The role of binocular disparity in the neural representation of multiple moving stimuli in the visual cortex

Anjani Sreeprada Chakrala*, Jianbo Xiao*, Xin Huang

Department of Neuroscience, School of Medicine and Public Health,
University of Wisconsin – Madison

* These authors contributed equally to this work.

Correspondence should be addressed to:

Xin Huang, Department of Neuroscience, University of Wisconsin, Madison, WI 53705, USA
Email: Xin.Huang@wisc.edu

Keywords: neural encoding, binocular disparity, transparent motion, divisive normalization, segmentation, figure-ground segregation, natural scenes, object-based attention, extrastriate cortex

Acknowledgment

We thank Dr. Greg DeAngelis for sharing their data set of the disparity tuning of a large number of MT neurons to be used in our model, Bryce Arseneau for animal training, and Drs. Jennifer Coonen and Kevin Brunner at the Wisconsin National Primate Research Center for excellent veterinary care and surgical assistance.

Grant support

This research was supported by National Eye Institute Grant R01 EY022443.

Conflict of Interest

None.

Abstract

Natural scenes often contain multiple objects and surfaces in 3-dimensional space. A fundamental process of vision is to segment visual scenes into distinct objects and surfaces. The stereoscopic depth and motion cues are particularly important for segmentation. However, how the primate visual system represents multiple moving stimuli located at different depths is poorly understood. Here we investigated how neurons in the middle temporal (MT) cortex represented two overlapping surfaces located at different horizontal disparities and moved simultaneously in different directions. We recorded the neuronal activities in MT of three male macaque monkeys while two of them performed a discrimination task to report the motion direction of an attended surface of two overlapping stimuli, and the third animal performed a behavioral task with the attention directed away from the receptive fields of MT neurons. We found that neuronal responses to overlapping surfaces showed a robust bias toward the horizontal disparity of one of the two surfaces. For all animals, the disparity bias in response to two surfaces was positively correlated with the disparity preference of the neurons to single surfaces. For two animals, neurons that preferred the near disparities of single surfaces (near neurons) showed a near bias to overlapping stimuli, and neurons that preferred the far disparities (far neurons) showed a far bias. For another animal, both near and far neurons showed a near bias to overlapping stimuli, although near neurons showed a stronger near bias. The disparity bias to overlapping stimuli was delayed relative to the response onset and was more delayed when the angular separation between two motion directions was smaller. Interestingly, for all three animals, both near and far neurons showed an initial near bias in comparison to the average of the responses to individual surfaces. We also found that the effect of attention directed to the disparity of one of two surfaces was object-based rather than feature-based. Although attention can modulate neuronal response to better represent the attended surface, the disparity bias cannot be explained by attention modulation. Our results can be explained by a unified model with a variable pooling size to weigh the response to individual stimulus components and divisive normalization. Our results revealed the encoding rule for multiple horizontal disparities and motion directions of overlapping stimuli. The disparity bias would allow subgroups of neurons to better represent different surfaces of multiple stimuli and therefore provide a population code that aids segmentation. The tendency for MT neurons to better represent the near-surface of overlapping stimuli in one animal and during the early response period in all three animals suggests that the neural representation of multiple stimuli at different depths may be beneficial to figure-ground segregation since figural objects are more likely to be in front of the ground in natural scenes.

Introduction

Natural scenes often contain multiple objects and surfaces in 3-dimensional visual space. Segmenting scenes into distinct objects and surfaces is fundamental to primate vision. Among other visual features, stereoscopic depth and motion velocity are important cues aiding segmentation. While depth (Wallace and Mamassian 2004) or motion cue alone (Wallace and Mamassian 2003) can lead to segmentation, when combined they further enhance segmentation (Snowden and Rossiter 1999; Goutcher 2016; Kohler et al. 2019; Nawrot and Blake 1989; Bradshaw and Rogers 1996; Ban et al. 2012). For example, psychophysical studies have shown that it becomes easier to segment overlapping surfaces moving in different directions when they are located at different depths than at the same depth (Hibbard and Bradshaw 1999; Greenwood and Edwards 2006b).

Although the neural representation of the horizontal disparity of a single stimulus in the visual cortex has been well characterized (Cumming et al. 1999; Prince et al. 2002; DeAngelis and Uka 2003), how the visual system represents multiple horizontal disparities of visual stimuli located at different depths is poorly understood. Previous neurophysiology studies have characterized how neurons in V1 and the middle-temporal (MT) cortex represent multiple motion directions of overlapping stimuli located at the same depth (Snowden et al. 1991; Qian and Andersen, 1994; Treue et al. 2000; Krekelberg and van Wezel 2013; Xiao et al. 2014; McDonald et al. 2014; Xiao and Huang 2015; Wiesner et al 2020). Relatively little is known about the neural representation of multiple stimuli moving at different depths. In one study, Bradley and colleagues (1995) showed that the responses of MT neurons became stronger as the difference between the horizontal disparities of two overlapping stimuli moving in opposite directions increased. This suggests that increasing the disparity difference between two surfaces reduces inhibition between motion signals. However, this study investigated the interaction between motion and disparity cues only for one set of motion directions, i.e., the preferred and null directions of each neuron. To fully understand the neural representation of multiple moving stimuli and the impact of stereoscopic depth cues, it remains to be determined how neurons respond to a full range of motion directions and how depth cues influence the direction tuning to multiple moving stimuli. In response to a single stimulus, MT neurons encode motion direction and disparity independently such that the response tuning for a single motion direction does not change its shape as the disparity of the moving stimulus is varied (Smolyanskaya et al. 2013). It is still unclear how the disparities of multiple stimuli influence the response tuning to multiple motion directions.

Attention can bias neuronal responses elicited by multiple 2-dimensional stimuli in the receptive field (RF) in favor of the attended stimulus (Desimone and Duncan 1995; Reynolds et al. 1999; Li and Basso 2005; Treue and Maunsell 1996; Ferrera and Lisberger 1997; Treue and Martinez-Trijillo 1999; Recanzone and Wurtz 2000; Wannig et al. 2007). The effects of selective attention on the neural representation of multiple moving stimuli located at different depths in 3D space remain to be determined.

We aimed to determine the roles of stereoscopic depth, in particular horizontal disparity, and selective attention in the neural representation multiple stimuli moving in different directions. To address this question, we characterized how neurons in area MT of macaque monkeys represented two overlapping surfaces located at different horizontal disparities and moved simultaneously in different directions. Area MT is known for its importance in motion and depth

perception and MT neurons are tuned to both motion direction in the frontoparallel plane and horizontal disparity (Maunsell and Van Essen 1983b; Bradley et al. 1995; DeAngelis and Newsome 1999; DeAngelis and Uka 2003). We recorded the activities of MT neurons while two animals performed a novel direction-discrimination task and reported the motion direction of an attended surface of two overlapping stimuli located at different disparities. To determine the role of attention, we also recorded MT activities while the attention of a third animal was directed away from the RFs. We characterized the full direction tuning curves of MT neurons to multiple motion directions at different disparities, as we manipulated the angular separation of two motion directions and selective attention. We tested the hypothesis that disparity cues bias the direction tuning curves of MT neurons elicited by multiple motion directions to better represent one surface of overlapping stimuli and therefore facilitate segmentation. We further hypothesized that such disparity bias was not caused by selective attention but can be modified by attention. We found that MT responses to overlapping stimuli showed a robust bias toward the horizontal disparity of one of two surfaces, and the disparity bias can be modified but not dependent on selective attention. We also found evidence that the effect of attention in our study was object/surface-based rather than feature-based. We devised a unified model with a variable pooling size to weigh the response to individual stimulus components and divisive normalization. The model was able to explain our finding of disparity bias and individual differences across animals.

Results

We asked how the visual system utilizes motion and stereoscopic depth cues jointly to segment multiple visual stimuli. For this, we characterized how neurons in the middle-temporal (MT) cortex of macaque monkeys represent overlapping stimuli that were located at different depths and moving in different directions, while monkeys performed a direction discrimination task.

Behavioral task and performance

We trained two male macaque monkeys (B and G) to perform a direction discrimination task (see Methods). Visual stimuli were two overlapping random-dot surfaces moving simultaneously in two different directions at two depths of -0.1° (near) and 0.1° (far) horizontal disparity. The two motion directions were separated by either 60° or 120° and the vector average (VA) direction was varied randomly across 360° in a step of 30°. The monkey's task was to report the motion direction of the surface at a cued depth (either near or far) by making a saccadic eye movement to one of twelve reporting targets (a 12 AFC task) (Fig. 1A, B). To perform the task well, the monkeys must segment two overlapping surfaces and identify the direction of the cued surface. We randomly interleaved the experimental trials in which either the near surface (Fig. 1A) or the far surface (Fig. 1B) was cued. We referred to these trials as "Attend Near" and "Attend Far" trials, respectively. Also randomly interleaved were trials in which only a single surface was presented at either the near (-0.1°) or far (0.1°) disparity ("Single Near" and "Single Far"). Our data set included 49 behavior and neural recording sessions from monkey B, and 44 sessions from monkey G. Across the single- and two-surface trials, monkey B performed at a mean correct rate of 75% (STD = 6.6) (single-surface 87%, two-surface 72%), and monkey G performed at 72% correct (STD = 8.1) (single-surface 87%, two-surface 68%). As expected, both monkeys

performed significantly better in the single-surface trials than in the two-surface trials (signed rank test, $p < 10^{-8}$). Notably, while monkey B performed equally well in the "Attend Near" and "Attend Far" trials of overlapping stimuli (signed rank test, $p = 0.56$) (Fig. 1C, D), monkey G performed significantly better at discriminating the motion direction of the near-surface (mean correct rate 72%) than the far-surface (mean correct rate 65%) of overlapping stimuli (signed rank test, $p = 4.3 \times 10^{-13}$) (Fig. 1E, F).

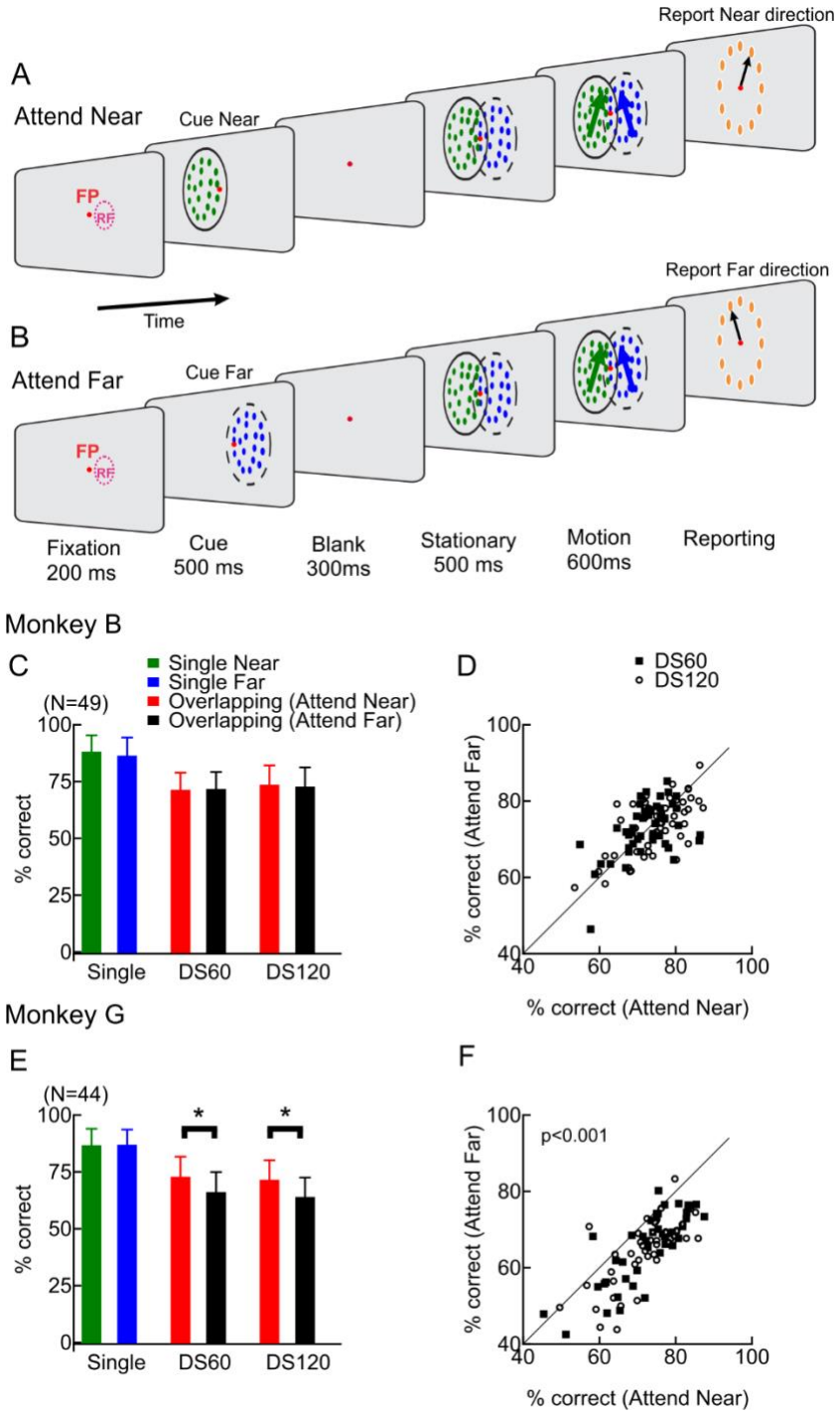


Figure 1. Discrimination task and behavioral performance.

A, B. Each trial started with fixation on a target at the center of the monitor, then a stationary random dot patch (RDP) was presented for 500 ms at either -0.1° (near) disparity (A) or at 0.1° (far) disparity (B) randomly as a cue. After 300ms blank period, either two overlapping RDPs- one at near and one at far disparity ("Attend Near" and "Attend Far" trials as shown in (A) and (B)) or a single RDP ("Single Near" and "Single Far" trials not shown in the figure) appeared stationary for 500ms, and then moved for 600ms. At the end of each trial, the monkey made an eye movement to report the motion direction of the RDP at the cued (Near or Far) disparity to receive a juice reward. **C.** Average % of correct choice trials across 49 recording sessions of Monkey B for Single (green: "Single Near", blue: "Single Far") and Overlapping motion directions (red: "Attend Near", black: "Attend Far"). DS60 and DS120 indicate 60° and 120° separations between overlapping motion directions. The error bars show standard deviations. **D.** % of correct choice trials of "Attend Near" conditions (or Cue Near) and "Attend Far" (or Cue Far) conditions are plotted on x and y axes, respectively. Each symbol represents one recording session. Solid squares and open circles are for DS60 and DS120 conditions. **E.** Same as **C** for Monkey G across 44 sessions. (*) indicates $p<0.001$, paired signed rank test. **F.** Same as **D** for Monkey G. p value is from one-sided signed rank test comparing x and y values.

Direction tuning curves of example neurons to overlapping stimuli moving at different depths.

Figure 2 shows the direction tuning curves (red and black) of four example MT neurons to overlapping stimuli moving in different directions at different depths, referred to as the "bi-directional stimuli". The direction tuning curves to the constituent near-surface and far-surface when they were presented alone, referred to as the "near component" and "far component" were shown in green and blue, respectively. If a neuron's maximum response of the directing tuning curve to the near component was greater than that to the far component, we classified the neuron as near-preferred ("near neuron" in brief), and vice versa as far-preferred ("far neuron").

