Mechanisms of Neural Representation and Segregation of Multiple Spatially Separated Visual Stimuli

Steven Wiesner, Bikalpa Ghimire, Xin Huang

Department of Neuroscience, University of Wisconsin-Madison, Wisconsin 53705, USA

Correspondence should be addressed to:

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

Acknowledgment

We thank Bryce Arseneau for animal training, 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, and McPherson Eye Research Institute at the University of Wisconsin – Madison.

Conflict of Interest

None.

Abstract

Segregating objects from one another and the background is essential for scene understanding, object recognition, and visually guided action. In natural scenes, it is common to encounter spatially separated stimuli, such as distinct figure-ground regions, adjacent objects, and partial occlusions. Neurons in mid- and high-level visual cortex have large receptive fields (RFs) that often encompass multiple, spatially separated stimuli. It is unclear how neurons represent and segregate these stimuli within their RFs. We investigated this question by recording the neuronal responses in the middle temporal (MT) cortex from two male macaque monkeys to multiple moving stimuli. We placed a motion border between two spatially separated random-dot patches within the RFs that moved in two different directions. We varied the vector average direction of the stimuli to characterize the full direction tuning curves. Across motion directions, responses to multiple stimuli were systematically biased toward the stimulus located at the neuron's more-preferred RF subregion. The sign and magnitude of this spatial-location bias were correlated with the neuron's spatial preference measured with single patches presented in isolation. We demonstrated that neuronal responses to multiple stimuli can be captured by an extended divisive normalization model, as a sum of the responses elicited by individual stimuli, weighted by the neuron's spatial preference. We also proposed a circuit implementation for the extended normalization model. Our results indicate that MT leverages spatial selectivity within the RFs to represent spatially separated moving stimuli. The spatial-location bias in neuronal responses enables individual components of multiple stimuli to be represented by a population of neurons with heterogeneous spatial preferences, providing a neural substrate for segregating multiple visual stimuli.

Introduction

In natural vision, it is rare to encounter an isolated object presented on a blank background. Instead, natural scenes are often complex and contain multiple entities. The ability to segment objects from one another and from their background is central to scene interpretation, object recognition, and visually guided action. Visual motions provide potent cues for segmentation (Wertheimer, 1938; Braddick, 1993; Britten, 1999; Stoner and Albright, 1993; Wagemans et al., 2012). Visual entities that move differently are likely to be segmented. However, the neural mechanisms by which the visual system utilizes motion cues to achieve segmentation remain to be understood.

The extrastriate area MT in primates serves as a key hub for processing motion information and plays a crucial role in motion-based segmentation (Britten, 2003; Born and Bradley, 2005; Pasternak and Tadin, 2020). Previous neurophysiological studies have investigated how MT neurons represent and segregate overlapping moving stimuli in their receptive fields (RFs) that give rise to the perception of transparent motion (Snowden et al., 1991; Stoner and Albright, 1992; Qian and Andersen, 1994; Treue et al., 2000; Xiao et al. 2014; McDonald et al., 2014; Xiao and Huang, 2015; Wiesner et al., 2020; Chakrala et al., 2024; Huang et al., 2025). Our laboratory has found that the responses of MT neurons to overlapping moving stimuli often preferentially represent one stimulus component with a higher luminance contrast or motion coherence (Xiao et al., 2014), or one of two competing motion directions (Xiao and Huang, 2015), speeds (Huang et al., 2025), or binocular disparities (Chakrala et al., 2024). These studies suggest a common pattern: MT neurons tend to employ a strategy of biased mixing of stimulus components, rather than equally pooling them. This strategy enables neurons to better represent individual stimulus components, thereby facilitating segmentation.

Using transparent moving stimuli has the advantage of isolating motion cues from spatial cues, as stimulus components are superimposed and occupy the same spatial location. Yet, in natural environments, it is more common to encounter spatially separated stimuli, such as figure and ground regions, two objects abutting each other, and situations where one object partially occludes another. Previous neurophysiological studies on the segmentation of spatially separated stimuli have mainly focused on the interaction between the RF center and surround (Allman et al., 1985; Born, 2000; Huang et al., 2007, 2008). Britten and Heuer (1999) investigated the spatial summation of multiple stimuli moving in the same direction within the RFs of MT neurons. However, relatively little is known about the neural representation and segregation of multiple stimuli moving differently within the RFs of MT neurons.

The motion discontinuity at the border between two spatially separated stimuli provides key information for segmentation (Nakayama and Loomis, 1974; Regan and Beverley, 1984; Grossberg, 1991, 1994; Black and Fleet, 2000; Feldman and Weinshall, 2008; Yoonessi and Baker, 2011). In this study, we focused on investigating how MT neurons represent two directions at the motion border of spatially separated stimuli within the RFs. The spatial RF of an MT neuron can be roughly described with a Gaussian function (Raiguel et al., 1995; Kumano and Uka, 2010), and some MT neurons also exhibit a substructure with multiple distinct hotspots of sensitivity within the RFs (Livingstone et al., 2001; Richert et al., 2013). When two spatially separated stimuli are placed in the RF, according to the neuron's spatial RF profile, the neuron may respond to one stimulus more strongly than the other when each stimulus is presented alone, exhibiting a "spatial preference". We hypothesize that MT neurons leverage their spatial preferences for individual

stimuli – when multiple, spatially separated stimuli move simultaneously in different directions, MT responses are biased toward the stimulus at the neuron's more preferred location. Such spatial bias would enable different neurons to better represent the stimulus components at their preferred locations.

To test this hypothesis, we placed a motion border between two spatially separated random-dot patches within the RFs. The two stimuli moved in two directions separated by an angle of either 60° or 30°, which was significantly smaller than the mean tuning width (95º-102º) of MT neurons in response to single directions (Albright, 1984; Huang and Lisberger, 2009). At these small angular separations (AS), equal pooling of the responses elicited by individual stimulus components creates a challenging situation for segregating motion directions (Treue et al., 2000; Xiao and Huang, 2015) and therefore provides an opportunity to reveal the role of spatial cues in motion segmentation.

We found that MT neurons indeed exhibited spatial bias in response to spatially separated moving stimuli, consistent with our hypothesis. We further demonstrated that the spatial bias can be described by an extended normalization model, resulting from uneven weighting of the responses elicited by individual stimulus components according to the neuron's spatial preference for single stimuli. We also suggested a circuit implementation for the spatial preference-weighted normalization model. This study revealed that the visual system exploits spatial cues in multiple stimuli and leverages the spatial selectivity of MT neurons to weigh the stimulus components differentially, thereby facilitating visual segmentation.

Results

We investigated how neurons in area MT encode spatially separated stimuli moving simultaneously in different directions within the neurons' RFs. Specifically, we asked how MT neurons represent and segregate multiple visual stimuli moving in different directions, as well as the roles of spatial cues in this representation. To address these questions, we recorded the electrical signals from isolated neurons in MT from two male rhesus macaques (R and Z) while they performed a fixation task. Our dataset includes recordings from 117 MT neurons (81 from R and 36 from Z).

Our main visual stimuli consisted of two spatially separated random-dot patches, positioned adjacent to each other along a vertical border and moving in two directions separated by either 60° (Fig. 1A, B) or 30° (Fig. 3). The border was roughly centered on the RFs of the recorded MT neurons. Nevertheless, it was common for the neuronal responses to the left and right patches presented alone to be different, indicating a spatial preference for either the left or right patch of the neurons. We varied the vector average (VA) direction of the bidirectional stimuli across 360º to characterize the direction tuning curves. We also measured the direction tuning curves to the left and right patches presented alone.

We have previously found that, in response to overlapping random-dot stimuli, some MT neurons have a preference for the motion direction either at the clockwise (C) or counterclockwise (CC) side of two motion directions, referred to as the "direction side-bias" (Xiao and Huang, 2015). To control for the direction side-bias and examine the effects of the spatial cues, our main visual stimuli included conditions in which the motion direction of the right patch was either at the C side

(Fig. 1A) or CC side (Fig. 1B) of the motion direction of the left patch. This design allowed us to separate the effects of spatial cues from directional biases.

