Spatial arrangement drastically changes the interaction between visual stimuli that compete in multiple feature domains in extrastriate area MT

Steven Wiesner, Ian Baumgart, Xin Huang

Department of Neuroscience, School of Medicine and Public Health
McPherson Eye Research Institute
University of Wisconsin - Madison, WI 53705, U.S.A.

Correspondence should be addressed to:

Xin Huang, Department of Neuroscience, School of Medicine and Public Health, University of Wisconsin, Madison, WI 53705, USA
Email: *Xin.Huang@wisc.edu*

Acknowledgment

We thank Jianbo Xiao for assistance on electrophysiological recordings and data analysis, Jennifer Gaudio Carson for assistance on animal training. This work was supported by National Institutes of Health Grant R01EY022443.

ABSTRACT

Natural scenes often contain multiple objects. However, how neurons in the visual cortex represent multiple visual stimuli within their receptive fields (RFs) is not well understood. Previous studies have shown that, when multiple stimuli compete in one feature domain, the evoked neuronal response is dominated by the stimulus component that has a stronger signal strength, which can be explained by response normalization. Here we investigate how neurons in middle-temporal (MT) cortex of the macaque monkey represent multiple stimuli that compete in more than one feature domain within their RFs. Visual stimuli were two random-dot patches moving in different directions. One stimulus moved at high coherence with low luminance contrast, whereas the other moved at low coherence with high contrast. We found that, although the peak MT response elicited by the "low contrast & high coherence" stimulus alone was stronger than by the "high contrast & low coherence" stimulus, MT response to both stimuli when they were overlapping was almost completely dominated by the high-contrast stimulus. When two stimuli were spatially separated within the RF, the contrast dominance was abolished. We found the same results when using contrast to compete with motion speed. Computer simulations using a V1-MT model suggest that the contrast dominance is due to normalization occurring at input stage fed to MT and MT neurons cannot overturn it according to their own feature selectivity. Our results reveal new rules on stimulus competition and highlight the importance of hierarchical processing on the neural representation of multiple visual stimuli in the extrastriate cortex.

SIGNIFICANCE STATEMENT

We investigated the rules by which cortical neurons represent multiple visual stimuli that compete in more than one feature domain. We found that the interaction between multiple stimuli within the RFs of neurons in area MT depends on the spatial arrangement of the stimuli. When multiple stimuli are overlapping, the response tuning curves of MT neurons are strongly dominated by the stimulus component that has higher luminance contrast albeit lower motion coherence or a slower speed than a competing stimulus. When multiple stimuli are spatially separated, the contrast dominance is abolished. These results cannot be explained by response normalization only within MT, but reveal the importance of hierarchical processing on the neural representation of multiple visual stimuli in extrastriate cortex.

Introduction

In natural scenes, multiple visual stimuli are often present in a local spatial region. While it is generally well understood how neurons in the visual cortex encode a single visual stimulus, how neurons encode multiple visual stimuli within their receptive fields (RFs) remains to be elucidated. Because visual perception depends critically on the integration and segregation of multiple visual stimuli (Braddick, 1993), understanding the neural representation of multiple visual stimuli is of significant importance.

Extrastriate middle-temporal (MT) cortex is a brain region that is important for visual motion processing (Britten, 2003; Born and Bradley, 2005; Park and Tadin, 2018). Neurons in area MT receive feedforward inputs from direction selective neurons in V1 (Movshon and Newsome, 1996) and have RFs about ten times larger in size than those of V1 neurons at the same eccentricities (Gattass and Gross, 1981; Albright and Desimone 1987). Previous studies have shown that neuronal responses in area MT elicited by multiple moving stimuli follow a sub-linear summation of the responses elicited by the individual stimulus components (Snowden et al., 1991; Qian and Andersen, 1994; Recanzone et al., 1997; Ferera and Lisberger, 1997; Britten and Heuer, 1999; Heuer and Britten, 2002; McDonald et al., 2014), consistent with a model of divisive normalization (Simoncelli and Heeger, 1998; Britten and Heuer, 1999; Carandini and Heeger, 2011).

Work in our laboratory has shown that the direction tuning curves of MT neurons in response to overlapping random-dot stimuli moving transparently in two different directions can also be described as a weighted sum of the responses elicited by the individual stimulus components (Xiao et al., 2014; Xiao and Huang, 2015). When two stimulus components have different signal strengths in one feature domain, defined either by motion coherence or luminance contrast, MT neurons pool the stimulus component that has a stronger signal strength with a greater weight (Xiao et al., 2014). The response bias in area MT toward the stimulus component that has a stronger signal strength can be accounted for by a descriptive model of divisive normalization (Xiao et al., 2014), similar to the contrast normalization model used to describe neuronal responses in V1 (Carandini et al., 1997; Busse et al. 2009).