Figures 2A1-D1 and 2A2-D2 show the responses of a near neuron and a far neuron from monkey B, elicited by the stimuli shown in Fig 2A-D (single component stimuli not shown in the illustration). The response tuning curve to the bi-directional stimuli of the near example neuron (red curve) showed a clear bias toward the tuning curve to the near component (green curve) when the monkey attended to the near surface (Fig. 2A1-D1). We referred to this response bias as a "near bias". The near neuron responded more strongly to the near than the far component, and therefore the average of the responses to the near and far components (gray dashed line) had a peak toward the near component. However, the red curve showed more bias toward the near component than expected by averaging the component responses (Fig. 2A1-D1). This near bias was found at both angular separations of 60° (Fig. 2A1, B1) and 120° (Fig. 2C1, D1). The near bias was also found regardless of whether the motion direction of the near component was at the clockwise (Fig. 2A1, C1) or counterclockwise side (Fig. 2B1, D1) of the two motion directions, suggesting that this bias was not toward a specific direction component. When monkey B attended to the far surface, the response to the bi-directional stimuli of the near neuron (black curve) still showed a near bias in some stimulus conditions (Fig. 2A1, C1, D1), but to a lesser degree than when the monkey attended to the near surface. In contrast, the far example neuron showed a clear bias (black curve) toward the far component when monkey B attended to the far surface (Fig. 2A2-D2). When the monkey attended to the near surface (red curve), the far bias was still clearly present in some conditions (Fig. 2C2, D2) but was weaker than when the monkey attended to the far surface. We also noticed that, in addition to the preference for one of the two binocular disparities, some neurons preferred

one of the two motion directions as previously reported by Xiao and Huang (2015) using two-dimensional stimuli. For example, the far neuron in Figure 2A2-D2 showed a stronger far bias when the far component was at the counterclockwise side (Fig. 2A2) than at the clockwise side of the two motion directions (Fig. 2B2). Our experimental design allowed us to separate the effect of disparity bias from the directional bias as shown below.

Figures 2A3-D3 and 2A4-D4 show the responses of a near neuron and a far neuron from monkey G. Remarkably, regardless of whether a neuron was near- or far-preferred and whether monkey G attended to the near or far surface, the responses to the bi-directional stimuli (red and black curves) showed a clear bias toward the near component. The responses to the bi-directional stimuli of these two neurons also showed little difference between the attending near and far conditions (comparing the red and black curves).

Population-averaged tuning curves to bi-directional stimuli moving at different depths

The results shown in the example neurons were representative across populations of neurons for monkeys B and G. Figure 3 shows the tuning curves to bi-directional stimuli averaged across near neurons and far neurons, at 60° and 120° angular separation. The direction tuning curve of each neuron was first fitted with a spline, normalized, and then shifted such that vector average (VA) direction 0° of the bi-directional stimuli was aligned with the neuron's preferred direction (see Methods). The spline-fitted tuning curves were then averaged across near neurons and far neurons. To balance the potential bias toward one of two motion directions, we averaged the response tuning curves to bi-directional stimuli when the near component was at the clockwise side (Near C-Far CC in Fig 2) and counterclockwise side of the two motion directions (Near CC-Far C). To do so, before averaging we flipped the tuning curves relative to VA 0° when the near component was at the counterclockwise side of the two motion directions to the clockwise side (see Methods). As a result, in Figure 3 the near component is always at the "clockwise" side of the two motion directions.

For monkey B, population-averaged tuning curves to bi-directional stimuli showed a clear bias toward neurons' preferred disparity, i.e., near neurons showed a bias toward the near component (Fig. 3A, C), whereas far neurons showed a bias toward the far component (Fig. 3B, D). The disparity bias was modulated by attention – near neurons showed a stronger near bias when monkey B attended to the near surface (Fig. 3A, C) and far neurons showed a stronger far bias when monkey B attended to the far surface (Fig. 3B, D). Nevertheless, the disparity bias cannot be explained by attention modulation. For example, far neurons still showed a clear far bias even when monkey B attended to the near surface (red curve in Fig. 3D).

In contrast, for monkey G, population-averaged tuning curves to bi-directional stimuli showed a robust bias toward the near component for both near neurons (Fig. 3E, G) and far neurons (Fig. 3F, H), and in trials when either the near or the far surface was cued for attention. We found very small attention modulation, if any, in the data from monkey G.

As an attention control, we asked whether MT neurons showed disparity biases without the involvement of attention. To do so, we trained a third monkey R to perform a fine direction discrimination task in the visual hemifield opposite from the RFs, while we presented the bi-

directional stimuli and individual stimulus components in the RFs (see Methods). The results from monkey R showed similar disparity bias as found in monkey B – near neurons tended to show a near bias (Fig. 3I, K) and far neurons tended to show a far bias (Fig. J, L). In other words, a neuron's response to bi-directional stimuli tends to show a bias toward the stimulus component located at the neuron's preferred disparity.

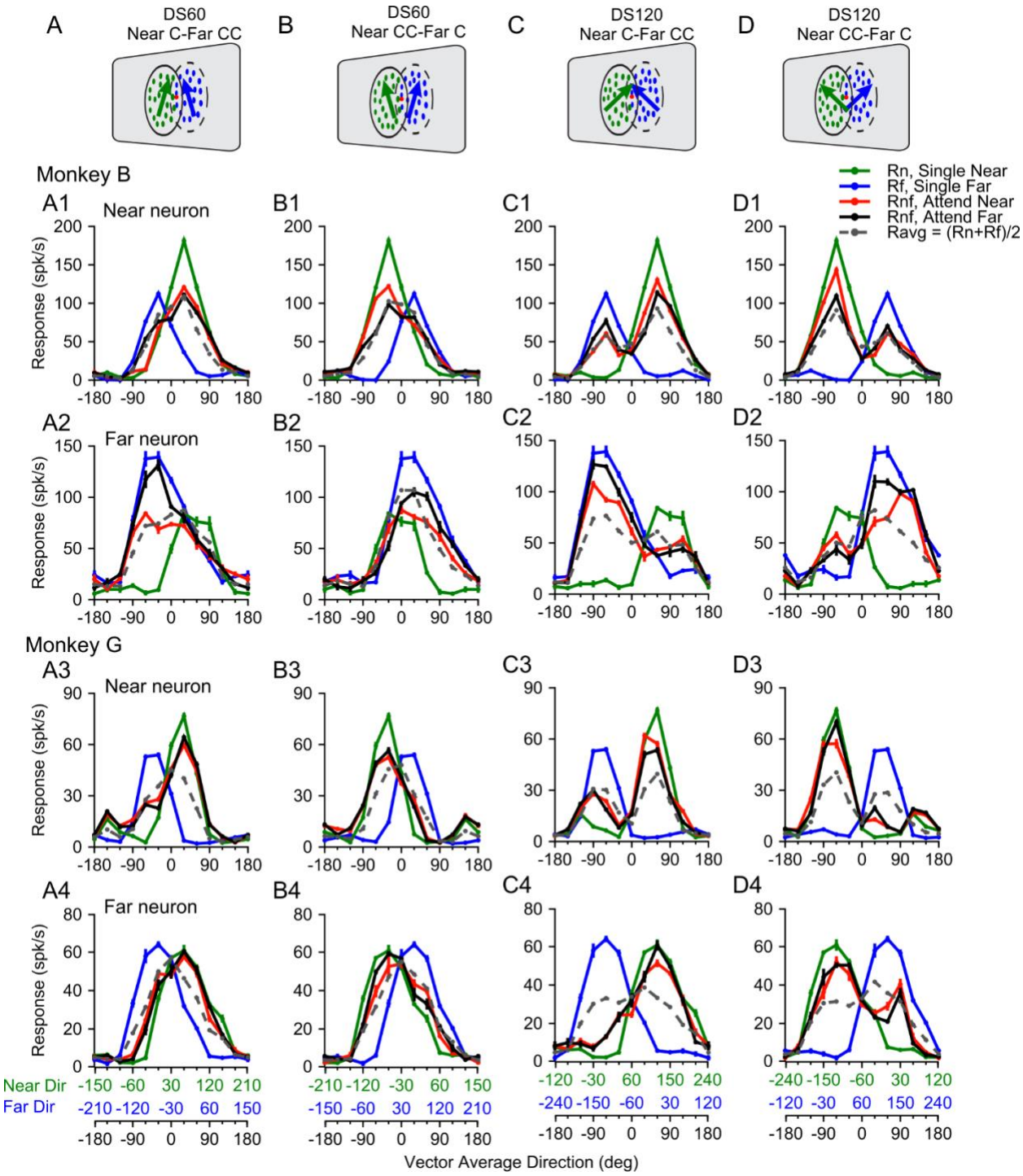


Figure 2. Example MT neurons' tuning curves to single and overlapping stimuli moving at different depths.

A-D. Schematic showing overlapping stimuli used in the discrimination task. Two RDPs (near and far surfaces in green and blue, respectively) moved in different directions such that their angle separation is either 60° or 120° (DS60 and DS120) with different angular arrangements: near motion direction on clockwise side and far motion direction on counterclockwise side (Near C-Far CC) and vice-versa (Near CC-Far C). Colors were used for illustrative purposes only. **A1-A4, B1-B4, C1-C4, D1-D4** are response tuning curves of example neurons to the stimuli shown in **A, B, C, D**, respectively, with the vector average direction (VA) of the two directions varied across 360° at a 30° step. **A1-D1.** Tuning curves of an example near neuron of monkey B to overlapping (bi-directional) stimuli (red: Attend Near and black: Attend Far conditions) and the single component directions (green: Single Near and blue: Single Far). Grey dashed lines show the average of the single component direction tuning curves (green and blue). **A2-D2.** Direction tuning curves of an example far neuron of Monkey B. **A3-D3.** Direction tuning curves of an example near neuron of Monkey G. **A4-D4.** Direction tuning curves of an example far neuron of Monkey G. The x-axes in green and blue show single near and far direction components, respectively, of the corresponding bidirectional stimuli for which the vector average (VA) direction is shown by the x-axis in black. A VA direction of 0° is aligned with the neuron's preferred direction (PD).

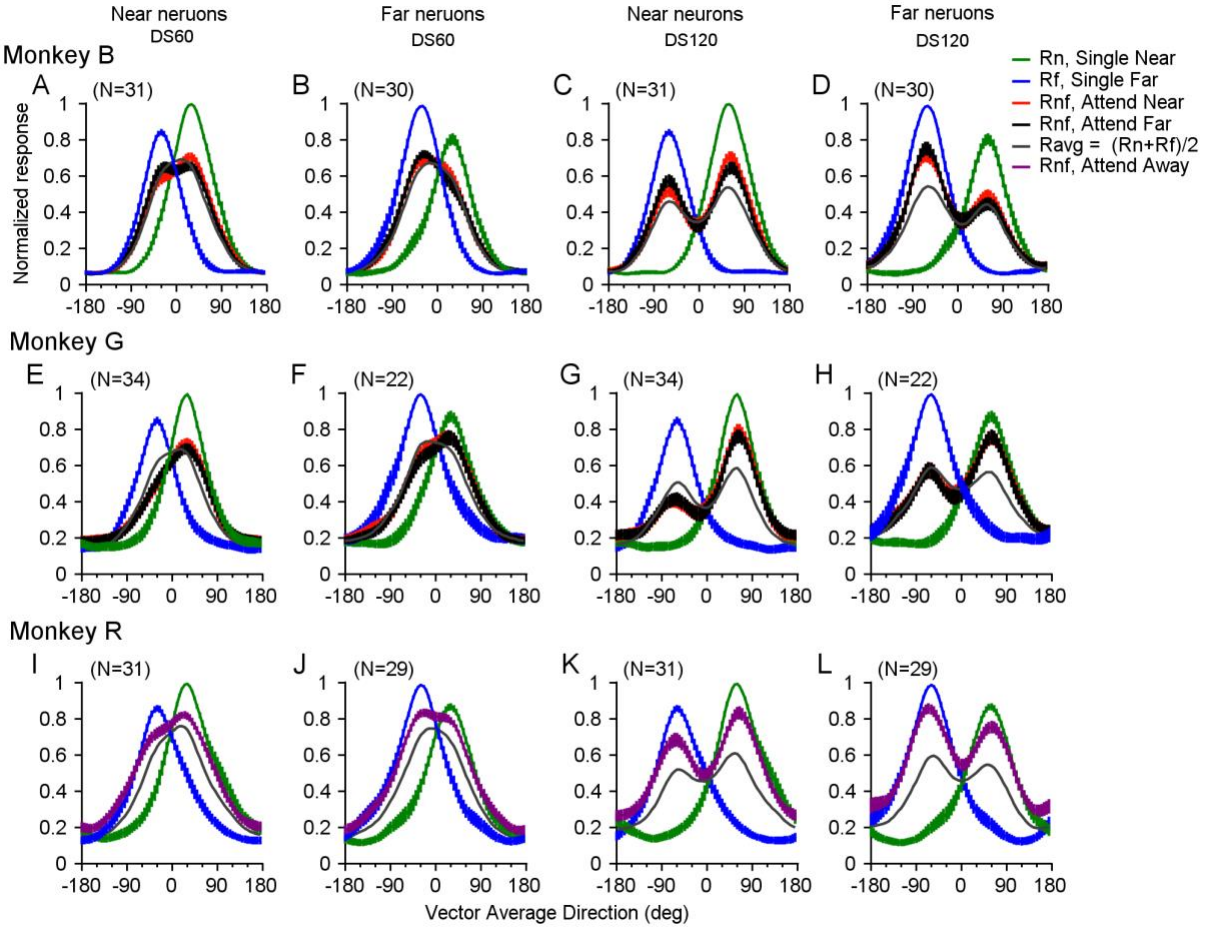


Figure 3. Population-averaged tuning curves to single and overlapping stimuli moving at different depths.

A, E, I. Tuning curves to bidirectional stimuli with 60° direction separation (DS60) averaged across all near neurons. **B, F, J.** Averaged across all far neurons. **C, G, K.** Tuning curves to bidirectional stimuli with 120° direction separation (DS120) averaged across all near neurons. **D, H, L.** Averaged across all far neurons. **A-D.** Monkey B. **E-H.** Monkey G. **I-L.** Monkey R from the Attend away paradigm. The width of

each curve indicates standard error. For each panel, Near CC-Far C tuning curves (like in Fig 2) were spline-fitted, normalized, and flipped along the x-axis (-180° to 180° VA) and were averaged with Near C-Far CC spline-fitted, normalized tuning curves across near or far neurons.

Response bias toward the near or far component and effect of attention

To quantify the response biases toward either the near or far component, we fitted the bi-directional tuning curves of each neuron (red and black curves as in Fig 2) as a weighted sum of the responses to the individual stimulus components (green and blue curves as in Fig 2). We found that a model of weighted summation plus a nonlinear interaction term (Eq. 1), referred to as the SNL model, provided a significantly better fit than a linear weighted summation (LWS) model shown in Eq. 13 (Supplementary Fig. 1, one-sided signed rank test, $p < 10^{-7}$, SNL model mean PV Monkey B =93%, Monkey G =83%, Monkey R =88%; LWS model mean PV: Monkey B PV =91%, Monkey G PV =81%, Monkey R PV =82%, see Eq.16 for PV).

$$R_{pred}(\theta_n, \theta_f) = w_n R_n(\theta_n) + w_f R_f(\theta_f) + b R_n(\theta_n) R_f(\theta_f) \qquad (1)$$

in which R_{pred} is the model-predicted response to bi-directional stimuli moving in directions θ_n and θ_f; R_n and R_f are the direction tuning curve to single motion directions at the near and far surface, respectively; w_n and w_f are two model parameters and represent the response weights for the near and far component, respectively; b is a third model parameter that determines the strength and sign of a multiplicative interaction between two component responses.

We defined a weight bias index (*WBI*) as: $WBI = (w_n - w_f)/(w_n + w_f)$. A positive *WBI* indicates a bias toward the near component, whereas a negative *WBI* indicates a bias toward the far component. Figure 4 shows the distributions of *WBI* across near and far neurons and the effect of attention. In the scatter plots (Fig. 4A, B), each neuron contributed to four data points, for 60° and 120° angular separation and the near component moving at either the clockwise or counterclockwise side of two motion directions, respectively. For monkey B, near neurons had significantly larger *WBI* than the far neurons regardless of whether the monkey attended to the near or far surface (one sided sign rank test, $p < 0.001$) (Fig. 4A). When the monkey attended to the near surface, near neurons showed a significant near bias (mean *WBI* = 0.12, p= 1.4310e-08, one sided sign rank test comparing mean to zero), and far neurons showed a significant far bias (mean *WBI* = -0.17, p= 1.9459e-08). When the monkey attended to the far surface, far neurons showed a significant far bias (mean *WBI* = -0.30, p= 5.7378e-15), and near neurons had a mean *WBI* of -0.01 which was not different from zero (p = 0.0537). Attention had a significant effect on *WBI*, giving rise to a larger *WBI* when monkey B attended to the near surface than the far surface across near neurons ($p = 5.4 \times 10^{-9}$), far neurons (p = 0.007), and all neurons combined (p= 3.0086e-10) (Fig. 4A).