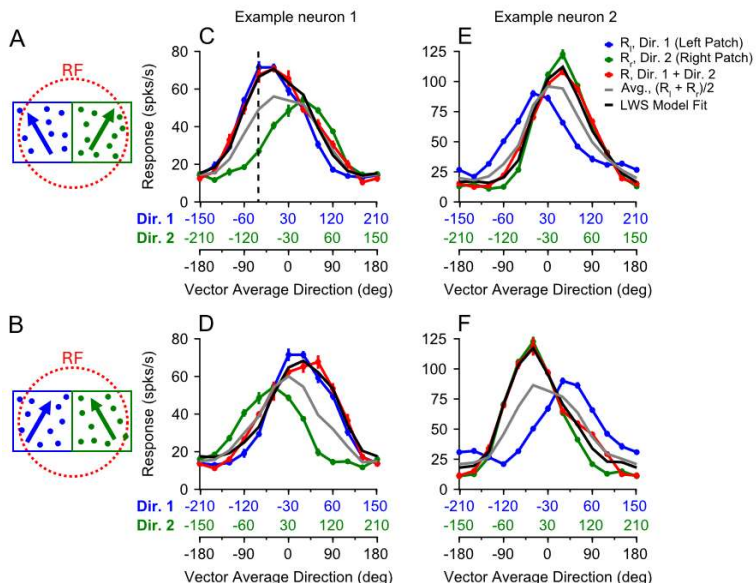

Figure 1. Visual stimuli and response tuning curves of two example neurons. A, B. Cartoons depict spatially separated stimuli placed side by side with a vertical motion boundary and different arrangements of motion directions. Dots were achromatic. The blue and green arrows correspond to the left and right components, respectively. **C.** Response tuning curves from example neuron one. The blue and green abscissas represent the unidirectional components for Dir. 1 and Dir. 2, respectively. The black abscissa corresponds to the VA direction of the bidirectional stimuli. This neuron shows a strong spatial bias to the stimulus on the left side of the RF, with the red curve closely following the blue curve. The black curve is the fit of the linear weighted summation (LWS) model to the bidirectional response ($w_l = 0.98, w_r = 0.27$). **D**: Same neuron with the component directions flipped ($w_l = 0.96, w_r = 0.38$). **E**: Response tuning curves from example neuron two showing a strong spatial bias to the component on the right side of the RF ($w_l = 0.26, w_r = 0.98$). **F**: Same example neuron as in E, but with component directions flipped ($w_l = 0.23, w_r = 0.94$). Error bars indicate the standard error (SE).

Spatial bias in the response tuning to bidirectional stimuli

Figure 1 shows the direction tuning curves of two representative MT neurons in response to two stimulus patches with an angular separation (AS) of 60º. The blue curve indicates each neuron's responses to the left patch presented alone and moving in single directions, whereas the green curve indicates the responses to the right patch alone. The red curve indicates the responses to both patches moving simultaneously. The three tuning curves (blue, green, and red) are plotted in a way that the VA direction of the bidirectional stimuli and the corresponding component directions of the individual patches are aligned vertically (see the vertical line in Fig. 1C). The responses to the bidirectional stimuli of the first example neuron showed a strong bias toward the motion directions of the left patch (Fig. 1C, D), significantly different from the average of the responses to the two component directions (the gray curves). The response bias occurred regardless of whether the left patch moved in the CC side (Fig. 1A, C) or the C side (Fig. 1B, D) of the two component directions, indicating that the bias cannot be explained by a direction side-bias. Note

that this neuron's responses to the individual patches presented alone showed a preference for the left patch – the blue curve had a higher peak amplitude than the green curve. The response bias toward the left patch of the bidirectional stimuli is consistent with the neuron's spatial preference for the individual patch. Another example neuron (Fig. 1E, F) had a spatial preference for the right patch in the RF when the individual patches were presented alone. In response to the bidirectional stimuli, the response of the neuron was strongly biased toward the right patch, regardless of whether the right-side component was moving in the direction at the C or CC side of the two motion directions (Fig. 1E, F).

The responses to the bidirectional stimuli can be well described by a weighted sum of the responses to the individual components, referred to as the linear weighted summation (LWS) model:

$$R_{pred}(\theta_l, \theta_r) = w_l R_l(\theta_l) + w_r R_r(\theta_r) + c \quad (1)$$

The model has three free parameters. w_l and w_r are the response weights for the left and right components, respectively; c is an offset term. θ_l and θ_r are the component directions of the left and right patches, and R_l and R_r are the measured responses to the direction components, respectively. The model-predicted response to the bidirectional stimuli R_{pred} captured, on average, 90.6% of the response variance across the neurons (std = 9.43%, N = 89). The black curves in Figure 1 illustrate the LWS model fits of the two example neurons.

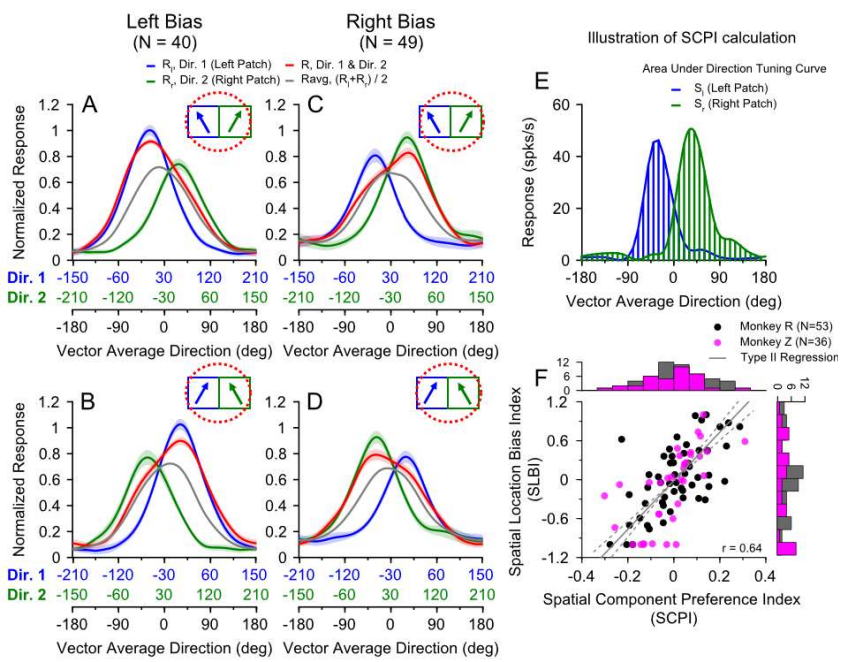

Figure 2. Spatial bias in responses to the bidirectional stimuli with an AS of 60°. **A-D.** Population-averaged response tuning curves. The average of the responses elicited by individual stimulus components is shown by the gray curve. The abscissa convention and arrangement of the direction tuning curves are the same as in Figure 1. After spline fitting, the preferred direction (PD) of each neuron is aligned with the VA direction of 0° and averaged across neurons. **A, B.** Averaged responses from neurons that show a spatial bias towards the left side of the RF (N = 40) (i.e., *SLBI* < 0) when the left patch moved at the counter-clockwise (CC) side (**A**), or clockwise (C) side (**B**) of the two motion directions. **C, D.** Averaged responses

of neurons showing a spatial bias towards the right side of the RF (N=49) (i.e., *SLBI* > 0). Error bands indicate the SE. **E:** Illustration of the calculation of the spatial component preference index (*SCPI*). The blue shade depicts the area under the tuning curve (AUC) for the left stimulus component (S_l), while the green curve for the right stimulus component (S_r) of an example neuron. **F:** Scatter plot comparing *SLBI* as a function of *SCPI* across neurons and for each animal. Response weights for each neuron were calculated based on the LWS model. The solid gray line indicates the best fit from Type II regression, while the dotted lines indicate the lower and upper bounds of the 95% confidence interval.

To isolate the effects of the spatial cues from the direction side-bias, we devised a "spatial location bias index": $SLBI = \frac{(w_{r_c} - w_{l_{cc}}) + (w_{r_{cc}} - w_{l_c})}{(w_{r_c} + w_{l_{cc}}) + (w_{r_{cc}} + w_{l_c})}$, in which w is the response weight for a component patch. The subscripts l and r represent left and right stimulus patches, and C and CC represent motion at the C and CC sides of the two stimulus directions. We always compared the weights for the left and right stimuli, w_l and w_r, while balancing the motion directions. A positive *SLBI* indicates a response bias toward the right component patch, whereas a negative *SLBI* indicates a left-patch bias.

