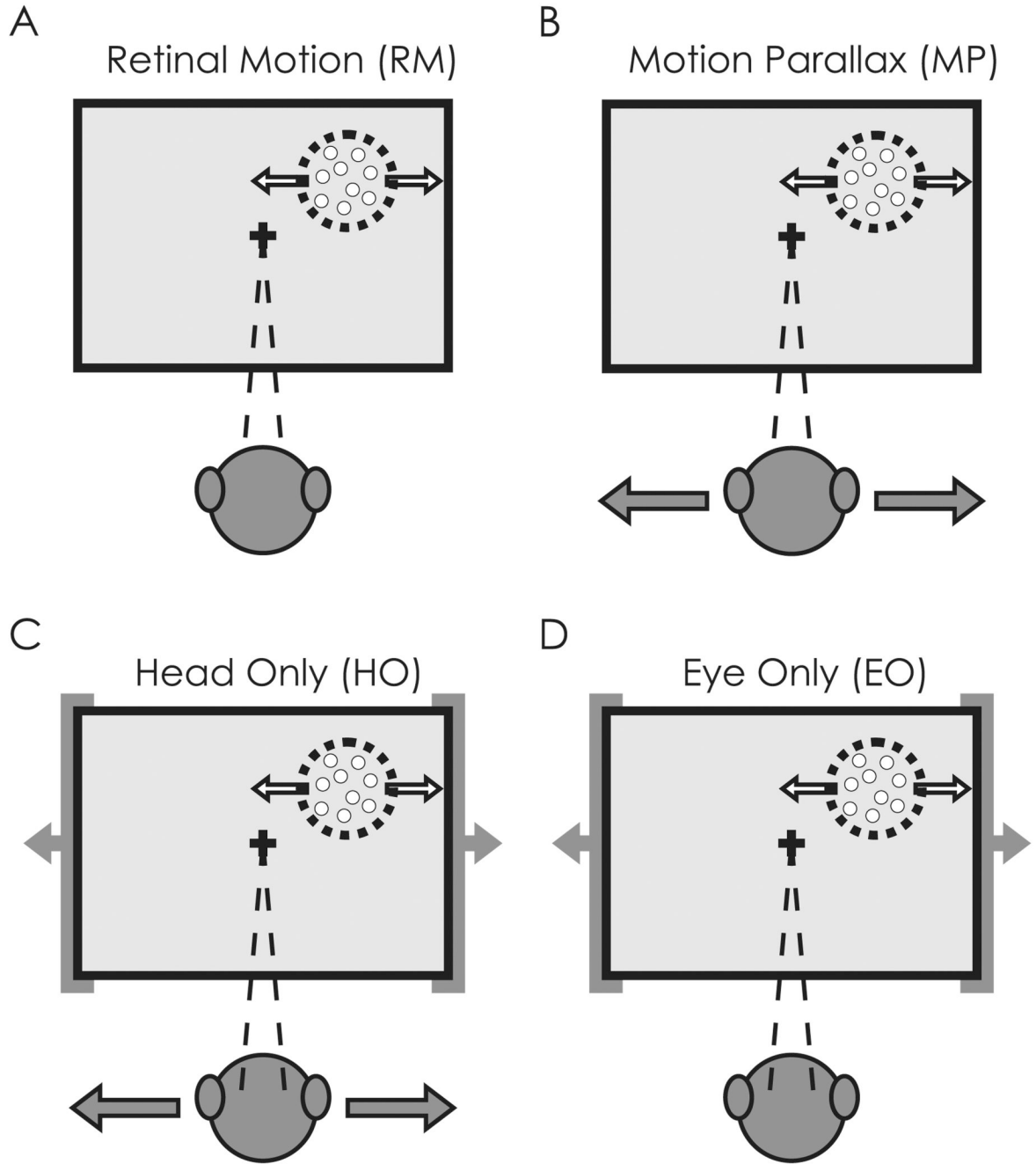
Figure 1.