For monkey G, however, both near and far neurons showed a significant near bias, regardless of whether the near surface or the far surface was cued for attention (p= 6.0931e-56). Furthermore, attention did not change *WBI* significantly across the near neurons (p = 0.29), far neurons (p = 0.59), and all neurons combined (p = 0.41). Nevertheless, near neurons still showed a significantly larger *WBI* than the far neurons when monkey G attended to the near surface (p <

0.01) and when the attending near and attending far conditions were pooled (one-sided rank sum test, p=0.002).

For monkey R, we found a similar result that the near neurons showed a significantly larger *WBI* than the far neurons when monkey R attended away from the RFs (p= 3.9412e-05) (Fig. 4C). Near neurons showed a significant near bias (mean *WBI* = 0.07, p = 4.3×10^{-4}) and far neurons showed a significant far bias (mean *WBI* = -0.03, p = 0.003). In summary, MT neurons in monkeys B and R showed a significant bias toward the stimulus component moving at the neurons' preferred disparity. Although neurons in monkey G showed a predominant near bias, near neurons also tended to show a stronger near bias than the far neurons.

Relationship between disparity bias in responses to two surfaces and disparity preference to single surface

To quantify the disparity bias and to balance the effect of potential directional bias, we defined a two-surface disparity bias index (*TDBI*) that pooled the response weights obtained when the near component moved either in the clockwise (*nC*) or counterclockwise (*nCC*) direction of the two motion directions from Eq.1 :

$$TDBI = \frac{(w_n - w_f)_{nC} + (w_n - w_f)_{nCC}}{(w_n + w_f)_{nC} + (w_n + w_f)_{nCC}} \qquad (2)$$

A positive *TDBI* indicates a bias toward the near component of two overlapping stimuli and a negative *TDBI* indicates a bias toward the far component. We examined the relationship between the disparity bias to one of the components of overlapping stimuli and a neuron's preference for the disparity of a single surface, measured by a single surface disparity preference index (*SDPI*):

$$SDPI = \frac{\max(R_n) - \max(R_f)}{\max(R_n) + \max(R_f)} \qquad (3)$$

in which R_n and R_f *are* the direction tuning curves to the near and far surfaces when presented alone, as defined in Eq. 1. A positive and negative *SDPI* indicate a preference for the near and far disparity, respectively. Figure 5 shows that, for all three animals, the *TDBI* is significantly correlated with the *SDPI* ($p < 0.01$) (Pearson correlation coefficient r = 0.57, 0.31, and 0.24 for monkeys B, G, and R, respectively). The positive correlation was robust regardless of whether monkeys B and G attended to the near or far surface (Fig. 5A, B), or whether attention was directed away from the RF (Fig. 5C). For monkey G, even though most of the neurons showed a near bias and preferred near disparity, the positive correlation between *TDBI* and *SDPI* still holds, indicating a common trend across animals.

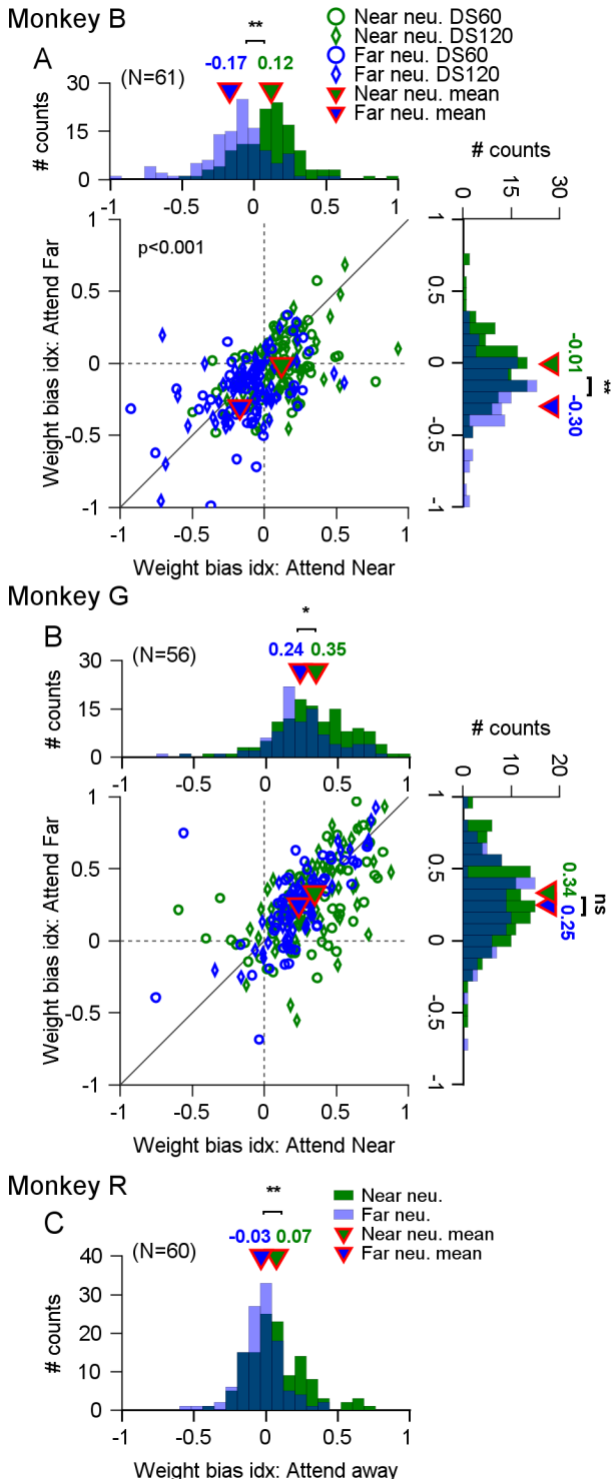


Figure 4. Effects of attention on the response weights of the near and far components.

A. Comparison of the weight bias indices (WBIs, $wbi = (w_n - w_f)/(w_n + w_f)$, Eq. 1) of Attend Near and Attend Far conditions, respectively from monkey B. Each dot represents the result from one bi-directional tuning curve of a neuron. Open circles and solid diamonds correspond to DS60 and DS120 conditions, respectively. Each neuron contributes to two of each of the symbols- one from Near C-Far CC and the other

from Near CC-Far C conditions. Green and blue dots denote near and far neurons. The triangles show the mean values for each group. p value in the scatter plot is of paired signed rank test. The histograms show the distributions of WBIs for the Attend Near (top) and Attend Far (right) conditions for near (green) and far (blue) neurons **B.** Same as **A** for monkey G. **C.** The histograms show the distributions of WBIs for the Attend away paradigm for near (green) and far (blue) neurons from monkey R. (**) indicates $p<0.001$, (*) indicates $p<0.01$, (ns) indicates $p>0.05$, one-sided signed rank test.

Time course of the response tuning to bi-directional stimuli moving at different depths

To understand the nature of disparity bias and the underlying mechanisms, we asked the following questions. First, did the disparity bias emerge at the beginning of the neural response or build up over time? Second, what was the relationship between the disparity bias and attention modulation? Third, did the neuronal population in MT show an overall bias to the near or the far disparity of two surfaces? To answer these questions, we characterized the time course of the response tuning curves to the bi-directional stimuli moving at different depths.

Figure 6 shows the population-averaged tuning curves of monkey B to the bi-directional stimuli, calculated within a sliding time window of 50 ms at a step of 10 ms. As in Figure 3, we combined the conditions in which the near component moved at either the clockwise or the counterclockwise side of the two directions (see Methods). In the combined responses, we plotted the near component at the clockwise side of the two motion directions. As such, at the right side of VA direction 0º (i.e., positive VA directions), the near component moved in a more preferred direction of the neurons and elicited a stronger response than the far component. Specifically, at the VA direction of 30º and 60º, the near component moved in the preferred direction (0º) of the neuron when the angular separation was 60º and 120º, respectively. We referred to VA 30º or 60º as "Near PD". Conversely, at the left side of VA direction 0º, the far component moved in a more preferred direction and elicited a stronger response (Figs. 6A1-D1). At the VA direction of -30º and -60º, the far component moved in the preferred direction (0º) at the angular separation of 60º and 120º, respectively. We referred to VA -30º or -60º as "Far PD". When the response tuning to the bi-directional stimuli is biased toward the near component, the tuning curve should skew to the positive VA directions. Likewise, when the bias is toward the far component, the tuning should skew to the negative VA directions.

At 60º angular separation, the near neurons showed a slightly delayed near bias, starting about 40 ms after the response onset (Fig. 6A2, A3). Attending the near surface made the near bias strong and persistent over time (Fig. 6A2). Attending the far surface did not abolish the initial near bias, but later shifted the tuning back and forth between the near and far components and made the tuning broader than attending near (Fig. 6A3). Consistently, the far neurons showed a delayed far bias (Fig. 6B2, B3). Attending the far surface made the far bias more robust and persistent over time (Fig. 6B3) than attending near. Nevertheless, far neurons still showed clear far bias even when attention was directed to the near surface (Fig. 6B2). For both the near and far neurons, the average of the responses to the near and far components (*Ravg*) was roughly balanced over time around VA 0º, with a slight near (Fig. 6A4) and far (Fig. 6B4) skewness.

Similarly, at 120º angular separation, the near neurons showed a near bias (Fig. 6C2, C3) and the far neurons showed a far bias (Fig. 6D2, D3). The disparity bias appeared sooner than that at 60º angular separation. Selective attention modulated the responses but was not sufficient to

change the disparity bias according to attention (Fig. 6C3, 6D2). For both the near and far neurons, the *Ravg* was slightly skewed toward the neurons' preferred disparity during the early response (Fig. 6C4, 6D4).

To separate the disparity bias from the effect of selective attention, we averaged the tuning curves obtained from the attending near and attending far conditions (Fig. 7A1-D1). At both 60º and 120º angular separations, we found that the near neurons showed an earlier, initially stronger, and more transient near bias (Fig. 7A2, C2) than the far bias of the far neurons (Fig. 7B2, D2). Furthermore, the near bias at 60º angular separation was slightly delayed relative to that at 120º angular separation (comparing Fig. 7A2 with C2, and Fig. 7B2 with D2). We further examined the disparity bias as the deviation from the average of the component responses by subtracting *Ravg* from the attention-averaged responses (Fig. 7A2-D2). At 60º angular separation, the disparity bias was reflected in response facilitation at the biased side, and suppression near VA 0º and the other side of the VA directions (Fig. 7A2, B2). At 120º angular separation, the disparity bias was reflected in stronger response facilitation on the biased side than on the other side (Fig. 7C2, D2). The suppression near VA 0º was stronger for the near neurons than the far neurons (comparing Fig. 7A2 vs B2, and 7C2 vs D2).

To examine the effects of selective attention, we subtracted the response of one attention condition from the other (Fig. 7A3-D3). For near neurons, the attention modulation was reflected in response facilitation at positive VA directions around "Near PD", and/or suppression at negative VA directions around "Far PD" (Fig. 7A3, C3). The facilitation (Fig. 7A3) and suppression (Fig. 7A3, C3) occurred as soon as the response onset. In contrast, for far neurons, the attention modulation was mainly reflected in much-delayed facilitation at negative VA directions around "Far PD" (Fig. 7B3, D3).

Figures 8 and 9 show the results from monkey G. The latency of the response to two surfaces in depths was about 30 ms delayed relative to that of monkey B (indicated by white dashed lines in Fig. 8A2), whereas the response latency to 2D stimuli showed no difference between the two animals (results not shown). At 60º angular separation, the near neurons of monkey G showed the development of near bias over a period of about 80 ms, starting from around 110 ms to fully biased to the near side around 190 ms (Fig. 8A2, A3; Fig. 9A1, 9A2). At 120º angular separation, the near neurons showed an earlier and stronger near bias than at 60º angular separation (Fig. 8C2, C3; Fig. 9C1, C2). The near bias was stronger and more persistent when attention was directed to the near surface (Fig. 8A2, C2) than to the far surface (Fig. 8A3, C3). Near neurons showed a small effect of attention modulation, mainly reflected in response facilitation at the positive VA directions around "near PD" when the response of attending far was subtracted from attending near (Fig. 9A3, C3). There was small initial facilitation toward negative VA directions around far PD at about 100 ms after motion onset (Fig. 9A3, C3), suggesting that the monkey may have initially attended the far surface briefly before attending the cued near surface.

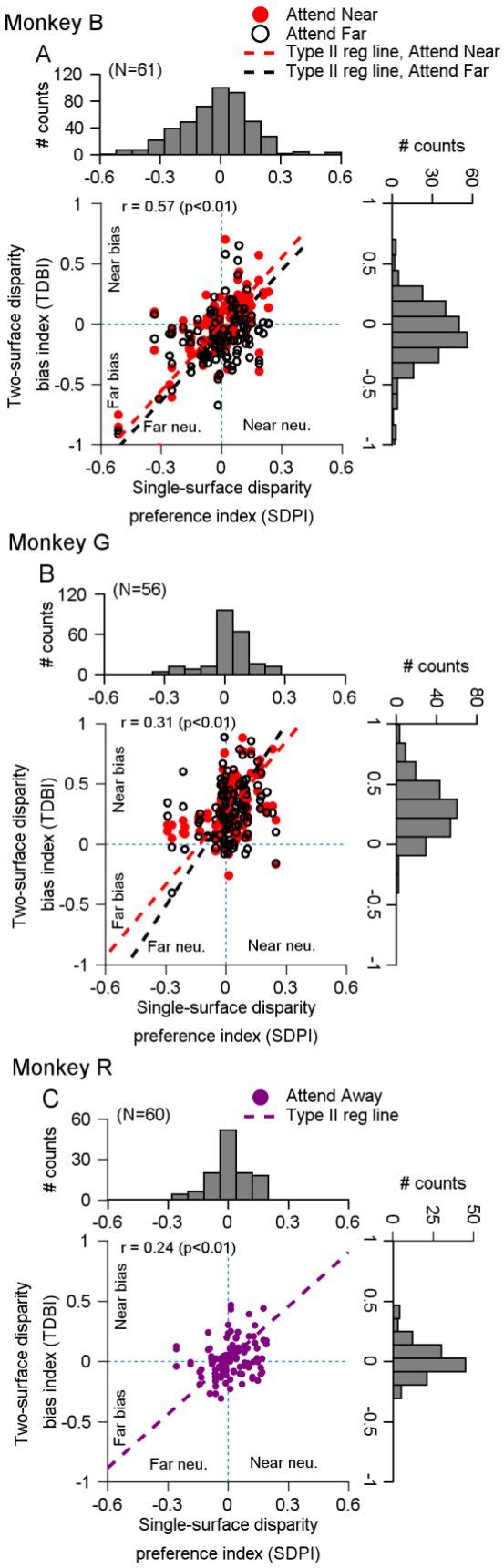

Monkey B
A
Attend Near
Attend Far
Type II reg line, Attend Near
Type II reg line, Attend Far
(N=61)
# counts
r = 0.57 (p<0.01)
Near bias
Far bias
Far neu.
Near neu.
Two-surface disparity bias index (TDBI)
Single-surface disparity preference index (SDPI)
Monkey G
B
(N=56)
r = 0.31 (p<0.01)
Monkey R
C
Attend Away
Type II reg line
(N=60)
r = 0.24 (p<0.01)

Figure 5. Relationship between MT neurons' response bias to the one of the two surfaces and their preference to single surfaces.

A. Comparison of Two-surface disparity bias indices (TDBIs) and Single-surface disparity preference indices (SDPIs) under Attend Near and Attend Far conditions, in red solid circles and black open circles, respectively from monkey B. Each neuron contributes to two symbols for each attention condition-one from DS60 and the other from DS120 conditions. The dotted lines are the type 2 regression fits. r is the Pearson correlation coefficient between x and y values of Attend Near and Attend Far (red and black symbols) together with the corresponding p values in brackets. The histograms show the distributions of TDBIs (top) and SDPIs (right). **B.** Same as **A** for monkey G. **C.** Same as **A** with Attend away paradigm from monkey R.