Among 89 neurons tested with an AS of 60º, 49 neurons showed a right bias (mean ± std: w_r = 0.70 ± 0.15, w_l = 0.34 ± 0.19), and 40 neurons showed a left bias (mean ± std: w_r = 0.25 ± 0.20, w_l = 0.77 ± 0.19). The population-averaged tuning curves of the right- and left-bias neurons are shown in Figure 2A-D. Although the average of the component responses (the gray curve) was slightly shifted toward the stronger component, the signals of the two component directions were still mixed. In contrast, the response to the bidirectional stimuli (the red curve) was substantially biased toward the stronger component, with the response peaking at the VA direction where one component direction was aligned with the neuron's preferred direction. Such a representation of the bidirectional stimuli would allow a better segregation of the component directions.

To characterize the neuron's spatial preference for single patches, we used a "spatial component preference index": $SCPI = \frac{(S_r - S_l)}{(S_r + S_l)}$, in which S_r and S_l are the integrals of the direction tuning curves to the right and left patches, respectively (illustrated in Fig. 2E). The integral of the tuning curve to single directions of a stimulus component is equivalent to the sum of the responses elicited by a single motion direction of a stimulus component, from a population of neurons that have identical shapes of the direction tuning curves, but with different preferred directions. We found that the bias in the neuronal response to the bidirectional stimuli ($SLBI$) is positively correlated with the neurons' spatial preference for single patches ($SCPI$) (Fig. 2F, Type II regression, r = 0.64). While the $SCPI$ values were mostly within the range from -0.2 to 0.2, the $SLBI$ values had a wider spread from -1.0 to 1.0 (variances are significantly different, Levene's test, p = 2.05×10^{-15}), suggesting that a small difference in the component responses can lead to a large spatial bias in the responses to the bidirectional stimuli.

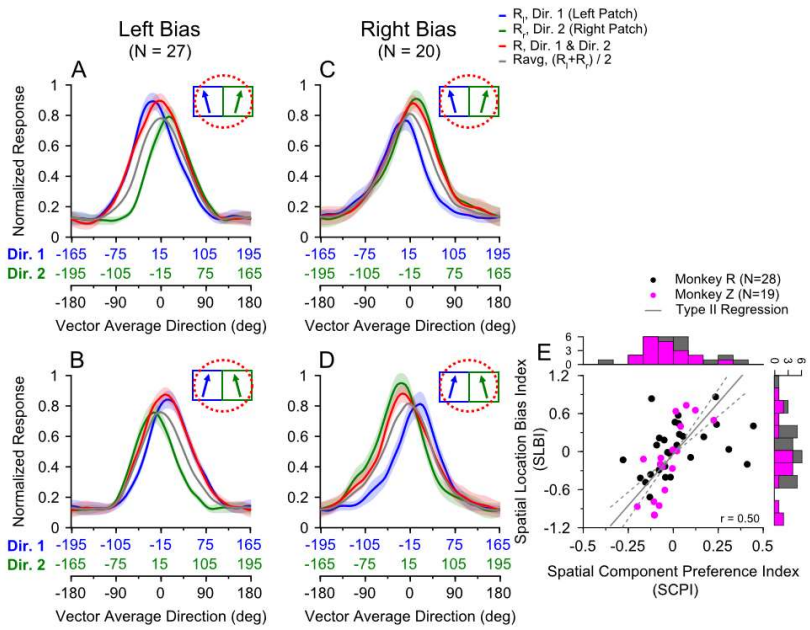

Figure 3. **Spatial bias in responses to the bidirectional stimuli with an AS of 30°**. **A-D.** Population-averaged response tuning curves. **A, B.** Tuning curves from a population of neurons showing spatial bias towards the left stimulus component (*SLBI* < 0). **C, D.** Neurons showing spatial bias towards the right stimulus component (*SLBI* > 0). **A, C**. The left patch moved in the CC-side direction. **B, D**. The left patch moved in the C-side direction. Error bands indicate the SE. **E.** Scatter plot comparing *SLBI* as a function of *SCPI* across neurons. The solid gray line shows the Type II linear regression fit, while the dotted lines indicate the lower and upper bounds of the 95% confidence interval.

As the AS between the motion components decreases, segregation of the multiple moving stimuli becomes more difficult. We asked whether spatial bias also occurred at smaller AS to facilitate segregation. We recorded from 47 MT neurons using an AS of 30º. Twenty-seven neurons had a negative *SLBI,* and the population-averaged response to the bidirectional stimuli showed a strong bias toward the left patch (Fig. 3A, B). Twenty neurons had a positive *SLBI,* and the population-averaged response to the bidirectional stimuli showed a strong bias toward the right patch (Fig. 3C, D). The LWS model also provided a good fit of the response tuning curves to the bidirectional stimuli with an AS of 30º (mean PV = 95.1%, std = 5.5%, n = 47). Similar to the AS of 60º, the *SLBI* for the AS of 30º was positively correlated with the *SCPI* (Type II regression, r = 0.50) (Fig. 3E, the variances are significantly different, Levene's test, $p = 9.96 \times 10^{-10}$). The spatial bias allowed neurons to represent the motion component at their preferred spatial location more effectively than would be possible if they simply averaged (or equally pooled) the responses to the stimulus components. For a population of neurons with different spatial preferences, different component directions with a small AS can be represented by different neurons, thereby facilitating segregation.

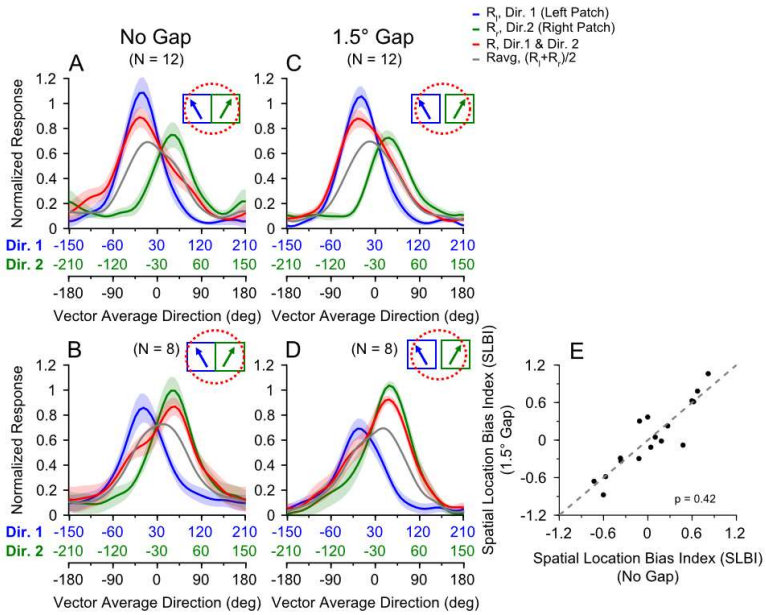

Figure 4. Comparing spatial bias in the responses to spatially separated stimuli with or without a gap between the two stimulus patches. The same group of neurons (N = 20) was tested with both stimulus conditions. The AS between the two motion directions was 60°. **A, B.** Population-averaged direction tuning curves from the experiment without the gap. **C, D.** With a 1.5 gap. The green and blue abscissae follow the same convention as in Figure 1. Error bands indicate the SE. **A-D**. The left patch moved in the CC-side direction, and the right patch moved in the C-side direction. **E.** Scatter plot of the *SLBI* with and without the gap between the stimuli (N = 20). Each dot represents *SLBIs* from one neuron. The gray dotted line indicates the unity line. The *SLBIs* were not significantly different between the two conditions (Wilcoxon Sign Rank test, p = 0.42).

Properties of spatial bias

In the following two experiments, we used bidirectional stimuli with an AS of 60º to further characterize the properties of the spatial bias. First, we asked whether the spatial bias in MT was contributed by the responses of V1 neurons with RFs at the motion border. With a sharp border between two moving stimuli as used in our main experiments, V1 neurons with RFs at the border can be stimulated by both stimulus components. A previous study from our lab suggested that normalization in V1 can have a strong effect on MT responses to multiple stimuli (Wiesner et al., 2020). To address this question, we placed a 1.5° gap between the two stimulus patches (Fig. 4C). At the eccentricities we recorded from MT neurons (mean = 9.3°± 4.4°), the RF sizes of V1 neurons are typically smaller than 1.5° (Dow et al., 1981). We recorded from 20 neurons under both gap (Fig. 4C, D) and no-gap (Fig. 4A, B) conditions. The response tuning and the spatial bias were very similar between the two conditions. Twelve neurons showed a left bias (Fig. 4A, C), and eight neurons showed a right bias (Fig. 4B, D). The *SLBI* values from the with and no-gap conditions were not significantly different (p = 0.42, Wilcoxon Signed Rank test) (Fig. 4F). These results suggest that the spatial bias did not originate from V1 neurons with RFs at the motion border.