However, natural scenes may have multiple visual stimuli that often differ in more than one feature domain. For example, one stimulus may have a stronger signal strength in feature A but a weaker signal strength in feature B, whereas another stimulus may have a weaker signal strength in feature A but a stronger signal strength in feature B. In this case, it is unclear which stimulus has an overall stronger signal strength and, more generally, how visual stimuli with multiple competing features interact within neurons' RFs.

One possibility is that, to neurons in a given visual cortical area, the overall signal strength of a visual stimulus is reflected in the evoked responses of a large population of neurons in that area. Due to divisive normalization in that visual area, a neuron might weigh a visual stimulus more strongly if the population neural response elicited by that stimulus is greater than the population response elicited by a competing stimulus. Alternatively, how neurons in a given

brain area weigh multiple competing stimuli may be the result of neural computations occurring in multiple stages along the hierarchical visual pathway and may not be explained by simply considering the population responses elicited in that brain area by the individual stimulus components.

Here, we investigate the rule of how neurons in area MT encode multiple moving stimuli that compete in more than one feature domain. We found that MT responses to multiple stimuli changed drastically when the spatial arrangement of the visual stimuli was varied. Our results reveal how visual stimuli that differ in multiple feature domains interact within neurons' RFs and shed light on how the neuronal responses in a given cortical area is shaped by neural processing along the visual hierarchy.

Materials and Methods

Two adult male rhesus monkeys (*Macaca mulatta*) were used 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., 2015). During sterile surgery with the animal under isoflurane anesthesia, a head post and recording cylinder were implanted to allow recording from neurons in cortical area MT.

We used tungsten electrodes (FHC) for electrophysiological recordings. We identified area MT by its characteristically large portion of directionally selective neurons, small receptive fields (RFs) relative to those of neighboring medial superior temporal cortex (area MST), and its location at the posterior bank of the superior temporal sulcus. 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 (https://sites.google.com/a/srscicomp.com/maestro/home). Visual stimuli were presented on a 25-inch CRT monitor at a viewing distance of 63cm. Monitor resolution was 1024x768 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.

Visual stimuli were achromatic random-dot patches presented within a stationary, circular aperture with a diameter of 3°. Individual dots were squares of 2 pixels 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 either 79 or 22 cd/m^2, presented on a uniform background with a luminance of 10 cd/m^2, which gives rise to a Michelson contrast of either 77.5% or 37.5%. Random dots in each

patch moved within the stationary aperture in a specified direction. The motion coherence of each random-dot patch was set to either 100% or 60%. To generate a random-dot patch moving at N% of motion coherence (after Britten et al. 1992; Newsome and Pare 1988), N% of the dots were selected to move coherently while the rest of the dots were repositioned randomly within the stationary aperture. Random selections of signal and noise dots occurred at each monitor frame. Therefore, a given dot would switch back and forth between a signal dot and a noise dot as it traveled across the circular aperture. The lifetime of each dot was as long as the motion duration.

Each stimulus was presented in a single trial, while the monkey maintained fixation within a 1° × 1° window around a small fixation point. The stimuli appeared after the monkey maintained fixation for 200ms. In order to separate the neuronal responses to the stimulus motion from those due to the stimulus onset, the dots remained stationary for 200ms before they began to move. The dots moved for 500ms, before being turned off. The monkeys maintained fixation for an additional 200ms. We varied the vector averaged (VA) direction of the bi-directional stimulus around 360° to characterize the response tuning curve.

The direction separation between the two random-dot patterns referred to as two stimulus components was fixed at 90°. Note that when a random-dot pattern moved at 60% coherence in a given direction, the visual stimuli were drastically different from a situation where 60% of the dots moved at 100% coherence and 40% of dots moved randomly. Because the random selections of signal and noise dots occurred at each monitor frame, a noise dot at one frame may have turned into a signal dot in the next frame and moved in the direction of the random-dot

pattern. As a result, the noise dots did not appear to be separable from the random-dot pattern. When a random-dot pattern of 60% coherence was superimposed with a random-dot pattern of 100% coherence, the frame-to-frame noise dots of the 60% coherence pattern did not appear to interfere with the random-dot pattern of 100% coherence. Trials with different vector averaged (VA) directions of the two stimulus components were randomly interleaved to characterize the response tuning to the bidirectional stimuli.