Remarkably, far neurons of monkey G also showed near bias (Fig. 8B2, B3, D2, D3; Fig. 9B1, D1) that had a similar time course as that of the near neurons (Fig. 9A2-D2). The responses of the far neurons to the bi-directional stimuli were robust and sustained at the positive VA directions around "near PD" throughout the motion period, making the tuning curve broader than the near neurons (comparing Fig. 9A1, B1). Far neurons showed little attention modulation at 60º angular separation (Fig. 9B3) and showed small and delayed facilitation at the negative VA directions around "far PD" at 120º angular separation when the response of attending near was subtracted from attending far (Fig. 10D3). Like near neurons, far neurons showed a small initial facilitation toward positive VA directions around near PD at about 100 ms (Fig. 9B3, D3), also suggesting that the monkey may have attended the other surface briefly before attending the cued surface. Regardless of the attention modulation, the near bias was robust throughout the motion response period.

When attention was directed away from the RF, the near and far neurons of monkey R showed disparity bias toward their preferred disparity roughly 50 ms after the response onset (Fig. 10). Within the early response period from 100 ms to 160 ms, the near bias of the near neurons was stronger than the far bias of the far neurons (Fig. 10A4-C4). For far neurons, the initial responses were roughly balanced between VA directions around the far PD and near PD and the far bias only became evident after 160 ms (Fig. 10B4, C4).

To address whether MT neurons showed an overall bias to the near or the far disparity of two surfaces, we pooled the results of near and far neurons for each of the three animals and examined the time course of $(R_{AttendNear} + R_{AttendFar})/2 - R_{avg}$ for monkeys B and G, and $R_{AttendAway} - R_{avg}$ for monkey R (Fig. 11 A1 to A3 and B1 to B3). To quantify the difference between the responses at the positive and negative VA directions, we flipped the responses at the negative VA directions relative to VA 0º and subtracted the flipped responses from the responses at the positive VA directions (Fig. 11 C1 to C3 and D1 to D3). For all three animals, there was an initial near bias (marked by the horizontal dashed lines in Fig 11). For monkey B, the later response after 120 ms showed a net far bias (Fig. 11 A1 to D1), whereas for monkey G, the near bias persisted in both earlier and later responses (Fig. 11 A2 to D2). For monkey R, the later response was roughly balanced, except for more near bias after 350 ms (Fig. 11 A3 to D3). Taken together, the time course results showed that regardless of whether attention was engaged, the neural responses across three animals were initially biased toward the near surface of two overlapping stimuli.

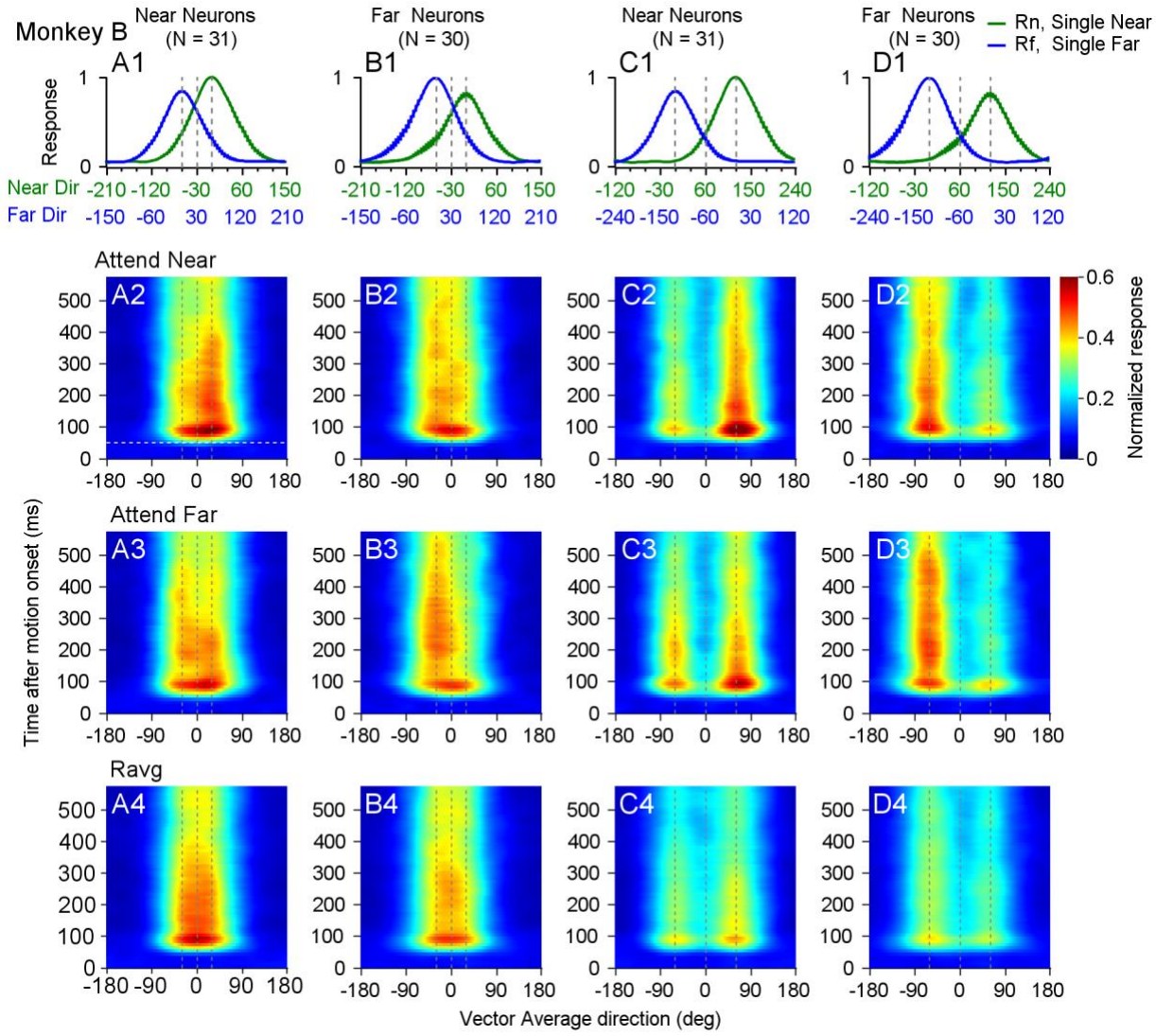


Figure 6. Time course of the response tuning to bi-directional stimuli moving at different depths from monkey B.

A1-D1. Responses to single component near and far directions. Motion directions and population averaged responses are plotted on x and y axes, respectively. The white dashed horizontal line in A2 indicates the response onset. **A2-D2.** Responses to bidirectional stimuli from "Attend Near" condition. Vector average (VA) direction of two motion directions, time after motion onset, and population averaged responses are plotted on x, y, and z axes, respectively. **A3-D3.** Responses to bidirectional stimuli from "Attend Far" condition. **A4-D4.** Responses to single component Near and Far directions in (A1 to D1) averaged together (R_{avg}). **A1-A4.** Averaged across near neurons for DS60 stimuli. **B1-B4.** Averaged across far neurons for DS60 stimuli. **C1-C4**. Averaged across near neurons for DS120 stimuli. **D1-D4.** Averaged across far neurons for DS120 stimuli. Vertical dashed lines in each panel indicate the responses to bidirectional stimuli with preferred direction at the near surface/less preferred direction at the far surface ("Near PD"), VA direction (x-axis) 0°, and preferred direction at the far surface/less preferred direction at the near surface ("Far PD") from right to left, respectively.

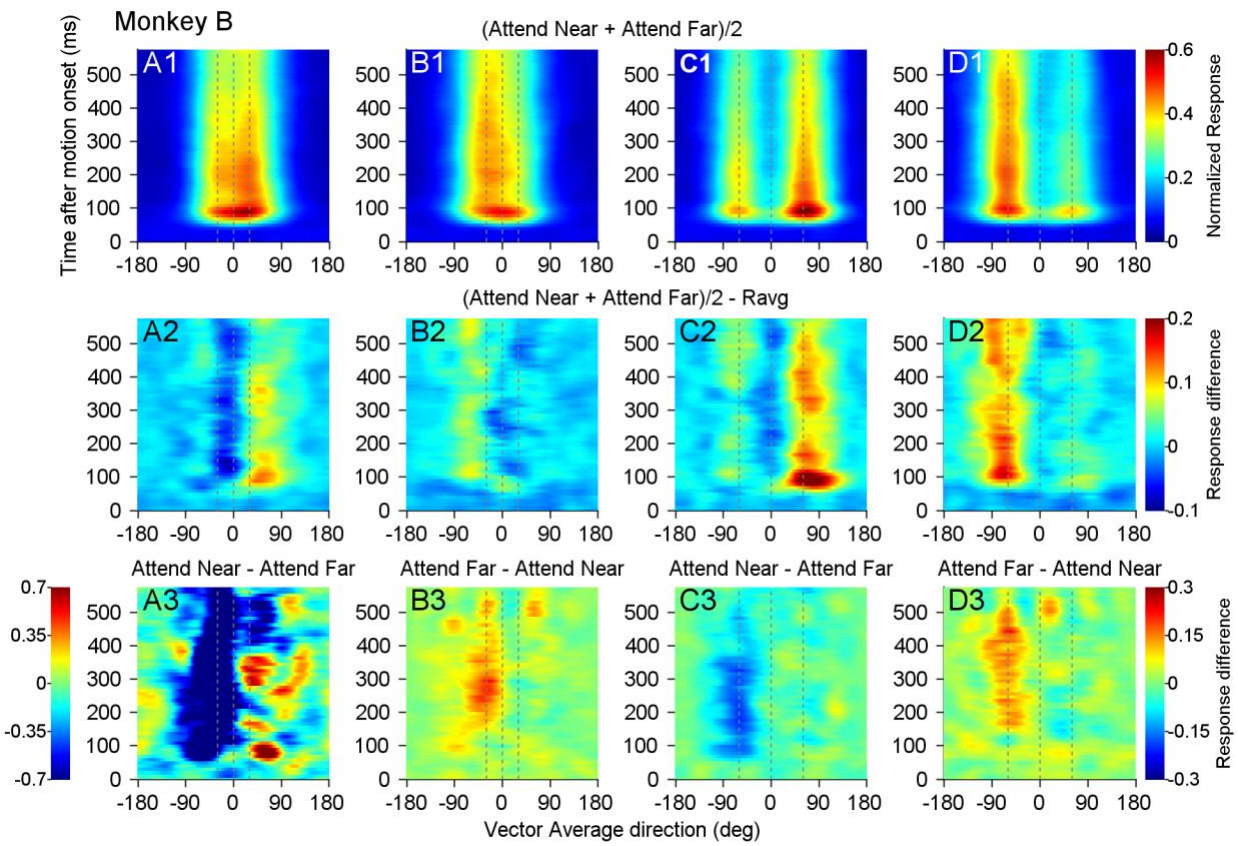


Figure 7. Temporal development of disparity bias and effects of selective attention from monkey B.

A1-D1. Responses to bidirectional stimuli in "Attend Near" (Fig 7 A2 to D2) and "Attend Far" (Fig 7 A3 to D3) conditions averaged together: $(R_{AttendNear} + R_{AttendFar})/2$. **A2-D2.** Difference between the responses from (Fig 7 A1 to D1) and (Fig 6 A4 to D4): $(R_{AttendNear} + R_{AttendFar})/2 - R_{avg}$. **A3, C3.** Difference between the responses from (Fig 6 A2, C2) and (Fig 6 A3, C3): $R_{AttendNear} - R_{AttendFar}$. **B3, D3.** Difference between the responses from (Fig 6 B3, D3) and (Fig 6 B2, D2): $R_{AttendFar} - R_{AttendNear}$. **A1-A3.** Averaged across near neurons for DS60 stimuli. **B1-B3.** Far neurons DS60. **C1-C3.** Near neurons DS120. **D1-D3.** Far neurons DS120. Vertical dashed lines in each panel indicate the responses to bidirectional stimuli with preferred direction at the near surface/less preferred direction at the far surface ("Near PD"), VA direction (x-axis) 0°, and preferred direction at the far surface/less preferred direction at the near surface ("Far PD") from right to left, respectively. Panels in each row have the same color bar scale shown on the right, except for A3 whose color bar is shown on the left of the panel.

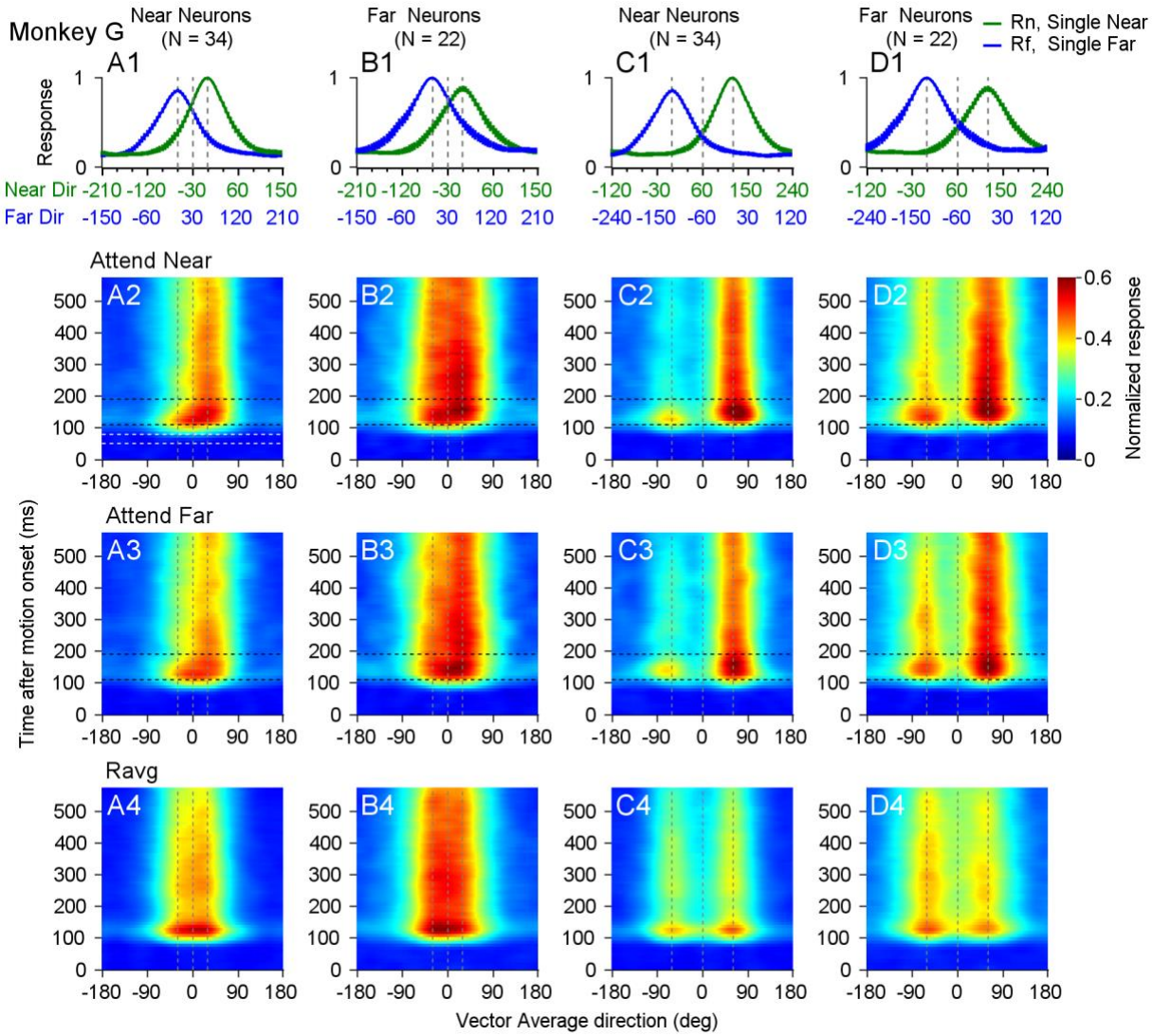


Figure 8. Time course of the response tuning bi-directional stimuli moving at different depths from monkey G.