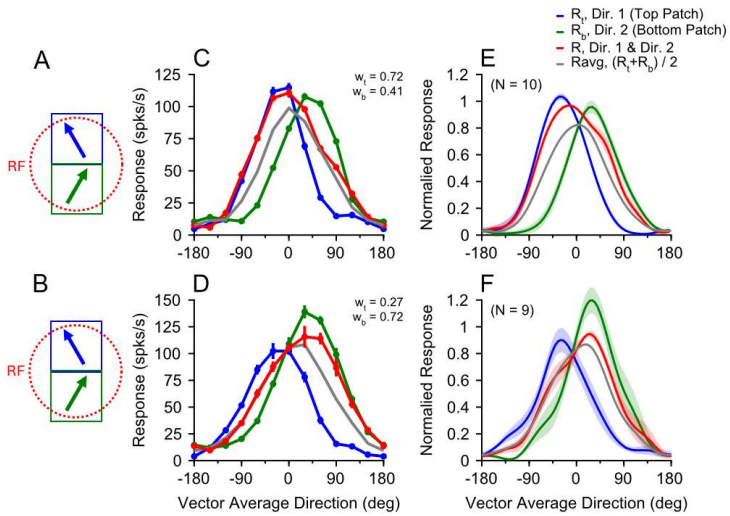

Figure 5. Spatial bias in responses to the bidirectional stimuli when the stimuli were placed in a top/bottom arrangement. The AS between the two motion directions was 60°. **A, B.** Cartoons depict the arrangement of the spatially separated stimuli with a horizontal border. The top patch always moved in the CC-side direction, and the bottom patch moved in the C-side direction. **C.** An example neuron showing a strong spatial bias towards the top stimulus component ($w_{top} = 0.72, w_{bottom} = 0.41$, weights from the LWS model fit). **D.** Another example neuron showing a strong spatial bias towards the bottom stimulus component ($w_{top} = 0.27, w_{bottom} = 0.72$). **E.** Population-averaged tuning curves of neurons showing a spatial bias towards the top stimulus component (N = 10). **F**. Population-averaged tuning curves of neurons showing a spatial bias towards the bottom stimulus component (N = 9). Error bands represent the SE.

In the second experiment, we asked whether the spatial bias depends on a specific placement of the stimulus components. We arranged the stimulus patches to be top/bottom with a horizontal motion border and no gap (Fig. 5A, B). We recorded from 19 neurons using this spatial arrangement. Figure 5C and 5D show the direction tuning curves of two example neurons. One neuron exhibited a strong top bias, although the response to the top patch was only slightly stronger than that of the bottom patch (Fig. 5C). Another neuron exhibited a bottom bias (Fig. 5D). We obtained the response weights for the top and bottom patches using the LWS model. Ten out of the 19 neurons had a larger response weight for the top patch and showed a top bias (Fig. 5E). Nine out of the 19 neurons had a larger response weight for the bottom patch and showed a bottom bias (Fig. 5F). These results suggest that the spatial bias of a neuron to bidirectional stimuli is a general phenomenon, not limited to a specific placement of stimulus components and border orientation.

Temporal development of spatial bias

To provide insight into the neural mechanism underlying the spatial bias, we characterized the time course of the neuronal responses to bidirectional stimuli with an AS of 60°, using the data set shown in Figure 2. We measured the direction tuning curves with a time window of 50 ms, sliding at a step of 10 ms. We pooled the tuning curves (with flipping as needed) to bidirectional stimuli that showed left and right biases toward the C and CC sides of the component directions (as illustrated in Fig. 2A-D), such that the spatial bias was all aligned to the negative VA directions

and the potential effect of direction side-bias was averaged out (see Methods). Figure 6 shows that the initial response tuning to the bidirectional stimuli (R_{12}) had little spatial bias during the time window centered on 45 ms. The spatial bias began to emerge during the time window centered on 55 ms and developed gradually over approximately 90 ms, reaching its maximum (Fig. 6A, B, E). The response peak of the tuning curve to the bidirectional stimuli eventually reached the VA direction of about -30°, where one of the component directions was at the neuron's preferred direction (i.e., 0°) (Fig. 6E, red curve). In other words, over time, the response peak to the bidirectional stimuli was aligned with the response peak to the component direction at the biased side, facilitating the representation of the component direction. In contrast, the average of the direction tuning curves to the individual stimulus components (R_{avg}) had a slight shift toward the stronger component response (Fig. 6C, D, E). This shift was relatively stable over time, and the response peak was around the VA direction of about -10° (Fig. 6E, black curve), much smaller than the actual response bias to the bidirectional stimuli.

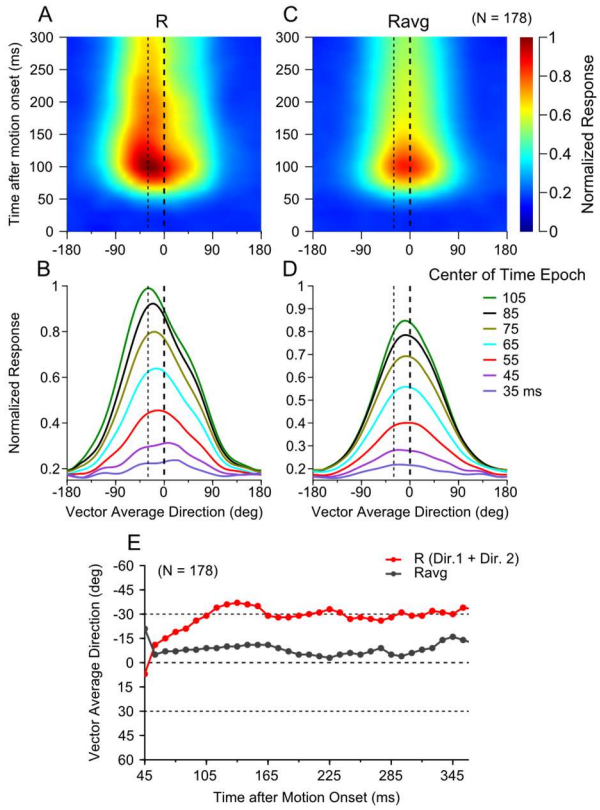

Figure 6. Time course of the neuronal response to the bidirectional stimuli. The AS between the bidirectional stimuli was 60°. Each neuron contributed to two tuning curves with the left patch moving in either the CC-side or C-side of two motion directions (total 178 tuning curves from 89 neurons). The tuning curves were aligned (and flipped relative to VA 0° as needed), such that the spatial bias of each neuron was always toward the negative VA directions. Direction tuning was calculated within a 50-ms time epoch, sliding with a step of 10 ms. **A, B.** Time course of the population-averaged response to the bidirectional stimuli. The thick dashed line shows VA 0°, which is aligned with each neuron's PD. The thin dotted line shows the VA direction of -30°, at which one of the stimulus components moved in the PD of the neuron. The ordinate in **A** indicates the center of each time epoch, relative to the motion onset. **C, D.** Time course of the average of the responses elicited by individual stimulus components. **E**: Time course of the response

peak of the tuning curves to the bi-directional stimuli (red) and the averaged component responses (black). The ordinate shows the peak location relative to the VA direction 0°. The abscissa shows the center of each time epoch after motion onset. The dashed line at -30° indicates the VA direction at which one stimulus component was moving in the PD of the neuron. The response to the bidirectional stimuli was biased toward this stimulus component.