Once a neuron was isolated, we began by characterizing its direction selectivity by interleaving trials of a 30° × 27° random-dot patch, moving at 10°/s in different directions at steps of 45°. The direction selectivity and preferred direction (PD) were determined on-line using MATLAB (The 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 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 screen was tiled in 5° steps to cover a 35° × 25° area of the visual field. 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 neuron's RF. For the main experiments, The raw RF map was used to determine locations within the RF that would allow the spatially-separated random-dot patches to be separated by at least 1° of visual angle. Once these locations were determined, they were used as the two possible locations for the overlapping bi-directional stimuli (site a and site b, respectively) .

In the first experiment, both stimulus components moved at the same speed, which was set at the neuron's PS if it was below 10°/s. If the PS was greater than 10°/s, the speed was set at 10°/s. One random-dot patch which we labeled the "low contrast & high coherence" component, had a motion coherence of 100%, with a luminance contrast of 37.5%. The other random-dot patch, which we termed the "high contrast & low coherence" component had a motion coherence of 60%, with a luminance contrast of 77.5%. Trials that contained a single stimulus component were interleaved with our bi-directional stimuli. The bi-directional stimuli were either overlapping in one of two possible locations within the RF (as described earlier), or they were spatially separated with at least 1° gap between the outer borders of the two random-dot patches. In a given trial with the bi-directional stimuli, the random-dot patches translated in two directions with an angle separation of 90°. In order to characterize the direction tuning curves, individual trials with a different vector averaged (VA) direction were randomly interleaved. For the majority of the experiments, the vector averaged direction was varied in 15° steps. In a small set of experiments, the VA direction was varied in 30° steps.

In the second experiment, we set the motion coherence of both random-dot patches to 100%, but altered the speed of the dots for each component independently. The dots of one stimulus component were given a faster speed (10°/s), with the same low luminance contrast of 37.5%. This was referred to as the "low contrast & high speed" component. The dots of the other stimulus component, the "high contrast & low speed" component, moved at a speed of 2.5°/s, with the same higher luminance contrast of 77.5%. The direction separation between the two components was again set at 90°. Trials with different VA directions were randomly interleaved in steps of 30° to characterize tuning curves. Trials containing a single stimulus component were

also randomly interleaved. For the bi-directional stimuli, the random-dot patches were either overlapping or spatially-separated within the RF in the same manner as the first experiment.

Data analysis

For each individual stimulus component and each VA direction of the bi-directional stimuli, the response firing rate was taken as the spike count during the 500ms motion period, and was averaged across repeated trials. We fit the raw tuning curves for each individual component and the bi-directional stimuli with a cubic spline at a resolution of 1°, then rotated the spline-fitted tuning curve for the bi-directional stimuli so that the PD of each neuron was aligned with the VA direction of 0°. The responses for each neuron were normalized by the maximum response to the "low contrast & high coherence" component, then each of the rotated, normalized tuning curves were averaged together across neurons to comprise our population average tuning curves.

To quantify the relationship between the responses elicited by the bi-directional stimuli and those elicited by the individual stimulus components, we fitted the direction tuning curves using a summation plus nonlinear interaction (SNL) model:

$$R_{pred}(\theta 1,\ \theta 2) = w_1 \cdot R_1(\theta 1) + \ w_2 \cdot R_2(\theta 2) + b \cdot R_1(\theta 1) \cdot R_2(\theta 2), \qquad (1)$$

where R_{pred} is the response to the bi-directional stimuli predicted by the model; θ_1 and θ_2 are the two component directions; R_1 and R_2 are the measured component responses elicited by the "low contrast & high coherence" component and the "high contrast & low coherence" component, respectively; w_1 and w_2 are the response weights for R_1 and R_2, respectively; and b is the

coefficient of the nonlinear interaction between the component responses. To determine the goodness of fit of the model for each neuron, we calculated the percentage of variance (PV) that the model accounted for as

$$PV = 100\ x\ (1 - \frac{SSE}{SST}), \qquad (2)$$

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.