A1-D1. Responses to single component near and far directions. Motion directions and population averaged responses are plotted on x and y axes, respectively. **A2-D2.** Responses to bidirectional stimuli from "Attend Near" condition. Vector average (VA) direction of two motion directions, time after motion onset, and population averaged responses are plotted on x, y, and z axes, respectively. The bottom and top white dashed lines in **A2** indicate the response onset of Monkey B and G, respectively, for comparison. **A3-D3.** Responses to bidirectional stimuli from "Attend Far" condition. The two black horizontal dashed lines in **A2-D2** and **A3-D3** mark the period of the development of near bias. **A4-D4.** Responses to single component near and far directions in (A1 to D1) averaged together (R_{avg}). **A1-A4.** Averaged across near neurons for DS60 stimuli. **B1-B4.** Averaged across far neurons for DS60 stimuli. **C1-C4.** Averaged across near neurons for DS120 stimuli. **D1-D4.** Averaged across far neurons for DS120 stimuli. Vertical dashed lines in each panel indicate the responses to bidirectional stimuli with preferred direction at the near surface/less preferred direction at the far surface ("Near PD"), VA direction (x-axis) 0°, and preferred direction at the far surface/less preferred direction at the near surface ("Far PD") from right to left, respectively.

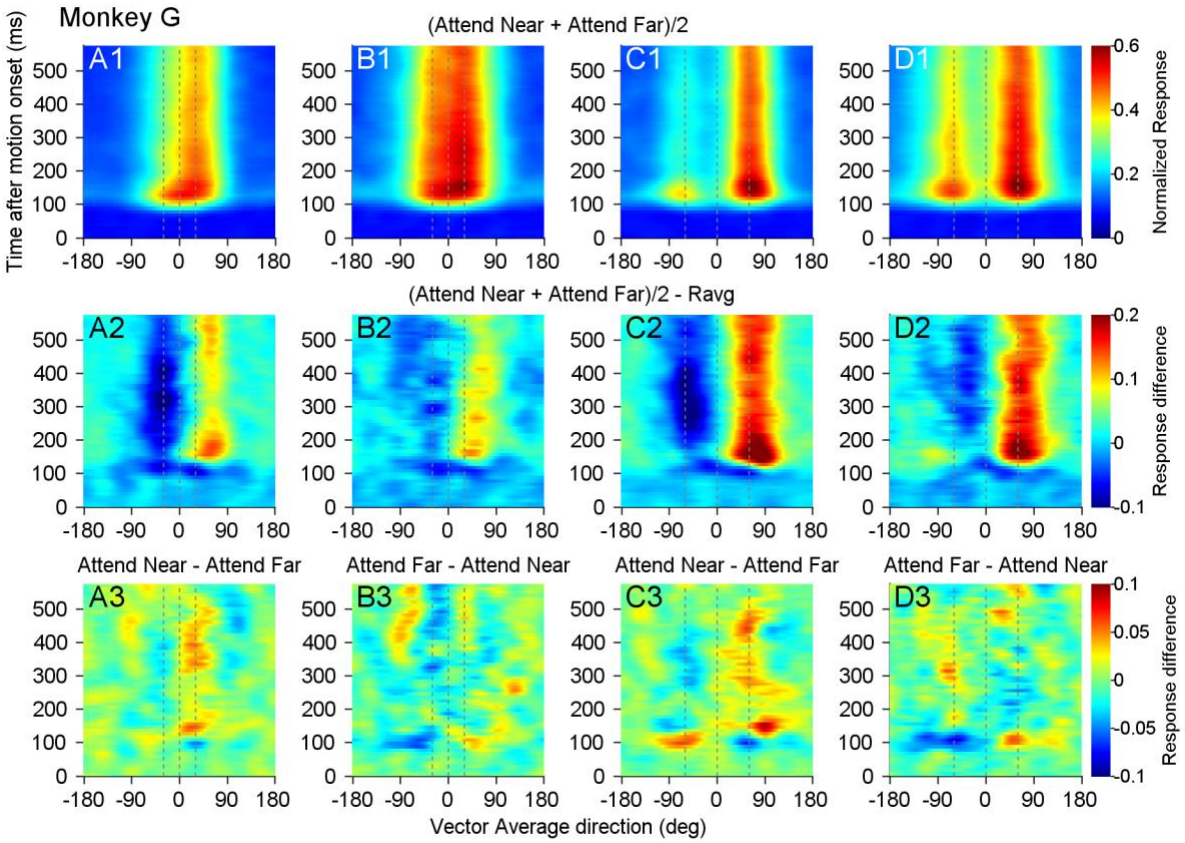


Figure 9. Temporal development of disparity bias and effects of selective attention from monkey G.

A1-D1. Responses to bidirectional stimuli in "Attend Near" (Fig 7 A2 to D2) and "Attend Far" (Fig 7 A3 to D3) conditions averaged together: $(R_{AttendNear} + R_{AttendFar})/2$. **A2-D2.** Difference between the responses from (Fig 9 A1 to D1) and (Fig 8 A4 to D4): $(R_{AttendNear} + R_{AttendFar})/2 - R_{avg}$. **A3, C3.** Difference between the responses from (Fig 8 A2, C2) and (Fig 8 A3, C3): $R_{AttendNear} - R_{AttendFar}$. **B3, D3.** Difference between the responses from (Fig 8 B3, D3) and (Fig 8 B2, D2): $R_{AttendFar} - R_{AttendNear}$. **A1-A3.** Averaged across near neurons under DS60 condition. **B1-B3.** Far neurons DS60. **C1-C3.** Near neurons DS120. **D1-D3.** Far neurons DS120. Vertical dashed lines in each panel indicate the responses to bidirectional stimuli with preferred direction at the near surface/less preferred direction at the far surface ("Near PD"), VA direction (x-axis) 0°, and preferred direction at the far surface/less preferred direction at the near surface ("Far PD"), from right to left, respectively.

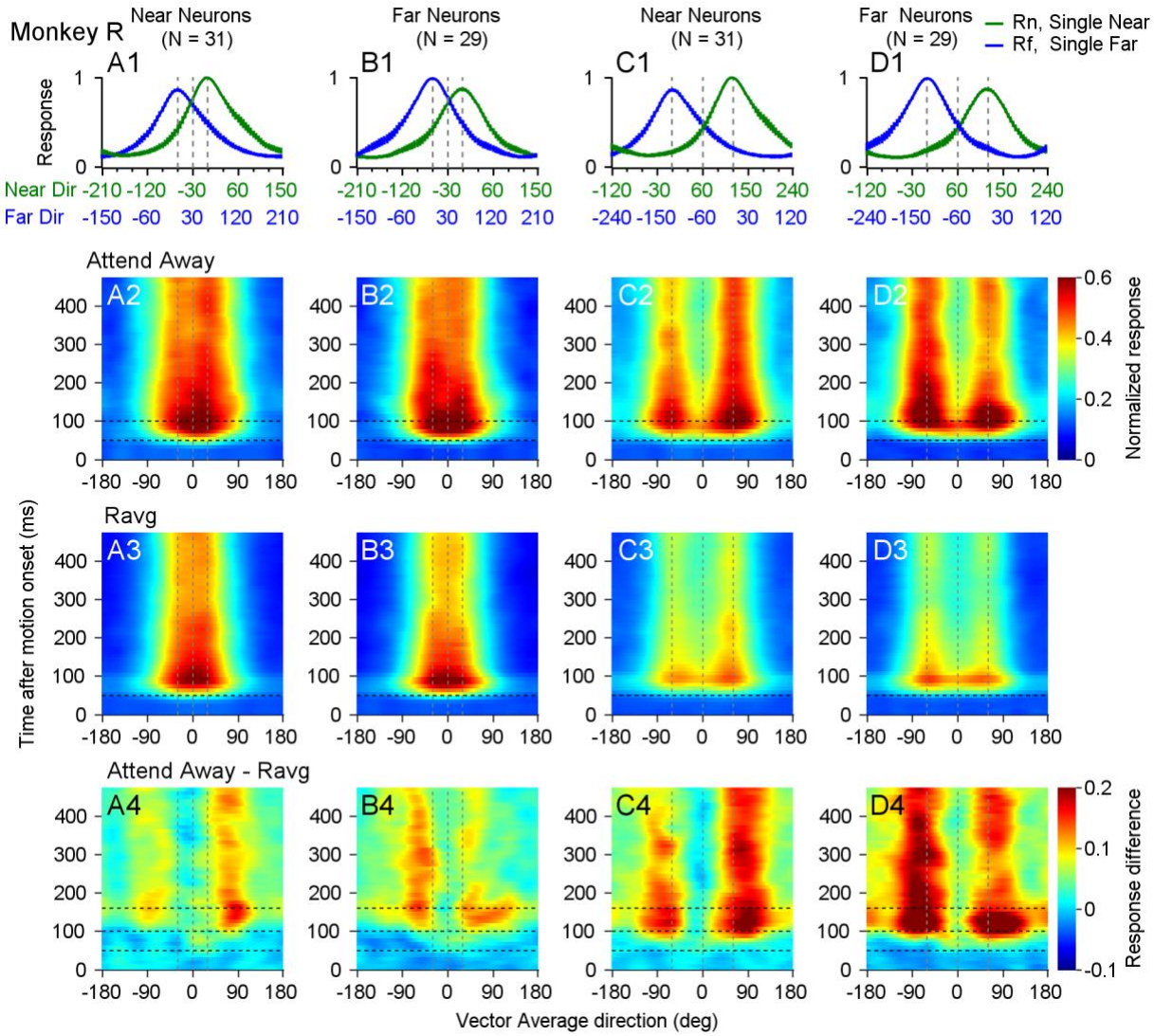


Figure 10. Time course of the response tuning to bi-directional stimuli moving at different depths in monkey R.

A1-D1. Responses to single component near and far directions. Motion directions and population averaged responses are plotted on x and y axes, respectively. **A2-D2.** Responses to bidirectional stimuli from "Attend Away" paradigm. Vector average (VA) direction of two motion directions, time after motion onset, and population averaged responses are plotted on x, y, and z axes, respectively. The bottom and top horizontal dashed lines mark 50 ms (response onset) and 100 ms (near bias onset), respectively. **A3-D3.** Responses to single component near and far directions in (A1-D1) averaged together (R_{avg}). The horizontal dashed lines mark 50 ms (response onset). **A4-D4.** Difference between the responses from (A1 to D2) and (A3 to D3): $R_{AttendAway} - R_{avg.}$ The bottom horizontal dashed lines mark 50 ms (response onset) and the middle and top lines mark 100 ms and 160 ms (window in which the near bias of the near neurons was stronger than the far bias of the far neurons), respectively. **A1-A4.** Averaged across near neurons for DS60 stimuli. **B1-B4.** Averaged across far neurons for DS60 stimuli. **C1-C4.** Averaged across near neurons for DS120 stimuli. **D1-D4.** Averaged across far neurons for DS120 stimuli. Vertical dashed lines in each panel indicate the responses to bidirectional stimuli with preferred direction at the near surface/less preferred direction at the far surface ("Near PD"), VA direction (x-axis) 0°, and preferred direction at the far surface/less preferred direction at the near surface ("Far PD") from right to left, respectively.

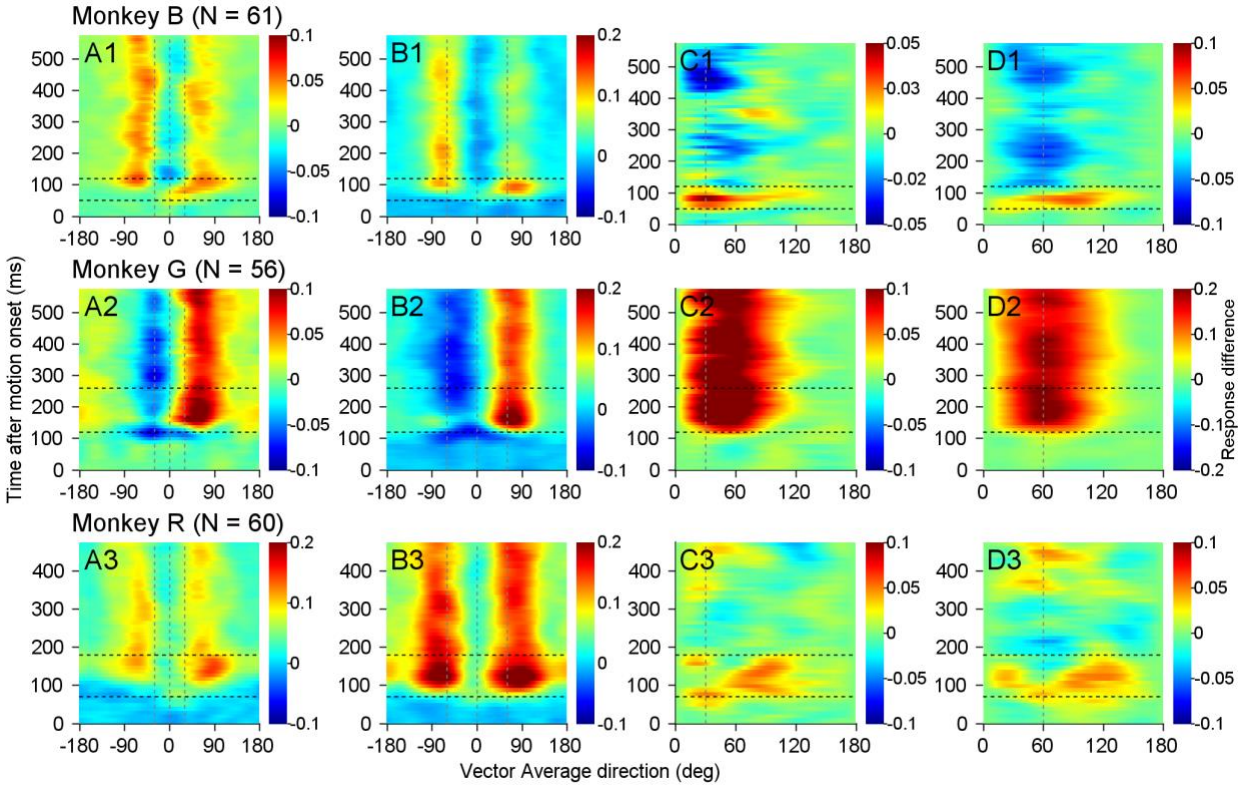


Figure 11. Temporal development of disparity bias across all neurons for three monkeys.

A1, A2. $(R_{AttendNear} + R_{AttendFar})/2 - R_{avg}$ averaged across near and far neurons for DS60 stimuli. **B1, B2**. $(R_{AttendNear} + R_{AttendFar})/2 - R_{avg}$ averaged across near and far neurons for DS120 stimuli. **A3.** $R_{AttendAway} - R_{avg}$ averaged across near and far neurons for DS60 stimuli. **B3.** $R_{AttendAway} - R_{avg}$ averaged across near and far neurons for DS120 stimuli. Vertical dashed lines in A1 to A3, B1 to B3 indicate the responses to bidirectional stimuli with preferred direction at the near surface/less preferred direction at the far surface ("Near PD"), VA direction (x-axis) 0°, and preferred direction at the far surface/less preferred direction at the near surface ("Far PD") from right to left, respectively. **C1-C3.** Difference between the responses from VA= 0° to 180° and those from VA= 0° to -180° in **A1-A3**, respectively. **D1-D3.** Difference between the responses from VA= 0° to 180° and those from VA= 0° to -180° in **B1-B3**, respectively. Vertical dashed lines in C1-C3 indicate VA= 30° and D1-D3 indicate VA= 60°. The horizontal dashed lines indicate the early near bias in all panels.

Attention modulation of neuronal response is object-based.