Normalization model fits of the responses to spatially separated stimuli

Our results demonstrate that MT neuronal responses to spatially separated bidirectional stimuli can be captured by a weighted sum of the neuron's responses to the individual stimulus components. A central question, then, is what determines these weights. The divisive normalization framework (Carandini and Heeger, 2012) has provided a powerful account of diverse neural phenomena, including responses to multiple visual stimuli (e.g., Britten and Heuer, 1999; Heuer and Britten, 2002; Busse et al., 2009; Xiao et al., 2014; Xiao and Huang, 2015; Ruff and Cohen, 2016; Bao and Tsao, 2018; Chakrala et al., 2024). In this model, the role of the denominator, which scales activity by the population response of the normalization pool, is well established. However, the nature of the numerator is less understood. We have previously proposed that the weight associated with a given stimulus component is proportional to the activity it elicits in a population of neurons, which we refer to as the weighting pool (Xiao et al., 2014; Wiesner et al., 2020; Huang et al., 2025). The composition of the weighting pool remains unclear. In this study, we assumed that the response of the weighting pool for each stimulus component reflected the spatial preference of a given MT neuron. We approximated the spatial preference of an MT neuron by summing the neuron's responses to each stimulus component moving in different directions across 360°. The summed response was invariant to the motion direction of the stimulus component and reflected the neuron's spatial preference for that component.

We used the following model to fit each MT neuron's tuning curve to the bidirectional stimuli (Eq. 2):

$$R_{pred}(\theta_1, \theta_2) = \frac{S_1^n}{S_1^n + S_2^n + \sigma} R_1(\theta_1) + \frac{S_2^n}{S_1^n + S_2^n + \sigma} R_2(\theta_2) + c \quad . \tag{2}$$

R_{pred} is the model-predicted response to the bidirectional stimuli. R_1 and R_2 are the recorded responses to the individual stimulus components. θ_1 and θ_2 are two component directions. S_1 and S_2 are the integrals of the direction tuning curves of the MT neuron to the two stimulus components:

$$S_i = \int_{\theta=0}^{2\pi} R_i(\theta) d\theta \ , \qquad i = 1 \ or \ 2 \ . \tag{3}$$

S_i does not depend on the stimulus direction θ because the direction tuning curve is circular invariant. The model has three free parameters n, σ, and c, with the constraints of $0.01 \leq n \leq 100$, and $0 \leq c \leq 100$. For a neuron that has a spatial preference for stimulus component 1, S_1 is greater than S_2. Therefore, the weight for R_1 is greater than that for R_2, making the spatial bias consistent with the neuron's spatial preference. The power-law function and divisive normalization in Eq. 2 contribute to the nonlinearity of the response weight.

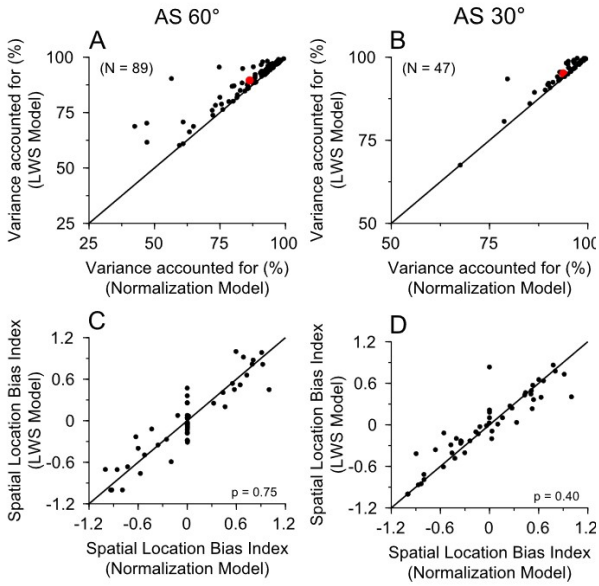

Figure 7. Comparison of the goodness of fit and *SLBI* between the normalization and LWS model. Each dot represents a single neuron. **A, B.** Scatter plot comparing the variance accounted for by the LWS (ordinate) and the normalization (abscissa) model for AS of 60° (A) and 30° (B). The red dot indicates the population means. **C, D.** Comparison of *SLBI* calculated using the weights from the LWS and normalization model, for AS of 60° (C) and 30° (D). The *SLBIs* from the two models are not significantly different (Wilcoxon Signed Rank test, p = 0.75 for AS 60°, p = 0.40 for AS 30°).

The normalization model provided a good fit of the neuronal responses to the bidirectional stimuli, accounting for, on average, 86.3% (std = 12.4%, N = 89) of the response variance for the AS of 60°, and 93.7% (std = 5.9%, N = 47) of the response variance for the AS of 30°. In addition, the exponent parameter (*n*) in the normalization model fit had a median value greater than 1 (4.4 for AS 60°, and 7.5 for AS 30°), indicating an expansive power law, which provided an explanation for why a small change in spatial preference could lead to a large change in spatial bias. The normalization model fits are comparable to the fits from the LWS model (90.6±9.4% for AS 60°, and 95.1±5.5% for AS 30°, Fig. 7), although the LWS model fits were significantly better (t-test, $p = 8.2\text{x}10^{-7}$ for AS 60°, $p = 5.16\text{x}10^{-5}$ for AS 30°). This difference is expected because the weights in the LWS are not constrained, whereas the weights in the normalization model must follow a prescribed formula based on the neuron's spatial preference. The weights for the stimulus components, and therefore the *SLBI*, derived from the normalization model were in close agreement with those derived from the LWS model (Fig. 7C, D, Wilcoxon signed-rank test, p = 0.75 and p = 0.40, respectively). These results provide a computational explanation for the spatial bias observed in responses to bidirectional stimuli.

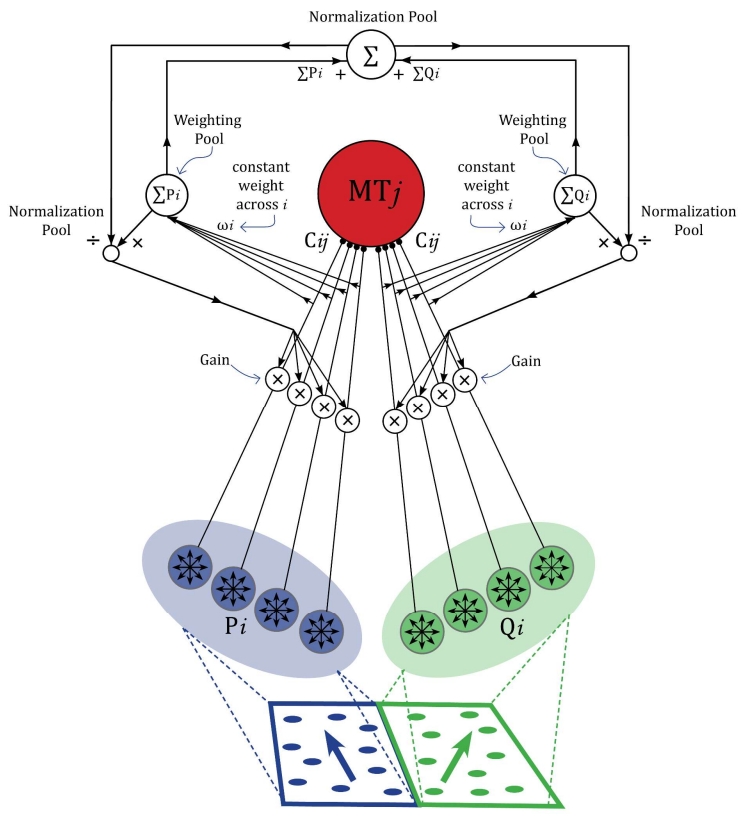

Figure 8. A schematic circuit diagram of the normalization model representing spatially separated stimuli. An MT neuron j (red circle) receives inputs from two populations of direction-selective neurons (P and Q), presumably in V1. Each population (P or Q) contains neurons with a broad range of direction preferences (black arrows). Neurons in population P have RFs covering one stimulus component (blue), and Q covering the other stimulus component (green). C_{ij} is the synaptic weight between neuron i in P (or Q) and the MT neuron j. The summed response from P (or Q), referred to as the weighting pool, multiplicatively changes the gain of the input from P (or Q) to the MT neuron. The summed response of both P and Q, referred to as the normalization pool, divisively modulates the gain. The normalization pool may also include other sources, such as neurons in MT (not shown).