Stimulus conditions used to test for depth-sign selectivity in area MT. Retinal image motion within the receptive field of the MT neuron (dashed circle) is the same in all conditions. A) In the Retinal Motion (RM) condition, the animal's body and eyes remain stationary while motion of a random-dot stimulus (white arrows) simulates a range of depths by motion parallax. B) In the Motion Parallax (MP) condition, the animal is translated (gray arrows) while maintaining fixation on a world-fixed target (cross). This condition involves both head movements and compensatory eye movements. C) In the Head Only (HO) condition, the animal's body and the virtual world are moved in phase (gray arrows) such that the same head movement occurs as in the MP condition, but the animal does not need to make any eye movements. D) In the

Eye Only (EO) condition, the animal remains stationary, but the entire virtual world (including the fixation target) is translated in front of the animal. Thus, the animal makes the same eye movements as in the MP condition but there is no head movement.
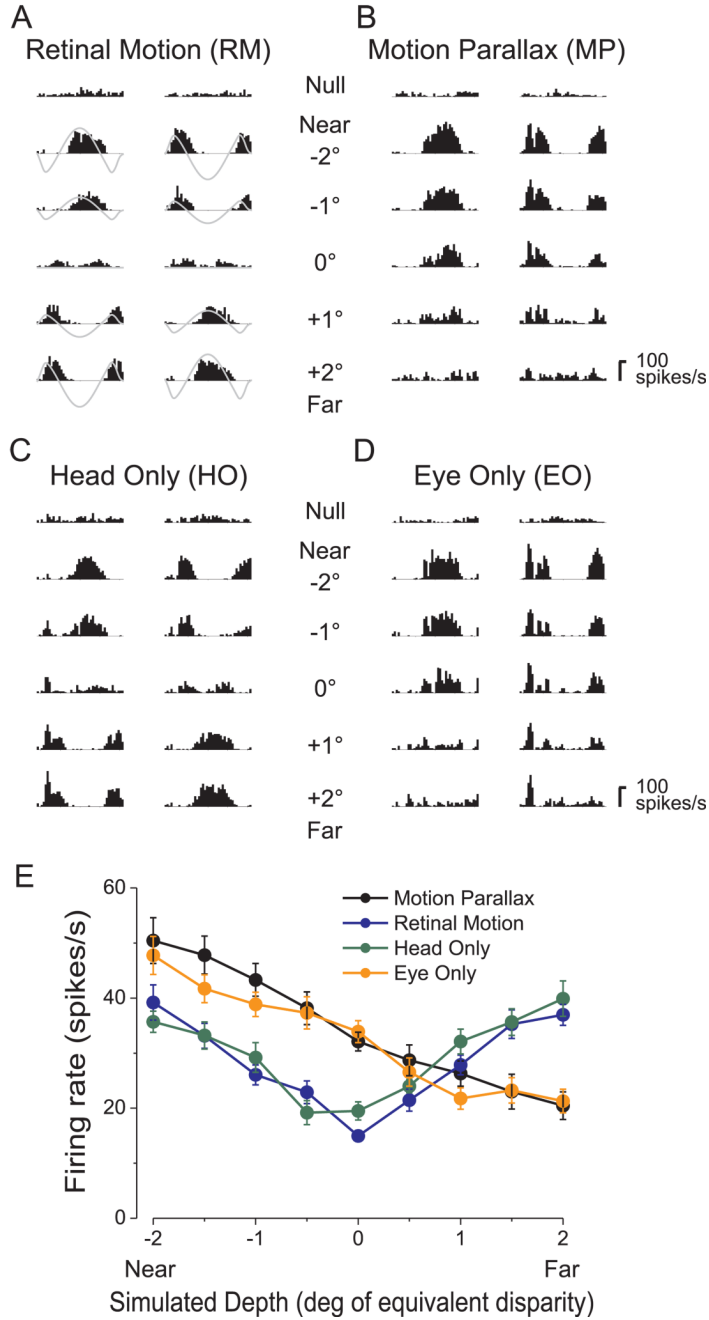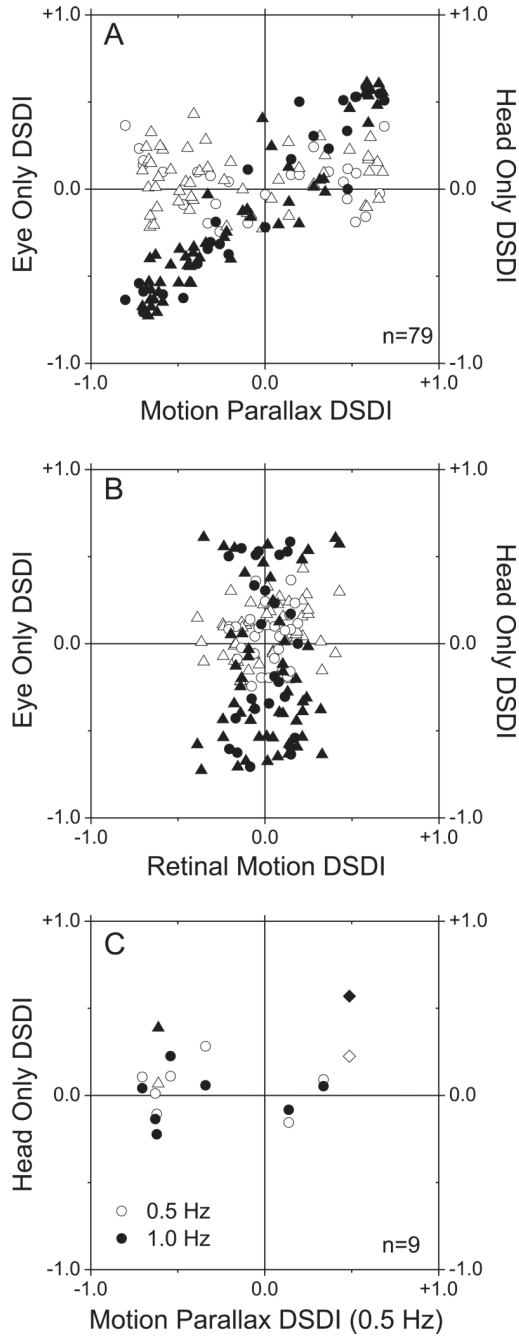
Figure 2.
Dissociation of extra-retinal inputs to an MT neuron. A) PSTHs in response to a 2 s stimulus presentation for five of nine simulated depths in the RM condition (for clarity, data for ±0.5° and ±1.5° are not shown). One column of PSTHs is shown for each starting phase of simulated observer motion. Gray traces represent retinal image velocity, with peaks and troughs representing motion in the preferred and anti-preferred directions, respectively. B) Responses of the same neuron in the MP condition. Note that responses to near stimuli are accentuated, while responses to far stimuli are suppressed relative to the RM condition. C) Responses in the HO condition are very similar to those seen in the RM condition. D) Responses in the EO condition are very similar to those in the MP condition. E) Depth tuning curves for each of the

four stimulus conditions: MP (black), RM (blue), HO (green), and EO (orange). Average firing rate over the 2s stimulus epoch is plotted as a function of simulated depth. Error bars represent SEs.