Figure 4 describes the effect of attention on the full response tuning curve to bi-directional stimuli, as measured by the impact on the weights for the near and far components. We asked, when an animal was cued to attend a stereoscopic depth (binocular disparity) of a surface, whether attention was disparity-based or object/surface-based. If attention was disparity-based (i.e., feature-based attention), we expected to find that when attention was directed to the preferred disparity of a neuron, the neuronal response would be enhanced regardless of the motion direction of the surface. On the other hand, if attention was object/surface-based, we expected to find that

when attention was directed to a stimulus component (i.e., a surface) that elicited a stronger response than the other component, the neuronal response would be enhanced regardless of whether the attended surface was located at the preferred or less preferred disparity of the neuron.

To answer this question, we measured the effect of attention when it was directed to a stimulus component that was located at either the preferred or less-preferred disparity and moving either in the preferred or less-preferred direction of the neuron. We quantified the effect using two attention modulation indices (AMI). One index *AMI_disparity* would be positive across conditions if the effect of attention was disparity-based. Whereas the other index *AMI_surface* would be positive if the effect of attention was surface-based.

$$AMI_disparity = \frac{R_{att_prefdsp} - R_{att_less_prefdsp}}{R_{att_prefdsp} + R_{att_less_prefdsp}} \quad (4)$$

in which $R_{att_prefdsp}$ was the response to the overlapping stimuli when attention was cued to the preferred disparity of the neuron, whereas $R_{att_less_prefdsp}$ was the response when attention was cued to the less-preferred direction.

$$AMI_\ surface = \frac{R_{att_strongerSurface} - R_{att_weakerSurface}}{R_{att_strongerSurface} + R_{att_weakerSurface}} \quad (5)$$

in which $R_{att_strongerSurface}$ was the response to the overlapping stimuli when attention was cued to a surface that elicited a stronger response, whereas $R_{att_weakerSurface}$ was the response when attention was cued to a surface that elicited a weaker response.

We found that the results from monkey B supported the idea that the attention effect was object/surface-based. Figures 12A and B showed the response tuning curves of two representative neurons from monkey B, one neuron preferred the near disparity (Fig. 12A) and the other preferred the far disparity (Fig. 12B). When attention was directed to a surface that elicited a stronger component response, regardless of whether the surface was located at the preferred or less-preferred disparity of the neuron, the response was always enhanced relative to attending the other surface (Fig. 12A, 12B). This trend was consistent across neuronal populations as captured by always positive *AMI_surface* values (Fig. 12C).

In contrast, when attention was directed to the surface located at the preferred disparity of the neuron (near in Fig. 12A, and far in 12B) and the surface moved in the less-preferred direction giving rise to a weaker component response, the response to both surfaces was reduced relative to attending the other surface (red curve at “Far PD” in Fig. 12A and black curve at “Near PD” in Fig. 12B). Furthermore, when attention was directed to the surface located at the less-preferred disparity of the neuron (far in Fig. 12A, and near in 12B) and the surface moved in the preferred direction giving rise to a stronger component response, the response was nevertheless enhanced (black curve at “far PD” in Fig. 12A and red curve at “near PD” in Fig. 12B). As the result, *AMI_disparity* averaged across the neuronal populations was not always positive (Fig. 12D), suggesting that the attention effect was not disparity-based.

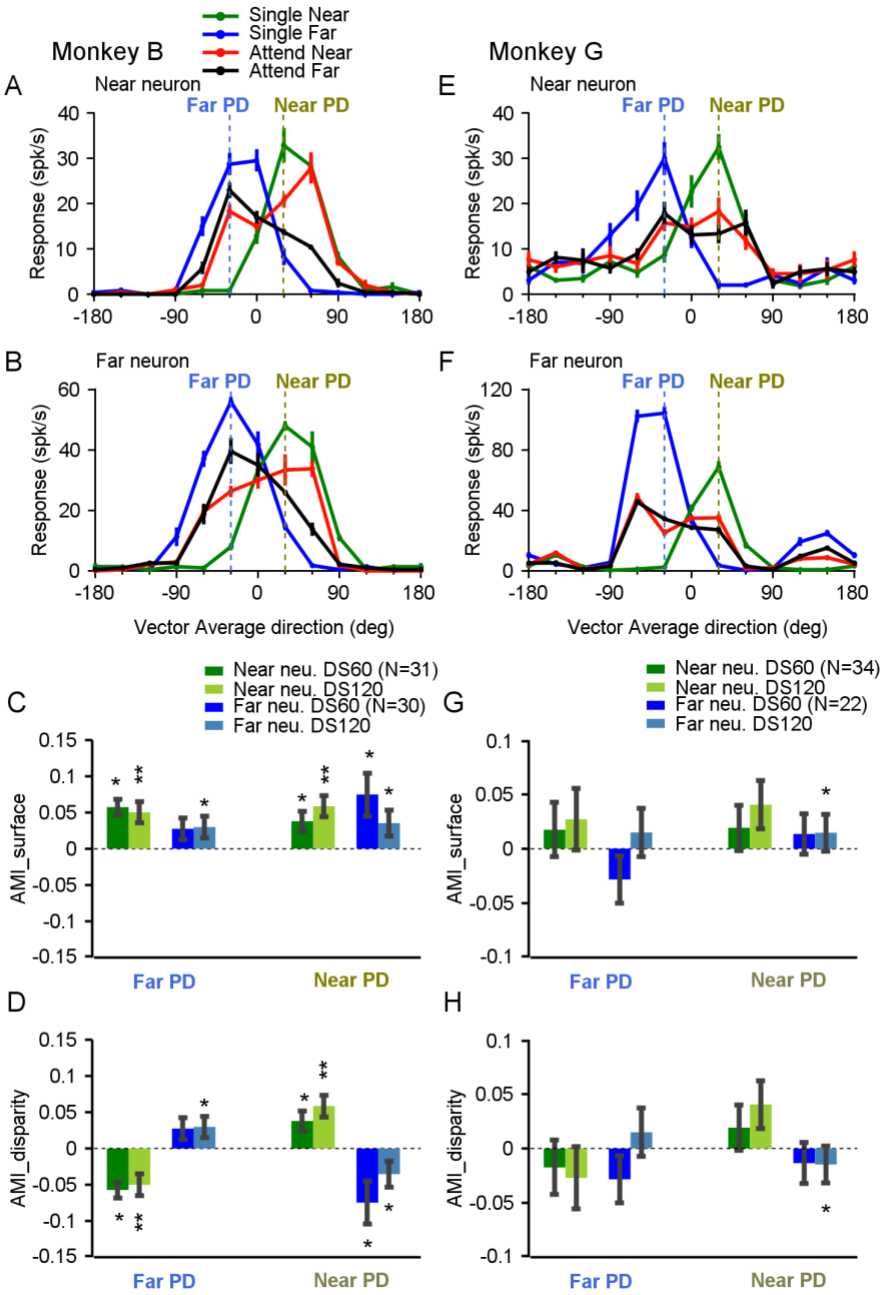


Figure 12. Attention modulation of MT neuronal responses.

A, B, E, F. Tuning curves of four example neurons to bidirectional stimuli separated in depth plotted in a format like that in Fig 2. The vertical dashed lines shown as "Near PD" and "Far PD" indicate responses elicited by the bidirectional stimuli with preferred direction at the near surface/less preferred direction at the far surface and by the stimuli with preferred direction at the far surface/less preferred direction at the near surface, respectively. **C, G.** Surface (object) based attention modulation indices (*AMI_surface* described by Eq. 5) calculated using responses to "Far PD" and "Near PD"- across Near neurons for DS60 and DS120 conditions (shown in two shades of green) and across Far neurons for DS60 and DS120 conditions (shown in two shades of blue). **D, H.** Disparity (feature) based attention modulation indices (*AMI_disparity* described by Eq. 4) calculated using responses to "Far PD" and "Near PD"- across Near

neurons for DS60 and DS120 conditions (shown in two shades of green) and across Far neurons for DS60 and DS120 conditions (shown in two shades of blue). **A-D**. Monkey B. Responses from 0 to 600 ms of motion period were used to calculate AMIs. **E-H**. Monkey G. Responses from 110 to 410 ms of motion period were used to calculate *AMI*s. (**) indicates p<0.01, (*) indicates p<0.05, no asterisk indicates p>0.05, one-sided signed rank test to indicate significance of difference from zero.

The attention effect found in monkey G was weaker than that in monkey B. Nevertheless, the effect of attention in monkey G was clearly visible in most of the conditions during the time window from 110 ms to 410 ms after the motion onset (Fig. 8A3, C3, D3). We therefore used the responses during this time window to calculate the direction tuning curves (Fig. 12E, F) and the AMI. Except for the far neurons at the angular separation of 60º, the trend of AMI of monkey G (Fig. 12G, H) was consistent with monkey B (Fig. 12C, D) and supported the idea that the attention modulation was object/surface-based. Note that, for most of the conditions, the effect found in monkey G did not reach statistical significance.

Fitting the neural responses to two moving surfaces using a unified model

From the results above, an MT neuron's response to overlapping moving stimuli separated in depth can be described by a weighted sum of the neuron's responses to the individual stimulus components presented alone. We found that the weights were dependent on the disparity preference of the neuron to single surface (Fig. 5) and selective attention (Fig. 4 and Fig. 12). While the general trend was consistent across animals, we also noted some pronounced individual differences. Specifically, for monkey G, both near and far neurons showed a dominant effect of near bias, albeit stronger near bias for the near neurons than the far neurons (Fig. 5). Whereas, for monkey B, the near neurons showed a near bias and the far neurons showed a far bias. We sought a unified model that can explain these results.

The divisive normalization model (Carandini and Heeger, 2011) has been used to explain a wide array of phenomena, including neuronal responses elicited by multiple visual stimuli (Britten and Heuer 1999; Heuer and Britten 2002; Busse et al. 2009; Xiao et al. 2014; Xiao and Huang 2015; Bao and Tsao 2018; Wiesner et al. 2020; Huang et al. 2023) and the effects of attention (Reynolds and Heeger 2009; Lee and Maunsell 2009; Ni et al. 2012; Ni and Maunsell 2019). However, in the normalization model, while the division by the activity of a population of neurons in the denominator (i.e., normalization) is well established, what accounts for the numerator is unclear. In the case of describing the response to multiple stimuli, the signal strength such as luminance contrast or motion coherence of each stimulus component was typically used in the numerator (e.g., Busse et al. 2009; Xiao et al. 2014). A caveat is that it is not always clear how to define the signal strength of a sensory stimulus (Wiesner et al. 2020). We have suggested that the weighting to a stimulus component of multiple stimuli is proportional to the activity of a population of neurons, referred to as the "weighting pool" (Xiao et al. 2014; Wiesner et al. 2020; Huang et al. 2023), and divisively normalized by the activity of the "normalization pool" as in the standard normalization model. The nature and scope of the "weighting pool" are not understood. Here, we further developed this idea and used the size of the "weighting pool" as a model parameter to account for the MT responses from monkeys B and G (Eq. 6 and 7).

$$R_{(n_a,f)} = \frac{\beta_{An} \bullet S^m_{near}}{\beta_{An} \bullet S^m_{near} + \alpha \bullet S^m_{far} + \sigma} R(\theta_n) + \frac{S^m_{far}}{\beta_{An} \bullet S^m_{near} + \alpha \bullet S^m_{far} + \sigma} R(\theta_f) + c \qquad (6)$$

$$R_{(n,f_a)} = \frac{S_{near}^m}{S_{near}^m + \alpha \bullet \beta_{Af} \bullet S_{far}^m + \sigma} R(\theta_n) + \frac{\beta_{Af} \bullet S_{far}^m}{S_{near}^m + \alpha \bullet \beta_{Af} \bullet S_{far}^m + \sigma} R(\theta f) + c \quad (7)$$

in which $R_{(n_a,f)}$ and $R_{(n,f_a)}$ are the model predicted responses to the bi-directional stimuli when the near surface and the far surface are attended, respectively. n_a *or* f_a denotes that either the near or far surface is attended. $R(\theta_n)$ and $R(\theta_f)$ are the measured responses elicited by the near and far components alone, moving in directions θ_n and θ_f, respectively. m, σ, and c are model parameters with the constraints of 0 < m < 100, σ > 0, and c > 0. A positive parameter α scales the relative contributions of S_n and S_f to the denominators as in tuned normalization (Ni et al. 2012; Rust et al. 2006; also see Carandini et al. 1997). β_{An} and β_{Af} are positive attentional weighting parameters when the near and far surfaces are attended, respectively. We derived the values of $\beta_{An}$and β_{Af} from the weights of the LWS model fitting (see Methods Eq. 15a and 15b).

In equations 17 and 18, S_{near} and S_{far} were the summed population neural responses of the "weighting pool" (***pR***) elicited by the stimulus disparity of the near and far components, respectively (Eq. 11, 12). Generally, for a given neuron *k*, S_k was the summed population neural responses of ***pR*** elicited by a single stimulus disparity d_{stim}, and weighted by a Gaussian window centered on the preferred disparity of the neuron *k* (i.e., d_{pref_k}) (Eq. 8-10). The standard deviation v of the Gaussian weighting function ***G*** was a model parameter and determined the broadness of ***G*** (Eq. 9), with the constraint of the disparity borders (from -1.6º to 1.6º) (see Methods). On one extreme when v was very large, the weighting pool included neurons covering a broad range of preferred disparities (orange lines in Fig. 13A for two example d_{pref_k} values), referred to as the "global pooling". On the other extreme when v was close to zero, the weighting pool only included neurons that had the same preferred disparity as the neuron *k*, referred to as the pure "local pooling" (brown lines in Fig. 13A). We determined the optimal v that gave rise to the best fit of the data of each animal and compared the optimal v between monkeys B and G (Fig. 13B, C) (see Methods).

The neuronal response of a population of neurons to a single stimulus disparity d_{stim} was a vector of length *N* and denoted as:

$$\boldsymbol{pR}(d_{\boldsymbol{stim}}) = [f_i(d_{stim},\ d_{pref,\ i})],\quad i = 1 \text{ to N}. \quad (8)$$

f_i was the disparity tuning curve of neuron *i*, in response to single disparities. $d_{pref,\ i}$ was the preferred disparity of neuron *i*. Because the number of neurons in our data set is relatively small, the percentage of the near and far neurons in our data may not reflect the true distribution of the disparity preference of MT neurons. We therefore used data of the disparity tuning curves of a large sample of MT neurons (*N* = 479) from DeAngelis and Uka (2003) to estimate ***pR***.

The Gaussian weighting function centered on the disparity d_{pref} of a neuron was a vector of length *N* and defined as:

$$\boldsymbol{G}(d_{pref}) = [\frac{1}{\sqrt{2\pi v}} e^{\frac{-(d-d_{pref})^2}{2v^2}}],\ d = -1.60\ to\ 1.60 \text{ and matched } dpref, i\ in\ \boldsymbol{pR} \quad (9)$$

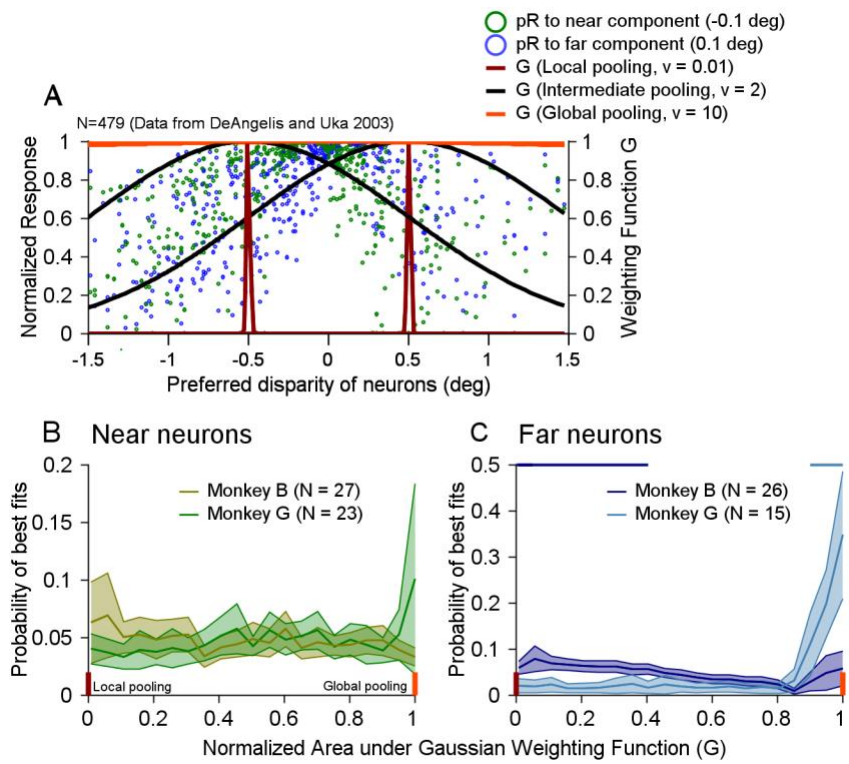


Figure 13. Probability distribution of sizes of the weighting pools from the divisive normalization model.