Although our model is descriptive, it has implications for possible circuit implementations, as illustrated in a schematic diagram (Fig. 8). We assumed that two populations of directional-selective neurons (P and Q) in V1 (or other early visual areas projecting to MT) have RFs covering the two stimulus components, respectively, and neurons in P and Q project to an MT neuron j. At each spatial location (circle within the shaded blue and green ovals), neurons in P and Q have preferred directions spanning 360°. Therefore, the summed response for P (or Q) (with an equal weight ω_i across neuron i) is independent of the component direction and reflects the spatial preference of the inputs to the MT neuron for the corresponding stimulus component. The inputs from P and Q to the MT neuron define the spatial preference of the MT neuron for subregions of the neuron's RF. C_{ij} is the synaptic weight between neuron i in population P (or Q) and the MT neuron j. C_{ij} is subject to gain control. The response of the weighting pool is the sum of the responses of the neurons in P (or Q), with an equal weight ω_i across neurons in P (or Q). The

response of the normalization pool is the sum of the population responses of P and Q. The normalization pool may also include other sources, such as a population of MT neurons (not shown in the diagram). The gain factor g for C_{ij} is controlled by multiplying the response of the weighting pool for P or Q, respectively, and divided by the response of the normalization pool. Since the normalization pool determines the common denominator, the exact value of the normalization pool would change the response weights for the stimulus components, but not their ratio.

Discussion

We found that MT neurons showed a spatial-location bias in their responses to multiple, spatially separated stimuli moving within their RFs, and this bias reflected each neuron's spatial preference for the individual stimuli when presented alone. The sign and magnitude of the spatial bias in response to multiple stimuli were positively correlated with the neurons' spatial preference for individual stimuli. Additionally, a small difference in responses to individual stimuli at different locations could lead to a substantial spatial bias. The spatial bias developed over time for approximately 90 ms before reaching its maximum, and then remained stable.

We demonstrated that MT neuronal responses to multiple spatially separated stimuli can be well described by an extended divisive normalization model, which is a sum of the responses elicited by individual stimuli, weighted by the neuron's spatial preference, and then divisively normalized. This finding reveals a key mechanism for representing and segregating multiple spatially separated stimuli, and expands our understanding of the canonical neural computation of divisive normalization (Carandini and Heeger, 2012). We proposed a circuit implementation for the extended normalization model, suggesting that the spatial preference of an MT neuron for a particular stimulus component can be estimated by the pooled activity of the inputs from a population of neurons that represent the stimulus component and project to the MT neuron.

The observed spatial bias cannot be explained by attention modulation

Selective spatial attention directed to one of multiple stimuli within MT neurons' RFs can significantly modulate neuronal responses in MT (Treue and Maunsell, 1996; Ferrera and Lisberger, 1997; Treue and Maunsell, 1999; Seidemann and Newsome, 1999; Lee and Maunsell, 2010). In this study, however, the spatial bias found in MT responses to multiple stimuli is unlikely to be contributed by spatial attention for the following reasons. First, in our main experiments, monkeys performed a simple fixation task. The neural recordings were conducted in the left MT of both monkeys. Therefore, all the RFs were located in the right visual hemifield. Although one might expect that spatial attention may favor the stimulus patch nearer the fixation point (the left patch), neurons in the main dataset showed nearly equal numbers of left- and right-biased responses. Second, some of our recordings were conducted using a multi-channel linear electrode array (S-Probe, Plexon). We found that simultaneously recorded neurons could exhibit left- and right-biases in the same experimental sessions. Third, we collected a small amount of data from R as the monkey performed a fine-direction discrimination task with attention directed to the visual hemifield opposite MT neurons' RFs. We placed the same visual stimuli with an AS of 60° in the RFs as in our main experiment. We found similar spatial biases using the attention-away-from-RFs task as with the fixation task (data not shown). Together, these observations suggest that the spatial bias in responses to spatially separated stimuli cannot be explained by attention modulation.

Representing multiple spatially separated stimuli by weighting the neuron's spatial preference

In the framework of divisive normalization, previous studies have established that neuronal responses to multiple visual stimuli can be described by weighting the responses to individual stimuli by their luminance contrasts, which has been found in V1 (Busse et al., 2009; MacEvoy et al., 2009), V4 (Reynolds and Desimone, 2003), IT (Bao and Tsao, 2018), and MT (Heuer and Britten, 2002; Ni et al., 2012; Xiao et al., 2014; Ni and Maunsell, 2017). Similarly, neuronal responses in MT to overlapping stimuli with different motion coherences can be described by weighting the responses to individual stimuli with their motion coherences (Xiao et al., 2014). In multisensory integration of visual and vestibular cues, MSTd neurons weigh the responses elicited by individual cues according to their reliability (Morgan et al., 2008; Fetsch et al., 2011; Ohshiro et al., 2011). More recently, our lab showed that neuronal responses in MT to two overlapping stimuli moving at different depths (binocular disparities) exhibit a bias toward one of the binocular disparities, and this disparity bias is positively correlated with the neuron's disparity preference to single stimuli presented alone (Chakrala et al., 2024). Although the spatial selectivity within the two-dimensional (2D) RFs of visual neurons is well known (e.g., Raiguel et al., 1995), previous studies have not linked the neuronal response to multiple spatially separated stimuli with the 2D spatial preference of the neurons.

Our finding that neuronal responses in MT to multiple spatially separated stimuli can be captured by weighting the 2D spatial preference of the neuron to the individual stimulus components enriches the current framework of divisive normalization. Our results also provide a critical insight into how the visual system represents and segregates multiple visual stimuli. By leveraging the spatial selectivity within the RFs, the visual system nonlinearly amplifies selectivity for single stimuli at different locations into a large spatial bias in the response to multiple stimuli. In this way, neuronal responses to multiple stimuli within the RFs are not equally mixed, but mainly represent the stimulus located at the more preferred location of the neurons. For a population of neurons that have various spatial preferences within the RFs, this coding strategy would allow individual components of multiple stimuli to be represented by different neurons in the population, thereby facilitating the segregation of multiple stimuli. This spatial location-based segregation is particularly useful when other features of the visual stimuli differ only slightly, such as the motion directions in our stimuli. Theoretical studies suggest that, rather than mixing multiple stimuli with equal weight, random mixing with different weights by individual neurons can significantly improve the ability to encode multiple stimuli in neuronal populations (Orhan and Ma, 2015; Huang et al., 2017). The spatial bias found in this study demonstrates that the visual system indeed takes this theoretically effective strategy to represent and segregate multiple visual stimuli.

Neural mechanisms underlying spatial bias in response to multiple stimuli

We previously suggested that MT responses to two moving stimuli with competing signal strengths (high-contrast/low-coherence vs. low-contrast/high-coherence) are governed by response normalization in V1 when individual V1 neurons are driven by superimposed stimuli, but not when V1 neurons are driven by only one stimulus component of spatially separated stimuli (Wiesner et al., 2020). In this study, a control experiment introduced a gap between the two stimuli (Fig. 4), ensuring that individual V1 neurons were driven by only one component. Our finding of similar spatial bias in MT using the stimuli with a gap suggests that the spatial bias in MT was not simply inherited from response normalization in V1.

Our model (Eq. 2) provides several insights into possible circuit mechanisms for representing multiple spatially separated stimuli (Fig. 8). First, the model suggests that weighting is applied to the response elicited by each individual stimulus component. Since the responses elicited by individual components are mixed at the level of MT neurons, we propose that weighting by spatial preference occurs at the inputs to the MT neuron in a form of gain control, before signals from multiple stimuli are mixed. Second, our model accounted for the entire tuning curve to bidirectional stimuli using fixed weights across stimulus directions, suggesting that the weights should be invariant to stimulus directions. Therefore, we propose that the weighting pool should include neurons with different preferred directions spanning 360°. This can be implemented by equally pooling the inputs from V1 neurons that have different preferred directions, as illustrated in Figure 8. Third, the weighting should reflect the spatial preference, rather than the direction preference of the MT neuron. One way to implement this is to pool across the synaptic inputs from the same V1 population (e.g., P_i) projecting to different MT neurons having the same RF location and the same spatial preference, but with different direction preferences. By pooling across the inputs to these MT neurons, the pooled activity preserves the spatial preference of the MT neurons and is invariant to the direction selectivity of the individual MT neurons. For simplicity, our circuit diagram did not depict multiple MT neurons with different direction preferences, but instead showed pooling from the V1 population (Fig. 8).