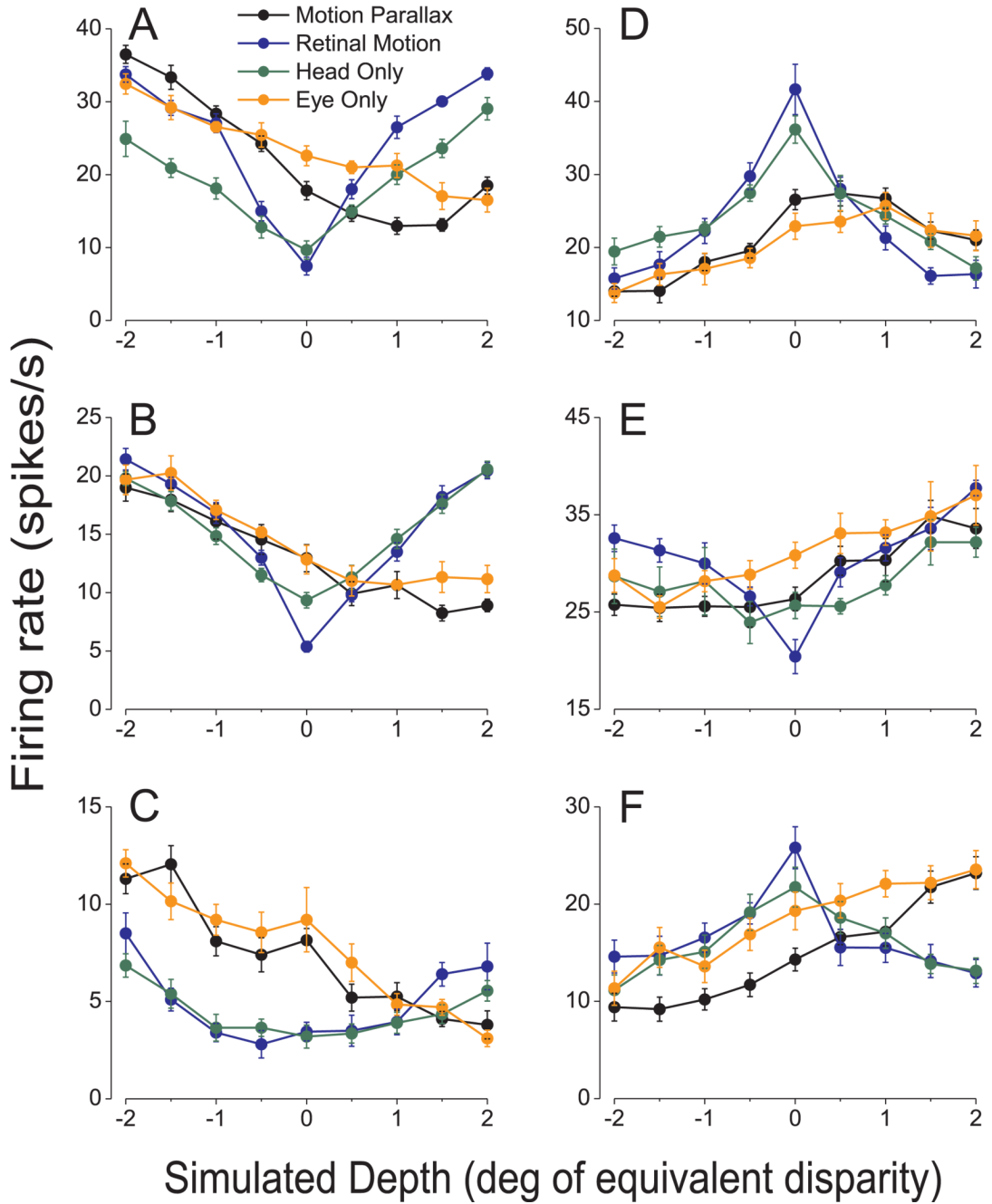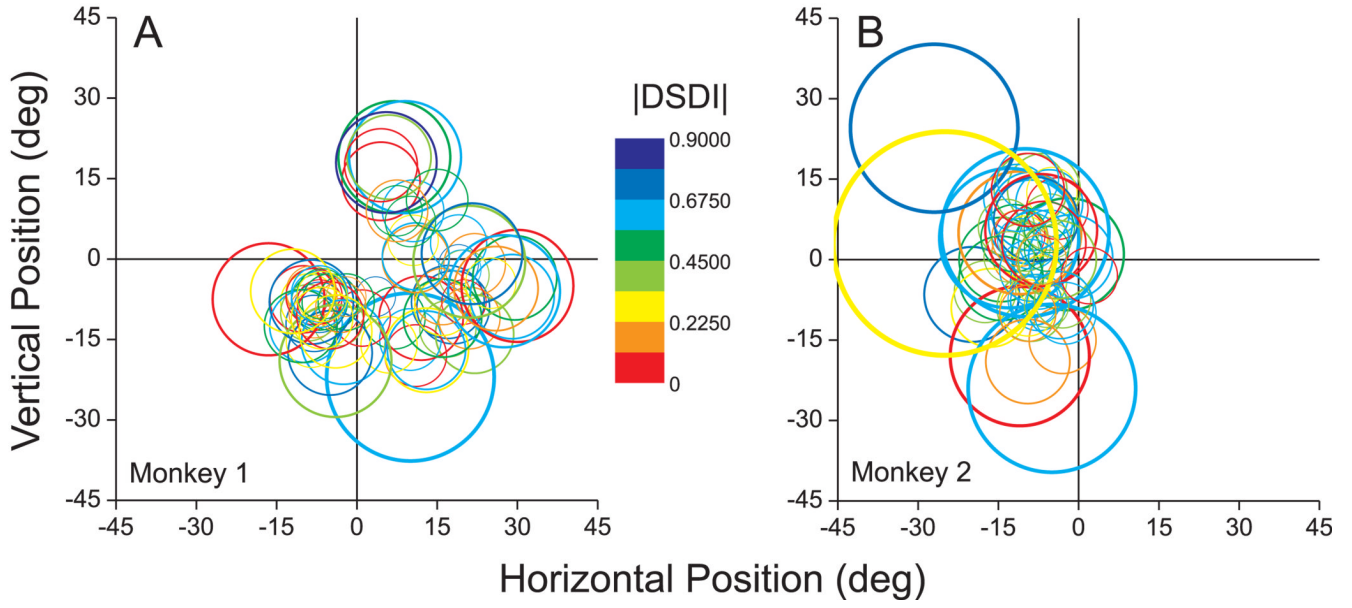
Figure 3.
Tuning curves from six additional MT cells that show selectivity for depth sign from motion parallax. Mean firing rates in the MP (black), RM (blue), HO (green), and EO (orange) conditions are shown for all nine simulated depths. Error bars are SEs. A, B, C) Neurons that prefer near stimuli in the MP condition. DSDI values for these near-preferring neurons in each stimulus condition were as follows, MP: −0.80, −0.67, −0.64; RM: 0.15, −0.35, 0.14; EO: −0.64, − 0.73, −0.63; HO: 0.36, 0.01, −0.20. D, E, F) Neurons that prefer far stimuli. DSDI values were as follows, MP: 0.66, 0.61, 0.68; RM: −0.13, 0.42, −0.23; EO: 0.55, 0.57, 0.56; HO: −0.02, 0.29, 0.09. Notice that two of these far cells (D, F) prefer slow speeds, and thus their RM tuning curves are symmetrically peaked instead of symmetrically U-shaped.