A. Green and blue dots show the normalized responses of 479 MT neurons (DeAngelis and Uka 2003) to -0.1° and +0.1° disparities moving in their preferred directions, denoted as pR (Eq. 8). Orange, brown, and black lines show examples of Gaussian weighting function (***G***) of global, local, and intermediate pooling strategies (Eq. 9). The area under ***G*** represents the size of the weighting pools. **B**. Probability distributions of the area under ***G*** s resulting from divisive normalization model (DNM) fits of the bidirectional tuning curves of individual neurons (see Methods). Yellowish green and green lines indicate the probability distributions from the fits of monkey B and G's near neurons, respectively. The error bands are 95% confidence intervals. **C**. Far neurons. Dark and light blue curves are for monkey B and G. The flat lines on top indicate the portions of significant difference between the monkeys. The colors of the flat lines match that of the monkey with significantly higher probability values (p<0.05, 95% confidence intervals estimated by bootstrapping).

For a given neuron k, S_k was the dot product of $G(d_{pref_k})$ and $pR(d_{stim})$, where d_{pref_k} is the preferred disparity of neuron k.

$$S_k(d_{stim}, d_{pref_k}) = \boldsymbol{G}(d_{pref_k}) \bullet \boldsymbol{pR}(d_{stim}) \qquad (10)$$

In our study, $d_{stim} = \pm 0.1^0$, so

$$S_{near}(d_{pref_k}) = S_k(-0.1^0, d_{pref_k}) \qquad (11)$$

$$S_{far}(d_{pref_k}) = S_k(0.1^0, d_{pref_k}) \qquad (12)$$

Due to the constraint of the range of preferred disparities in the data set, the area under the Gaussian weighting window ***G*** truncated by the disparity borders was a better descriptor of the size of the weighting window than the standard deviation ν (see Supplementary Fig. 2 for the relationship

between the two). We, therefore, varied the area under $\boldsymbol{G}$ to find the best fit of the neuronal response.

The model fitted the neuronal responses to two surfaces from monkeys B and G reasonably well. The mean PV was 89 % (STD = 11.2) for monkey B and 80% (STD = 17.5) for monkey G. We compared optimal weighting pool size that gave rise to the best fit across the near neurons (Fig. 13B) and the far neurons (Fig. 13C) for monkeys B and G. For the near neurons, monkey B had more neurons that were best fitted with local pooling and monkey G had more neurons that were best fitted with global pooling, but the differences did not reach the significance level ($p > 0.05$; 95% confidence intervals estimated by bootstrapping) (Fig. 13B). For the far neurons, local pooling fitted best for significantly more neurons from monkey B than monkey G (at normalized area of weighting window less than 0.4, $p < 0.05$), and global pooling fitted best for significantly more neurons for monkey G than monkey B (at weighting window area greater than 0.9, $p < 0.05$) (Fig. 13C). Remarkably, for 11 of 15 far neurons (70%) from monkey G, the optimal area of the weighting window was greater than 0.85, suggesting that global pooling was best to account for the results of far neurons from monkey G. In contrast, local and intermediate pooling was best to account for the results from monkey B. Taking together, we showed that a unified model with a variable size of the weighting pool and divisive normalization can explain our results of the disparity bias and the individual difference between the animals.

Discussion

We investigated the role of binocular disparity and selective attention in the neural representation of two transparently moving surfaces. We designed a novel motion discrimination task where macaque monkeys were cued using disparity to attend one of the two moving surfaces located at different depths. We found at least two schemes of how area MT encodes transparently moving surfaces separated in depth, both of which could be beneficial to the segmentation of the component surfaces. In one encoding scheme (seen in monkey B and R), in response to transparent motion surfaces separated in depth (one at a near disparity and the other at a far disparity), MT neurons showed a bias to the motion of the component surface closer to their preferred disparity. This response bias implies that subgroups of neurons defined by their disparity preferences (near and far neurons) can better represent the component surface located closer to their preferred disparities and create a distributed neural code to represent multiple surfaces. In the second encoding scheme (seen in monkey G), the responses of MT neurons to overlapping stimuli separated in depth were always biased to the motion of the near surface. The extent of this bias to the near surface was stronger for the near neurons than the far neurons. This near bias would benefit the segmentation of the near surface of overlapping stimuli. In natural scenes, if figural objects tend to be closer to the observers than the background (Huang et al. 2019 SfN abstract), the near bias of neuronal responses to multiple stimuli would benefit figure-ground segregation.

Role of disparity preference in the neural representation of multiple stimuli separated in depth

Previous studies showed that the neuronal response elicited by two stimuli within the RF can be described as a weighted sum of the responses elicited by the individual stimulus components (Qian and Andersen 1994; Recanzone et al.1997; Britten & Heuer 1999; Zoccolan et al. 2005; Xiao et al. 2014; Xiao and Huang 2015; Busse et al. 2009; Wiesner et al. 2020). Some of these studies have looked at what attributes of stimulus components might determine their response

weights, especially when the components differed in one or more features. For example, Xiao et al. (2014) found that rather than weighting each stimulus component equally, MT neurons weighted the stimulus component that had a higher signal strength more strongly, regardless of whether the signal strength was defined by motion coherence or luminance contrast. Analogously, in our study, the direction tuning curves of the near and far component surfaces had a similar selectivity but usually different amplitudes owing to the disparity tuning of MT neurons. In other words, neurons that preferred -0.1° disparity more than 0.1° show a higher response amplitude in the direction tuning to near component motion than the far component motion (near neurons), and vice-versa (far neurons). This distinction of two neuronal ensembles that preferred one component surface disparity over the other (across motion directions) when presented alone could be exploited by area MT. The distributed representations of the two surfaces by subgroups of neurons could add to the heterogeneity of the neuronal response, which has been shown to be beneficial for neural coding (Orhan & Ma, 2015).

Divisive normalization: taking contrast out of the equation

We used a divisive normalization model (DNM) framework to capture the heterogeneity of response weights across neurons. Several studies have used the normalization model to describe visual neurons' responses to two stimuli in their RFs that had different contrasts (Heuer & Britten 2002; Busse et al.2009; Xiao et al. 2014). The influence of adding a second stimulus to the first, on the neuronal responses, depended on the contrast values of the stimuli. To capture this contrast dependence, these studies used contrast in the numerator as well as in the denominator of the equation in the DNM model. Our stimuli had the same contrast but differed in motion direction and disparity. The neuronal responses to the overlapping stimulus across motion directions depended on neuronal disparity preferences. We took a novel approach to replace the contrast values of the stimulus components (referred to as "signal strength") in a typical DNM equation with population activity elicited by each component summed over a pool of neurons with varying disparity preferences. Additionally, a novelty of our model is its flexibility to define and vary the extent/size of the neuron pool that contributes to the weighting of component response in the numerator of DNM equation.

Disparity-based vs object/surface-based attention

An early psychophysics study (Snowden and Rossiter 1999) demonstrated that the role of stereo depth cues in transparent motion processing is distinct from other visual cues like color. In comparison to the color cues, depth cues gave a clearer impression of separate surfaces, which may then be selectively processed through an attentional mechanism. This emphasizes the importance of forming a percept of surfaces for the segmentation process. Is this surface-based perception important in our study? In our task, the monkeys selectively attended to the depth cues to report the motion direction of each surface. This process may be broken down to three steps: 1. conjugate the disparity and direction features forming a percept of a surface, 2. matching the disparity cue with the disparity of one of the overlapping surfaces, 3. retrieving the motion direction feature of that surface. It is important to note that even though disparity cues aided in the selection of one of the overlapping surfaces, disparity-based attention cannot explain the effects of attention on responses to preferred vs non-preferred direction seen in our data (Fig. 12). Instead, it appears that, to account for our results, attention may spread to the whole surface of a stimulus component. The attention effect seen in our data is consistent with object/surface-based attention (Wannig et al., 2007). Our study is unique to probe the effects of attention on the neural

representation of surfaces defined by two features (motion and disparity) that are conjointly encoded in one brain area (area MT).

Methods

Three male adult rhesus monkeys (Macaca mulatta) were the subjects in the neurophysiological experiments. We followed experimental protocols approved by the Institutional Animal Care and Use Committee of the University of Wisconsin–Madison and conformed to U.S. Department of Agriculture regulations and to the National Institutes of Health guidelines for the care and use of laboratory animals. Procedures for surgical preparation and electrophysiological recordings were akin to those described previously (Xiao et al. 2015). We implanted a head post and a recording cylinder during sterile surgery with the animal under isoflurane anesthesia. For electrophysiological recordings from neurons in area MT, we took a vertical approach and used tungsten electrodes (1–3 M; FHC). We identified area MT by its characteristically large portion of direction selective neurons, small RFs relative to those of neighboring medial superior temporal cortex (area MST), its location at the posterior bank of the superior temporal sulcus, and visual topography of the RFs (Gattass & Gross, 1981). We amplified electrical signals and identified single units with a real-time template-matching system and an offline spike sorter (Plexon). We monitored the animal's eye position using a video-based eye tracker (EyeLink, SR Research) at a rate of 1000 Hz. For some of recording sessions from monkey B, we tracked both the left and right eye positions which were then used to calculate the horizontal vergence values (left eye X position-right eye X position).

Experimental procedures

We used a real-time data acquisition program, "Maestro" to control stimulus presentation and data acquisition. For all experiments, we presented visual stimuli on a 25-inch CRT monitor at a viewing distance of 63 cm. Monitor resolution was 1024 x 768 pixels, with a refresh rate of 100 Hz. We generated the visual stimuli on a Linux workstation using the OpenGL application that communicated with an experimental control computer. We did the luminance calibration of the CRT monitor with a photometer (LS-110, Minolta) and then used a standard display gamma value of 2.2 for gamma correction.

After we isolated the spike waveforms of a single neuron, we first characterized the neuron's direction selectivity by interleaving trials of a 30° x 27° random-dot patch (RDP), covering the entire monitor, moving in different directions at a step of 45° and at a speed of 10°/s. The direction selectivity and preferred direction (PD) were determined using MATLAB (MathWorks). We then characterized the speed tuning of the neuron using a random dot patch moving at different speeds (1, 2, 4, 8, 16, 32, or 64°/s) in the PD of the neuron. Using a cubic spline, we took the preferred speed (PS) of the neuron as the speed that evoked the highest firing rate in the fitted speed-tuning curve. Next, we mapped the receptive field (RF) location of the neuron by recording responses to a series of 5° × 5° patches of random dots that moved in the PD and at the PS of the neuron whose location we varied randomly to tile the screen (35×25°) in 5° steps without overlap. We then interpolated the raw map of these responses at an interval of 0.5° to obtain the location with the highest firing rate response as the center of the RF.

We used green and red anaglyph filters (Kodak Wratten filters, nos. 25 and 58) to create stereoscopic depth on the monitor using horizontal disparity. To achieve a disparity of + n° or - n°, we used two identical dot patterns, one in green and the other in red with the center of the red dot pattern offset horizontally by +n° or -n° relative to that of the green dot pattern, respectively. As the animals viewed the stimuli through a green filter on the left eye and red filter on the right eye, only one set of dots was visible to each eye creating a binocular disparity that resulted in depth perception. Crosstalk between the two eyes, as viewed through the filters, was 14.5%, 9.6%, and 9.6% for monkey G, B, and R, respectively. We determined the selectivity of a MT neuron to binocular disparity by measuring its responses to a series of RDPs moving in its preferred direction and at its preferred speed (often within 5 to 20°/s) at fifteen different disparities from -1.5° to +1.5°.

Visual stimuli

Transparent motion stimuli constituted two overlapping circular random dot patches (RDPs) of size 15° moving at two different depths: one at -0.1° binocular disparity (near) and the other at 0.1° binocular disparity (far) relative to the fixation point on the monitor (Fig 2 A to D). We refer to the constituent RDPs as 'stimulus components': the one at near disparity as 'near component' and the one at far disparity as 'far component'. Each RDP's dot density was 0.89 dots/deg^2 with infinite dot lifetime and had 100% coherence. The visual stimuli were usually centered on the fixation spot and sometimes shifted toward the RFs for 1 to 2°. As a result, the visual stimuli covered the RFs but usually were not centered on the RFs. The speed of the motion stimuli was 10°/s in most sessions and 12°/s in a few sessions.

Behavioral tasks

In the discrimination task with attention cue (Fig 1 A and B), we trained Monkey B and G to report the motion direction of the RDP (stimulus component) at cued disparity (+ 0.1° or - 0.1°). Each trial started with a fixation dot in the center of the screen and stationary reference dots (not shown in the figure). The animal had to maintain fixation within 1°x 1° window around the dot to continue the trial. As a cue for attention, a static circular RDP with a diameter of 15° appeared for 500 ms at either near plane (-0.1° disparity) or far plane (0.1° disparity) referred to as "Cue Near/Attend Near" and "Cue Far/Attend Far" conditions, respectively. To aid depth perception, stationary reference dots at zero disparity covered the rest of the monitor. This was followed by 300 ms of blank screen with a fixation dot. Then, a single component RDP at the cued disparity (+0.1°: Single Near or -0.1°: Single Far) or two overlapping RDPs at +0.1° and -0.1° disparities appeared stationary for 500ms. And they started moving for 600ms in one of the 12 directions: 0 to 360° at 30° step. After the motion period, the stimuli were off, and 12 reporting targets arranged from 0 to 360° at 30° step along a circle of 10° diameter, appeared for 450 ms for monkey B and 600 ms for monkey G. During this time window, the animals had to report the motion direction of the cued component (Near or Far) by making a saccade to the corresponding target and holding fixation within 1.5°x 1.5° target window for 350ms to receive a juice reward. We used two angle separations between the overlapping motion directions (60°/DS60 or 120°/DS120) to reduce the chances of the animals predicting the direction of the component at cued disparity just by attending to the other/un-cued component. For each angle separation, we used two configurations to control the effects of clockwise (c) or counter-clockwise arrangement of overlapping stimuli on MT neuronal responses (Xiao et al 2015): 1. Near component moving clockwise and far component

moving counter-clockwise (Near C-Far CC), and 2. Near component moving counter-clockwise and far component moving clockwise (Near CC-Far C).

In the attention-away task, the stimuli were centered on MT neurons RFs. To direct the animal's attention away from the RF, we trained Monkey R to perform a two alternative forced choice (2-AFC) task of discriminating whether the motion direction of an RDP, presented in the opposite visual hemifield, was clockwise (C) or counter-clockwise (CC) to vertically upward direction. Every trial began with the appearance of a fixation dot at the center of the screen and the reference stationary dots at zero disparity. The animal fixated within 1°x 1° window of the dot for 200 ms. Then the stimuli in the RFs ("Single" or "Overlapping" disparity motion stimuli) appeared along with an "attention stimulus": a random-dot patch with a diameter of 5°, placed at 10° eccentricity in the opposite hemifield. Both stimuli remained stationary for 250 ms and then the "attention stimulus" moved either clockwise or counterclockwise from upward direction for 500 ms by 10°, 15°, or 20° offset in randomly interleaved trials. Then, all the stimuli were turned off, and two reporting targets appeared on the left and right side of the FP. The animal made a saccade to the left (or the right) side to indicate if the motion of the "attention stimulus" was counterclockwise (or clockwise) of upward direction and maintained fixation on the correct target for 200ms to receive a juice reward.