Materials and Methods

Two adult male rhesus monkeys (*Macaca mulatta*) participated in the neurophysiological experiments. Experimental protocols were approved by the Institutional Animal Care and Use Committee of UW-Madison and conform 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 routine and similar to those described previously (Xiao et al., 2014). A headpost and a recording cylinder were implanted during sterile surgery with the animal under isoflurane anesthesia. We took a vertical approach to access area MT. We identified area MT by its characteristically large portion of directionally 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 and Gross, 1981). For most of the electrophysiological recordings, we used tungsten electrodes (1-3 MΩ, FHC). For some experiments, we used a 16-channel laminar electrode array (S-Probe, Plexon Inc.) to record multiple neurons at the same time. Electrical signals were amplified, and single units were identified with a real-time template-matching system and an offline spike sorter (Plexon). Eye position was monitored using a video-based eye tracker (EyeLink, SR Research) with a rate of 1000 Hz.

Visual stimuli and experimental procedure

Stimulus presentation and data acquisition were controlled by a real-time data acquisition program "Maestro" (https://sites.google.com/a/srscicomp.com/maestro/home). Visual stimuli were presented on a 25-inch CRT monitor at a viewing distance of 63 cm. Monitor resolution was 1024 x 768 pixels (or 1152 x 864 pixels) with a refresh rate of 100 Hz. Stimuli were generated by a Linux workstation using an OpenGL application that communicated with an experimental

control computer. The luminance of the video monitor was measured with a photometer (LS-110, Minolta) and was gamma-corrected.

During initial data collection, for each stimulus patch, achromatic random dots were presented and moving within a 10° x 10° static square aperture. For most of the data collected, the size of each random-dot patch was reduced to 5° x 5°. Individual dots were squares extending ~0.08° on each side, and each random-dot patch had a dot density of 2.7 dots/deg^2. The dots had a luminance of 22 cd/m^2, presented on a uniform background with a luminance of 10 cd/m^2, giving rise to a Michelson contrast of 37.5% (Lum dot - Lum bkg)/(Lum dot + Lum bkg). Random dots in each patch moved within the stationary aperture in a specified direction, with a motion coherence of 100%. The lifetime of each dot was as long as the motion duration, such that if a dot exited the aperture during the motion period, it was immediately replotted on the opposite side.

In each experimental trial, the monkey maintained fixation within a 1° × 1° electronic window around a small fixation point. After a neuron was isolated, we first characterized its direction selectivity by interleaving trials of random dots moving within a 30° × 27° aperture, in single directions at a step of 45°, at a fixed speed of 10°/s. The direction selectivity and the preferred direction were determined by on-line analysis 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 neuron's PD. Using a cubic spline, the preferred speed (PS) of the neuron was taken as the speed that evoked the highest firing rate in the fitted speed tuning curve. Next, we used a series of 5° × 5° random-dot patches moving in the PD and at the PS of the neuron to map the neuron's receptive field (RF). The location of the patch was randomized, and the visual display was tiled in 5° steps without overlapping in a 30° x 30° aperture. The RF map was interpolated at 0.5° intervals, and the location giving rise to the highest firing rate was taken as the center of the RF. RF eccentricity of our neuronal population averaged 9.3° (std = ±4.4°) away from the fovea.

The visual stimuli were presented after the monkey maintained fixation for 250ms. To isolate the neuronal responses to the stimulus motion from those due to the stimulus onset, the visual stimuli were first turned on and remained stationary for 300ms before moving in a specific direction for 500ms. The visual stimuli were then turned off. The monkeys maintained fixation for an additional 250ms after the stimulus offset. In some trials, two random-dot patches moved in two different directions, referred to as the bidirectional stimuli, were presented simultaneously. We refer to each of the random-dot patches as a stimulus component. In the main experiment, the two stimulus components were placed side by side, with the border between them centered within the RF, and the AS between the two motion directions either 60° or 30°. Additional manipulations included adding a 1.5° gap between the two stimulus components (Fig. 4) or placing them in a top/bottom orientation so that the motion discontinuity boundary was horizontal (Fig. 5). We varied the VA direction of the bidirectional stimulus around 360° to characterize the response tuning curve. In other trials, only one stimulus component was presented on either the left or right side of the RF, and the direction was varied to characterize the tuning curve for each stimulus component. For most experiments, the VA direction of the bi-speed stimuli and the direction of the component stimulus were varied in 30° increments. At the AS of 30°, the directions were varied in a step of 15°. The trials presenting bidirectional stimuli and individual stimulus components were randomly interleaved. To control for a possible direction side bias toward either the clockwise or counter-clockwise component (Xiao and Huang, 2015), we included two sets of trials for the bidirectional stimulus. In one trial set, the component on the left side of the RF moved in the

direction on the counter-clockwise side of the two component directions, whereas the component on the right side moved in the direction on the clockwise side. In the other trial set, the left component moved on the clockwise side and the right component moved on the counterclockwise side.

Experimental design and statistical analysis

Response firing rate was calculated during the period of stimulus motion and averaged across repeated trials. We fitted the raw direction tuning curves for the bidirectional stimuli and the individual stimulus components using splines at a resolution of 1°. We then rotated the spline-fitted tuning curve to the bidirectional stimuli so that the VA direction of 0° was aligned with the PD of each neuron. The responses of each neuron to the bidirectional stimulus and individual stimulus components were normalized by the maximum spline-fitted response to the bidirectional stimuli. We then averaged the rotated and normalized tuning curves across neurons to obtain population-averaged tuning curves. To determine a neuron's spatial preference, we calculated the area under the response tuning curve (AUC) for both the left and right stimulus components (S_l and S_r, respectively). To calculate the AUC of each stimulus component, we used the spline-fitted response, then summed the firing rates across 360° (Wiesner et al., 2020).

To characterize the temporal development of the spatial bias regardless of whether the bias was toward the left or right patch in the RF, we combined direction tuning curves showing either left or right bias, and with different direction arrangements. To achieve this, we flipped the response tuning curves of the left-bias (or right-bias) neurons when the left (or right) component moved in the clockwise direction along the X-axis relative to VA at 0°. This flip changed the response bias toward the negative VA directions. For the tuning curves that showed either left (or right) bias and the left (or right) component moved in the counter-clockwise direction, the response bias was already toward the negative VA directions. We therefore did not flip these tuning curves. We then averaged all the flipped and not flipped tuning curves across the neuron population.

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) accounted for by the model as:

$$PV = 100 \times \left(1 - \frac{SSE}{SST}\right) , \tag{4}$$

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 (Morgan et al., 2008).

References

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*(6), Article 6. https://doi.org/10.1152/jn.1984.52.6.1106

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, *14*(2), 105–126. https://doi.org/10.1068/p140105

Bao, P., & Tsao, D. Y. (2018). Representation of multiple objects in macaque category-selective areas. *Nature Communications*, *9*(1), 1774. https://doi.org/10.1038/s41467-018-04126-7

Black, M. J., & Fleet, D. J. (2000). Probabilistic Detection and Tracking of Motion Boundaries. *International Journal of Computer Vision*, *38*(3), 231–245. https://doi.org/10.1023/A:1008195307933

Born, R. T. (2000). Center-surround interactions in the middle temporal visual area of the owl monkey. *J Neurophysiol*, *84*(5), 2658–2669. https://doi.org/10.1152/jn.2000.84.5.2658

Born, R. T., & Bradley, D. C. (2005). STRUCTURE AND FUNCTION OF VISUAL AREA MT. *Annual Review of Neuroscience*, *28*(1), 157–189. https://doi.org/10.1146/annurev.neuro.26.041002.131052

Braddick, O. (1993). Segmentation versus integration in visual motion processing. *Trends Neurosci*, *16*(7), 263–268. https://doi.org/10.1016/0166-2236(93)90179-p

Britten, K. H. (1999). How are moving images segmented? *Current Biology: CB*, *9*(19), R728-730. https://doi.org/10.1016/s0960-9822(99)80469-2

Britten, K. H. (2003). *The Middle Temporal Area: Motion Processing and the Link to Perception*. https://direct.mit.edu/books/edited-volume/5395/chapter/3948247/The-Middle-Temporal-Area-Motion-Processing-and-the

Britten, K. H., & Heuer, H. W. (1999). Spatial summation in the receptive fields of MT neurons. *J Neurosci*, *19*(12), 5074–5084.