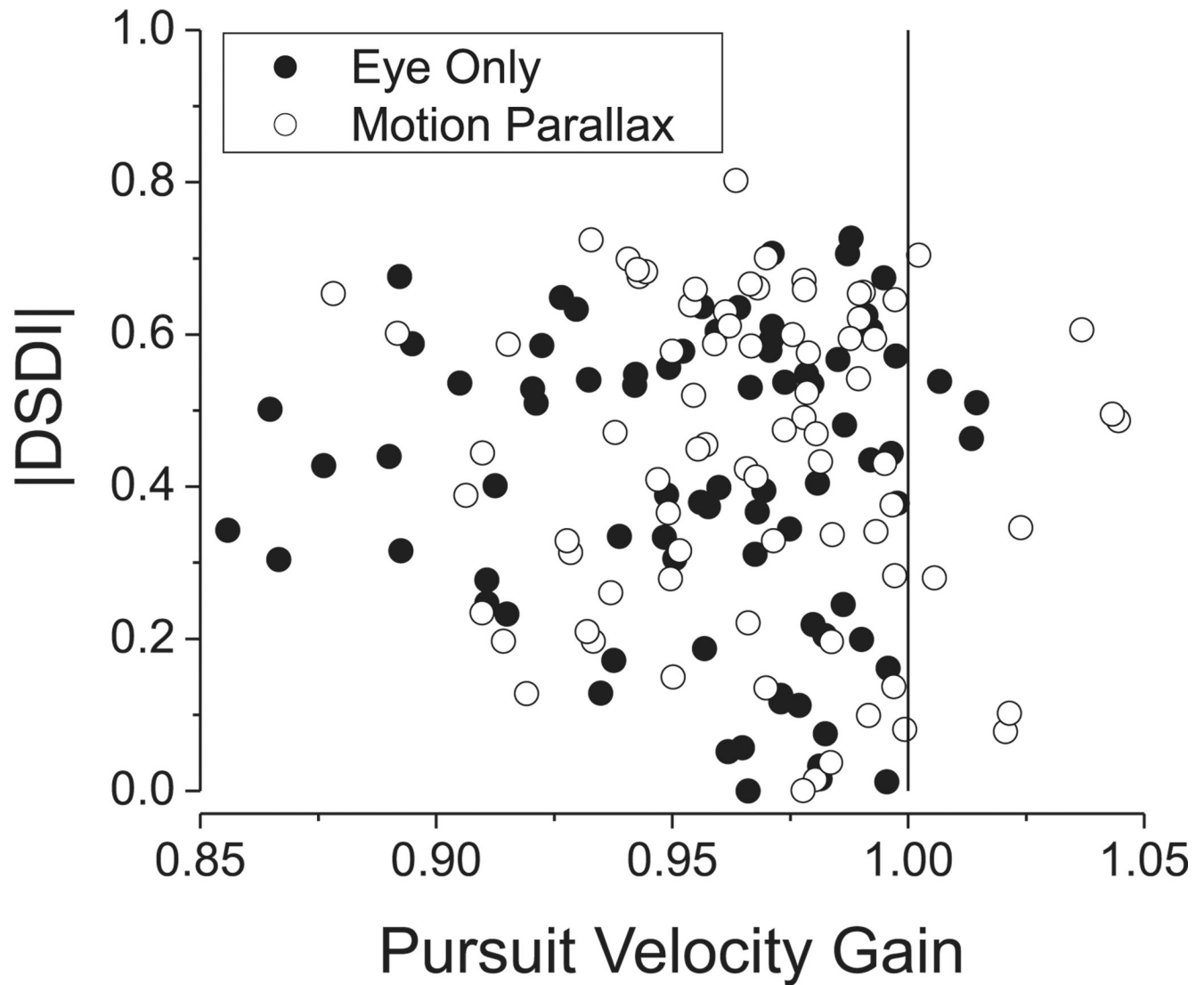
Figure 4.
Eye movements are necessary and sufficient to produce depth-sign tuning. A) This scatter plot compares DSDIs for the EO condition (filled symbols) and the HO condition (open symbols) to DSDIs obtained in the MP condition (data from 79 neurons from two monkeys; M1: circles, M2: triangles). Only DSDIs obtained in the EO condition are correlated with those obtained in the MP condition. B) DSDI values for the EO and HO conditions are plotted against those measured in the RM condition. C) Effect of temporal frequency in HO condition. This scatterplot compares DSDIs for the HO condition at 1.0Hz (filled circles) and 0.5Hz (open circles) with DSDIs obtained in the MP condition at 0.5Hz (data from 9 neurons from M2).

Only two neurons, denoted by triangles and diamonds, showed significant depth-sign tuning in the HO condition at 1.0 Hz.

Figure 5.
Graphical summary of the relationship between depth-sign selectivity and the receptive field location and size of MT neurons. Each circle represents the receptive field of a single MT neuron, as determined from quantitative receptive-field mapping. The color of each circle represents the absolute value of the DSDI in the MP condition, as indicated by the color scale. The origin (0, 0) in these graphs represents the location of the fixation target (i.e., the fovea). A) Data from monkey 1. B) Data from monkey 2.

Figure 6.
Depth-sign selectivity is not affected by small variations in pursuit gain. The absolute value of DSDI is plotted as a function of pursuit velocity gain for each of the 79 MT neurons tested. Filled and open symbols represent data collected in the EO and MP conditions, respectively.