Experimental design and statistical analysis

Neuronal response (firing rate) was calculated during the stimulus motion period (600 ms for the discrimination task with attention and 500 ms for attention away task) averaged across repeated trials. We constructed the response tuning curves to unidirectional stimuli (Single component) and to bidirectional stimuli (Overlapping) and fitted the raw direction tuning curves using cubic splines at a resolution of 1°. For each vector averaged (VA) direction of the overlapping motion directions, we determined the responses elicited by the bidirectional stimuli and the constituent unidirectional stimulus components. To average the direction tuning curves across neurons, we rotated the spline-fitted tuning curve elicited by the bidirectional stimuli such that the VA direction of 0° was aligned with the PD of each neuron. We then normalized neuronal responses by each neuron's maximum of response to near and far components and averaged the aligned, normalized tuning curves across neurons. We pooled the responses of Near C-Far CC and Near CC-Far C conditions for population averaged tuning curves in Fig 3. The resulting tuning curves always had the near component at C side and the far component at CC side for visualization. To do so, we first horizontally flipped the tuning curves to the bidirectional and unidirectional stimuli of neurons of Near CC-Far C conditions, along the axis of VA direction. We then averaged the flipped tuning curves together with the tuning curves of Near C-Far CC condition.

To examine the time-course of the response tuning to the bidirectional stimuli, we calculated the response tuning for each individual neuron using the trial-averaged firing rates within a 50ms time window, sliding at a step of 10-ms, and normalized the responses by the maximum firing rate of the near and far components response across all time windows. We then pooled the responses of Near C-Far CC and Near CC-Far C conditions in the same way as we did for the populations averaged tuning curves mentioned above.

LWS model

To quantify the relationship between the responses elicited by the bidirectional stimuli and those elicited by the individual stimulus components, we fitted the direction tuning curves using a linear weighted summation (LWS) model (Xiao et al. 2014) by minimizing the sum of squared error:

$$R_{nfpred}(\theta_n, \theta_f) = w_n R_n(\theta_n) + w_f R_f(\theta_f) + c \qquad (13)$$

R_{nfpred} is the response to the bidirectional stimuli predicted by the model; R_n and R_f are the measured responses elicited by the near and far components, respectively, when presented alone; θ_n and θ_f are the motion directions of the near and far components; w_n and w_f are the response weights of the near and far components, respectively; and c is a positive number. b in the SNL model is the "nonlinear interaction coefficient" that determines the sign and strength of the multiplicative interaction between the component responses.

Divisive normalization model (DNM)

We fitted each neuron's response tuning curves to the bidirectional stimuli using a modified divisive normalization model (see Results). Equations 17 and 18 were used to fit each of the four bidirectional tuning curves of a neuron (described in Fig 2: DS60 Near C-Far CC, DS60 Near CC-Far C, DS120 Near C-Far CC, and DS120 Near CC-Far C) under "Attend Near" and "Attend Far" conditions, respectively. Hence, each neuron contributed to the fits of eight bidirectional tuning curves presented in the results.

When fitting the DNM model (Eq. 17 and 18) to each neuron's bi-directional response tuning, we divided the search space for parameter v (with bounds from v =0.01 to 40) to result in equal steps (s) of the area covered by the gaussian weighting function $\boldsymbol{G}$ (See Eq. 20). This is because the area covered by the gaussian weighting function $\boldsymbol{G}$ better represents the size of the weighting pool than the standard deviation (v) (see Supplementary Fig 2). For each v value in the search space (v =0.01 to 40 at a step of s), we used the constrained minimization tool "fmincon" (MATLAB) to minimize sum of squared error (SSE) of Eq 17 or 18, varying all other free parameters of the model. To identify the v values resulting in the least SSE, we plotted the resulting SSE as a function of the values of v in the search space, referred to as SSE(v). Based on SSE(v), we calculated a probability distribution of the v values (pdf(v)) using the steps below, accounting for multiple v values resulting in similar fits (i.e., SSE values). First, we normalized the SSE(v) function between 0 and 1 to get $\mathrm{SSE}_{\mathrm{norm}}(v)$. Using a threshold of 0.05, for v vlaues with $\mathrm{SSE}_{\mathrm{norm}}(v)$<=0.05, we set $f(v)$=1 and for all other v values, $f(v)$=0. We then normalized $f(v)$with sum of $f(v)$ across all v values to get the probability of best fit at each v value i.e., $\mathrm{pdf}(v) = \frac{f(v)}{\Sigma f(v)}$. The pdf($v$) functions obtained for each tuning curve of each neuron are averaged across near and far neurons to get $\mathrm{pdf}_{\mathrm{near}}(v)$ and $\mathrm{pdf}_{\mathrm{far}}(v)$ for each monkey plotted against area under $\boldsymbol{G}$ (Fig 13B and 13C). To evaluate the statistical significance of the difference of pdfs between the monkeys, we used bootstrapping where we randomly sampled the pdf values (of a neuronal set) 1000 times with replacement and calculated the 95% confidence intervals for each v value.

<u>Attentional weighting parameters</u>*:* We derived the values of $\beta_{An}$and β_{Af} used in Eq. 17 and 18 from the weights of the LWS model fitting by rewriting Eq 1 as below:

$$R_{(An,f)pred} = w_{nAn}R_n + w_{fAn}R_f + c \qquad (14a)$$

$$R_{(n,Af)pred} = w_{nAf}R_n + w_{fAf}R_f + c \qquad (14b)$$

w_{nAn}, w_{fAn} and w_{nAf}, w_{fAf} are the weights of near and far components in the two attention conditions, respectively. c >0. We assume that attention to a component of bidirectional stimuli only scales the weight of the attended component (scaled by β_{An} and β_{Af} factors), but not the weight of the other/un-attended component (scaled by a factor of 1).

$$w_{nAn} = w_n * \beta_{An}$$

$$w_{fAn} = w_f * 1$$

$$w_{nAf} = w_n * 1$$

$$w_{fAf} = w_f * \beta_{Af}$$

$$\beta_{An} = \frac{w_{nAn}}{w_{nAf}} \;\; ; \; \beta_{Af} = \frac{w_{fAf}}{w_{fAn}} \qquad (15a; 15b)$$

To evaluate the goodness of fit of a model for the response tuning curve to the bidirectional stimuli, we calculated the percentage of variance (PV) the model accounts for as follows:

$$PV = 100 * (1 - SSE)/SST \qquad (16)$$

where SSE is the sum of squared errors between the model fit and the neuronal data, and SST is the sum of squared differences between the data and the mean of the data.

References

Bao P, Tsao DY. 2018. Representation of multiple objects in macaque category-selective areas. *Nat Commun* 9: 1774

Ban H, Preston TJ, Meeson A, Welchman AE. 2012. The integration of motion and disparity cues to depth in dorsal visual cortex. *Nat Neurosci* 15: 636-43

Bradley DC, Qian N, Andersen RA. 1995. Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature* 373: 609-11

Bradshaw MF, Rogers BJ. 1996. The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Res* 36: 3457-68

Britten KH, Heuer HW. 1999. Spatial summation in the receptive fields of MT neurons. *J Neurosci* 19: 5074-84

Busse L, Wade AR, Carandini M. 2009. Representation of concurrent stimuli by population activity in visual cortex. *Neuron* 64: 931-42

Carandini M, Heeger DJ. 2011. Normalization as a canonical neural computation. *Nat Rev Neurosci* 13: 51-62

Cumming BG, Parker AJ. 1999. Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. J Neurosci 19, 5602–18.

DeAngelis GC, Newsome WT. 1999. Organization of disparity-selective neurons in macaque area MT. *J Neurosci* 19: 1398-415

DeAngelis GC, Uka T. 2003. Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J Neurophysiol* 89: 1094-111

Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18: 193-222

Ferrera VP, Lisberger SG. 1997. The effect of a moving distractor on the initiation of smooth-pursuit eye movements. *Vis Neurosci* 14: 323-38

Gattass R, Gross CG. 1981. Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque. *J Neurophysiol* 46: 621-38

Goutcher R. 2016. Motion direction influences surface segmentation in stereo transparency. *J Vis* 16: 17

Greenwood JA, Edwards M. 2006b. Pushing the limits of transparent-motion detection with binocular disparity. *Vision Res* 46: 2615-24

Heuer HW, Britten KH. 2002. Contrast dependence of response normalization in area MT of the rhesus macaque. *J Neurophysiol* 88: 3398-408

Hibbard PB, Bradshaw MF. 1999. Does binocular disparity facilitate the detection of transparent motion? *Perception* 28: 183-91

Huang X, Wang C, Arseneau B, Yerxa T, Cooper E.A. 2019. Natural scene statistics of depth and motion pertinent to figure-ground segregation. Society for Neuroscience annual meeting.

Huang X, Ghimire B, Chakrala AS, Wiesner S. 2023. Neural coding of multiple motion speeds in visual cortical area MT. bioRxiv. doi: https://doi.org/10.1101/2023.04.08.532456.

Kohler PJ, Cottereau BR, Norcia AM. 2019. Image Segmentation Based on Relative Motion and Relative Disparity Cues in Topographically Organized Areas of Human Visual Cortex. *Sci Rep* 9: 9308

Krekelberg B, van Wezel RJ. 2013. Neural mechanisms of speed perception: transparent motion. *J Neurophysiol* 110: 2007-18

Lee J, Maunsell JH. 2010. The effect of attention on neuronal responses to high and low contrast stimuli. *J Neurophysiol* 104: 960-71
Li X, Basso MA. 2005. Competitive stimulus interactions within single response fields of superior colliculus neurons. *J Neurosci* 25: 11357-73
Maunsell JH, Van Essen DC. 1983b. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol* 49: 1148-67
McDonald JS, Clifford CW, Solomon SS, Chen SC, Solomon SG. 2014. Integration and segregation of multiple motion signals by neurons in area MT of primate. *J Neurophysiol* 111: 369-78
Nawrot M, Blake R. 1989. Neural integration of information specifying structure from stereopsis and motion. *Science* 244: 716-8
Nguyenkim JD, DeAngelis GC. 2003. Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J Neurosci* 23: 7117-28
Ni AM, Maunsell JHR. 2019. Neuronal Effects of Spatial and Feature Attention Differ Due to Normalization. *J Neurosci* 39: 5493-505
Ni AM, Ray S, Maunsell JH. 2012. Tuned normalization explains the size of attention modulations. *Neuron* 73: 803-13
Prince SJ, Cumming BG, Parker AJ. 2002. Range and mechanism of encoding of horizontal disparity in macaque V1. *J Neurophysiol* 87: 209-21
Qian N, Andersen RA. 1994. Transparent motion perception as detection of unbalanced motion signals. II. Physiology. *J Neurosci* 14: 7367-80
Recanzone GH, Wurtz RH, Schwarz U. 1997. Responses of MT and MST neurons to one and two moving objects in the receptive field. *J Neurophysiol* 78: 2904-15
Reynolds JH, Heeger DJ. 2009. The normalization model of attention. *Neuron* 61: 168-85
Smolyanskaya A, Ruff DA, Born RT. 2013. Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable. *J Neurophysiol* 110: 2806-16
Snowden RJ, Rossiter MC. 1999. Stereoscopic depth cues can segment motion information. *Perception* 28: 193-201
Snowden RJ, Treue S, Erickson RG, Andersen RA. 1991. The response of area MT and V1 neurons to transparent motion. *J Neurosci* 11: 2768-85
Treue S, Hol K, Rauber HJ. 2000. Seeing multiple directions of motion-physiology and psychophysics. *Nat Neurosci* 3: 270-6
Treue S, Martinez Trujillo JC. 1999. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399: 575-9
Treue S, Maunsell JH. 1996. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382: 539-41
Wallace JM, Mamassian P. 2003. The efficiency of speed discrimination for coherent and transparent motion. *Vision Res* 43: 2795-810
Wallace JM, Mamassian P. 2004. The efficiency of depth discrimination for non-transparent and transparent stereoscopic surfaces. *Vision Res* 44: 2253-67
Wannig A, Rodriguez V, Freiwald WA. 2007. Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron* 54: 639-51
Wiesner S, Baumgart IW, Huang X. 2020. Spatial Arrangement Drastically Changes the Neural Representation of Multiple Visual Stimuli That Compete in More Than One Feature Domain. *J Neurosci* 40: 1834-48

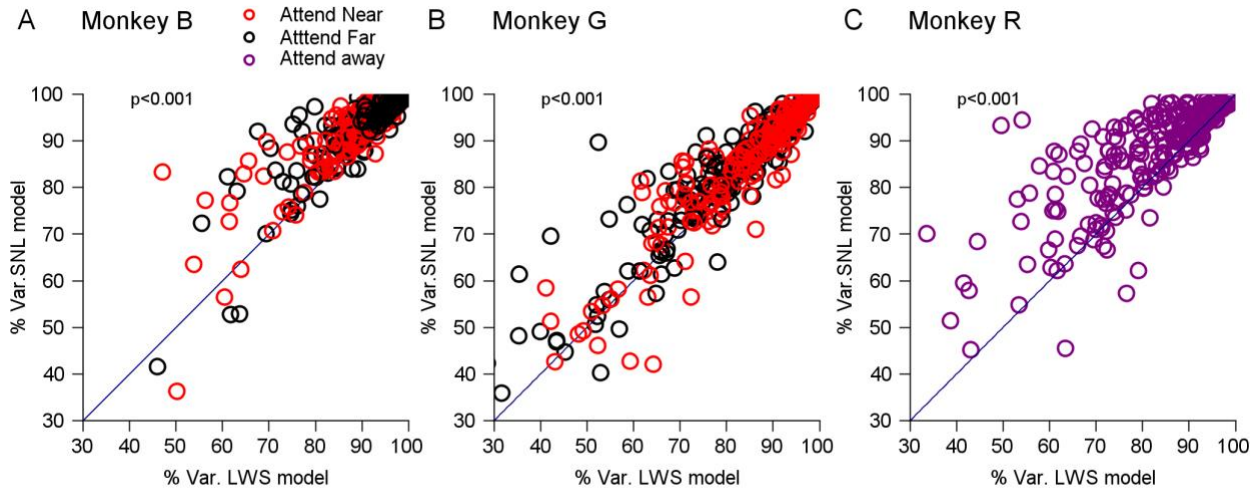
Xiao J, Huang X. 2015. Distributed and Dynamic Neural Encoding of Multiple Motion Directions of Transparently Moving Stimuli in Cortical Area MT. *J Neurosci* 35: 16180-98

Xiao J, Niu YQ, Wiesner S, Huang X. 2014. Normalization of neuronal responses in cortical area MT across signal strengths and motion directions. *J Neurophysiol* 112: 1291-306

Zoccolan D, Cox DD, DiCarlo JJ. 2005. Multiple object response normalization in monkey inferotemporal cortex. *J Neurosci* 25: 8150-64

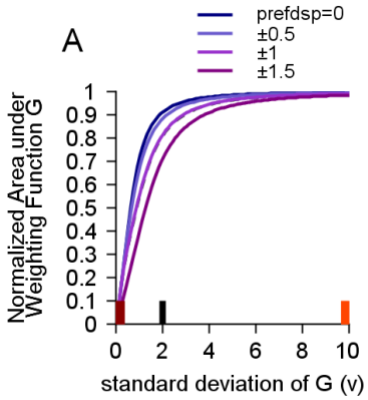
The role of binocular disparity in the neural representation of multiple moving stimuli in the visual cortex

Supplementary material



Supplementary Figure 1. Goodness of fits LWS model vs SNL model

A. Comparison of % Var. (percent variance *PV* Eq. 16) of SNL and LWS model fits of monkey B. Red and black dots indicate the fits of bidirectional tuning curves from "Attend Near" and "Attend Far" conditions. **B.** Monkey G. **C.** Monkey R from "Attend Away" paradigm. p values are from one-sided sign rank test. Blue lines indicate x=y line.



Supplementary Figure 2. Relationship between area under Weighting function ***G*** and standard deviation v

A. Area under Weighting function ***G*** vs standard deviation v in the divisive normalization model (Eq. 6, 7, 9). The different colors represent the different preferred disparity (prefdsp) values of neurons. The brown, black, and orange indicate the v values of Local, Intermediate, and Global pooling strategies shown in Fig 13.