Busse, L., Wade, A. R., & Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron*, *64*(6), 931–942. https://doi.org/10.1016/j.neuron.2009.11.004

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nat Rev Neurosci*, *13*(1), 51–62. https://doi.org/10.1038/nrn3136

Chakrala, A. S., Xiao, J., & Huang, X. (2024). *The role of binocular disparity and attention in the neural representation of multiple moving stimuli in the visual cortex* (p. 2023.06.25.546480). bioRxiv. https://doi.org/10.1101/2023.06.25.546480

Cherian, C., & Maunsell, J. H. R. (2025). *Neuronal normalization in monkey MT is an intensity-weighted average* (p. 2025.07.21.665985). bioRxiv. https://doi.org/10.1101/2025.07.21.665985

Feldman, D., & Weinshall, D. (2008). Motion Segmentation and Depth Ordering Using an Occlusion Detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(7), 1171–1185. https://doi.org/10.1109/TPAMI.2007.70766

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2011). Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci*, *15*(1), 146–154. https://doi.org/10.1038/nn.2983

Gattass, R., & Gross, C. G. (1981). Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*(3), 621–638. https://doi.org/10.1152/jn.1981.46.3.621

Grossberg, S. (1991). Why do parallel cortical systems exist for the perception of static form and moving form? *Perception & Psychophysics*, *49*(2), 117–141. https://doi.org/10.3758/BF03205033

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Percept Psychophys*, *55*(1), 48–121. https://doi.org/10.3758/bf03206880

Heuer, H. W., & Britten, K. H. (2002). Contrast dependence of response normalization in area MT of the rhesus macaque. *J Neurophysiol*, *88*(6), 3398–3408. https://doi.org/10.1152/jn.00255.2002

Huang, W., Huang, X., & Zhang, K. (2017). Information-theoretic interpretation of tuning curves for multiple motion directions. *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, 1–4. https://ieeexplore.ieee.org/abstract/document/7926142/

Huang, X., Albright, T. D., & Stoner, G. R. (2007). Adaptive surround modulation in cortical area MT. *Neuron*, *53*(5), 761–770. https://doi.org/10.1016/j.neuron.2007.01.032

Huang, X., Albright, T. D., & Stoner, G. R. (2008). Stimulus dependency and mechanisms of surround modulation in cortical area MT. *J Neurosci*, *28*(51), 13889–13906. https://doi.org/10.1523/JNEUROSCI.1946-08.2008

Huang, X., Ghimire, B., Chakrala, A. S., & Wiesner, S. (2025). Neural coding of multiple motion speeds in visual cortical area MT. *eLife*, *13*. https://doi.org/10.7554/eLife.94835.2

Huang, X., & Lisberger, S. G. (2009). Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. *J Neurophysiol*, *101*(6), 3012–3030. https://doi.org/10.1152/jn.00010.2009

Kumano, H., & Uka, T. (2010). The spatial profile of macaque MT neurons is consistent with Gaussian sampling of logarithmically coordinated visual representation. *J Neurophysiol*, *104*(1), 61–75. https://doi.org/10.1152/jn.00040.2010

Lee, J., & Maunsell, J. H. R. (2010). The effect of attention on neuronal responses to high and low contrast stimuli. *Journal of Neurophysiology*, *104*(2), 960–971. https://doi.org/10.1152/jn.01019.2009

Livingstone, M. S., Pack, C. C., & Born, R. T. (2001). Two-Dimensional Substructure of MT Receptive Fields. *Neuron*, *30*(3), 781–793. https://doi.org/10.1016/S0896-6273(01)00313-0

MacEvoy, S. P., Tucker, T. R., & Fitzpatrick, D. (2009). A precise form of divisive suppression supports population coding in the primary visual cortex. *Nature Neuroscience*, *12*(5), 637–645. https://doi.org/10.1038/nn.2310

McDonald, J. S., Clifford, C. W. G., Solomon, S. S., Chen, S. C., & Solomon, S. G. (2014). Integration and segregation of multiple motion signals by neurons in area MT of primate. *Journal of Neurophysiology*, *111*(2), 369–378. https://doi.org/10.1152/jn.00254.2013

Morgan, M. L., DeAngelis, G. C., & Angelaki, D. E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, *59*(4), 662–673. https://doi.org/10.1016/j.neuron.2008.06.024

Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception*, *3*(1), 63–80. https://doi.org/10.1068/p030063

Ni, A. M., & Maunsell, J. H. R. (2017). Spatially tuned normalization explains attention modulation variance within neurons. *Journal of Neurophysiology*, *118*(3), 1903–1913. https://doi.org/10.1152/jn.00218.2017

Ni, A. M., Ray, S., & Maunsell, J. H. R. (2012). Tuned Normalization Explains the Size of Attention Modulations. *Neuron*, *73*(4), 803–813. https://doi.org/10.1016/j.neuron.2012.01.006

Ohshiro, T., Angelaki, D. E., & DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nature Neuroscience*, *14*(6), 775–782. https://doi.org/10.1038/nn.2815

Orhan, A. E., & Ma, W. J. (2015). Neural population coding of multiple stimuli. *J Neurosci*, *35*(9), 3825–3841. https://doi.org/10.1523/JNEUROSCI.4097-14.2015

Pasternak, T., & Tadin, D. (2020). Linking Neuronal Direction Selectivity to Perceptual Decisions About Visual Motion. *Annual Review of Vision Science*, *6*(1), 335–362. https://doi.org/10.1146/annurev-vision-121219-081816

Qian, N., & Andersen, R. A. (1994). Transparent motion perception as detection of unbalanced motion signals. II. Physiology. *J Neurosci*, *14*(12), 7367–7380.

Raiguel, S., Van Hulle, M. M., Xiao, D. K., Marcar, V. L., & Orban, G. A. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *The European Journal of Neuroscience*, *7*(10), 2064–2082. https://doi.org/10.1111/j.1460-9568.1995.tb00629.x

Regan, D., & Beverley, K. I. (1984). Figure-ground segregation by motion contrast and by luminance contrast. *Journal of the Optical Society of America. A, Optics and Image Science*, *1*(5), 433–442. https://doi.org/10.1364/josaa.1.000433

Reynolds, J. H., & Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron*, *37*(5), 853–863. https://doi.org/10.1016/s0896-6273(03)00097-7

Richert, M., Albright, T. D., & Krekelberg, B. (2013). The complex structure of receptive fields in the middle temporal area. *Frontiers in Systems Neuroscience*, *7*, 2. https://doi.org/10.3389/fnsys.2013.00002

Ruff, D. A., & Cohen, M. R. (2016). Stimulus Dependence of Correlated Variability across Cortical Areas. *Journal of Neuroscience*, *36*(28), 7546–7556. https://doi.org/10.1523/JNEUROSCI.0504-16.2016

Seidemann, E., & Newsome, W. T. (1999). Effect of spatial attention on the responses of area MT neurons. *Journal of Neurophysiology*, *81*(4), 1783–1794. https://doi.org/10.1152/jn.1999.81.4.1783

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *J Neurosci*, *11*(9), 2768–2785.

Stoner, G. R., & Albright, T. D. (1992). Neural correlates of perceptual motion coherence. *Nature*, *358*(6385), 412–414. https://doi.org/10.1038/358412a0

Treue, S., Hol, K., & Rauber, H.-J. (2000). Seeing multiple directions of motion—Physiology and psychophysics. *Nature Neuroscience*, *3*(3), 270–276. https://doi.org/10.1038/72985

Treue, S., & Maunsell, J. H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, *382*(6591), 539–541. https://doi.org/10.1038/382539a0

Treue, S., & Maunsell, J. H. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *19*(17), 7591–7602.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wertheimer, M. (1938). Laws of organization in perceptual forms. In *A source book of Gestalt psychology* (pp. 71–88). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-005

Wiesner, S., Baumgart, I. W., & 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*(9), 1834–1848. https://doi.org/10.1523/JNEUROSCI.1950-19.2020

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*(49), 16180–16198. https://doi.org/10.1523/JNEUROSCI.2175-15.2015

Xiao, J., Niu, Y. Q., Wiesner, S., & Huang, X. (2014). Normalization of neuronal responses in cortical area MT across signal strengths and motion directions. *J Neurophysiol*, *112*(6), 1291–1306. https://doi.org/10.1152/jn.00700.2013

Yoonessi, A., & Baker, C. L., Jr. (2011). Contribution of motion parallax to segmentation and depth perception. *Journal of Vision*, *11*(9), 13. https://doi.org/10.1167/11.9.13