Model

We adapted the feedforward model proposed by Simoncelli and Heeger (1998) (available http://www.cns.nyu.edu/~lcv/MTmodel/) to reconstruct the experimental paradigm. Briefly, stages of the model correspond to one of the consecutive computational operations, shown simplified in Figure 7. Certain stages are interpreted as V1 simple, V1 complex, and MT cell firing rates. Based on monitor dimensions and viewing distance, we determined that 21 pixels constituted 1 visual degree. We constructed stimuli using two-by-two pixel dots with 3 degree (63 pixel) diameter circular apertures. We set neuron RFs by Gaussian convolutional filters as described in Table 1 and estimated the size of the RF for each neuron type by summing the lengths of the incorporated filters. For the spatially-separated condition, we set the gap between stimulus components to the size of the V1 complex neuron, 1.2 degrees, to ensure that no neuron would be driven by both components. We used directional cell populations such that the frequency domain approximately tiled a sphere in the frequency domain. We tuned the contrast

response functions by adjusting C50 values for V1 and MT neurons. These C50 values are represented in the model as σ^2 in the normalization equation (Equation 3), which is applied to both V1 and MT stages of the model.

$$R'_n(t) = \frac{\lfloor R_n(t) \rfloor}{K \sum_m \lfloor w_m \cdot R_m(t) \rfloor + \sigma^2} \tag{3}$$

$R_n(t)$ represents the nth neuron's linear filter response, $R'_n(t)$ represents its normalized response, $\lfloor\ \rfloor$ denotes the half-wave rectification operation, K represents the normalization strength, m represents the nth neuron's local population pool, and w represents the spatial weighting profile of the normalization pool. We fit simulated contrast response functions to physiological functions for V1 and MT from equations described in Sclar, Maunsell, and Lennie (1990), as shown in Figure 7. Similarly, we tuned coherence responses by varying the spatial scale of the normalization pool (m), the weights within the pool (w), and the size of the V1 linear receptive fields (n) (Figure 7). The MT coherence response function was fitted to data replotted from a figure published in Britten et al. (1998). Although only the MT coherence response function was fitted, changes to V1 parameters were required to simulate the response.

Table 1. Fitted model parameters

Variable for V1 neurons	Value	Variable for MT neurons	Value
V1 Frequency Scales	3		
Simple Cell RF	15 px		
Baseline	0	Baseline	0.1
C50 (σ^2)	0.04	C50 (σ^2)	0.007
Complex Pool Size	25 px	Spatial Pool Filter Size	187 px
Spatial Pool (sd)	7 px	Spatial Pool (sd)	53.4 px
Normalization Pool Size	15 px	Normalization Pool Size	153 px
Normalization Pool (sd)	5 px	Normalization Pool (sd)	51.7 px

The model's architecture was based on current understanding of cortical physiology and was adapted to better fit the electrophysiology data. The model parameters were fit before and after each architectural change, and the following changes enabled the model to capture the feature competition trends found in our data. V1 linear receptive fields were 3rd order Gaussian derivatives with length defined by the above simple cell RF. We used area-normalized Gaussian functions to set the weights for the spatial pooling and local population normalization operations. Multiple frequency scales were computed by tripling the sigma of the underlying Gaussian, similar to the doubling suggested in Simoncelli and Heeger (1998). This change was made after spectral analysis of stimuli showed that a wider range of scales was necessary to capture incoherent motion. V1 afferent weights were not adjusted to zero mean, allowing MT neurons to have variable proportions of positive and negative inputs. Importantly, rectification and static nonlinearity was applied to the MT level after spatial pooling and before normalization, which is both more physiologically plausible and better models the electrophysiology data.

Results

We asked the question of how neurons in extrastriate area MT represent multiple visual stimuli that compete in more than one feature domain. To address this question, we conducted neurophysiological experiments and computer simulations. We recorded from isolated single neurons in area MT of two macaque monkeys while they performed a fixation task. Visual stimuli were two random-dot patches moving simultaneously in different directions within the RFs. In the first experiment, we used luminance contrast and motion coherence as two competing features. One stimulus had a high luminance contrast but moved with low coherence, whereas the other stimulus had a low luminance contrast but moved with high coherence (see Methods). We manipulated the spatial arrangement of our visual stimuli to investigate the contributions of both earlier visual areas and area MT in mediating the competition between multiple visual stimuli. In a second experiment, we used luminance contrast and motion speed as the two competing features. We first present the results from the neurophysiological recordings and then computer simulations.

Neurophysiological experiments

We measured the direction tuning curves of MT neurons in response to two stimuli that had competing visual features and moved simultaneously in different directions. Our dataset includes recordings from 76 MT neurons, 43 from monkey G and 33 from monkey B. We set the angular separation between the motion directions of two individual stimuli, referred to as the stimulus components, at 90° and varied the vector averaged (VA) direction. In the first

experiment, one stimulus component had a low luminance contrast of 37.5% and moved at a high motion coherence of 100%. The other component had a high luminance contrast of 77.5% and moved at a low coherence of 60%. Figure 1 shows the direction tuning curves of two representative neurons. The red curve shows the neuronal response elicited when both stimulus components were present, as a function of the VA direction of the stimulus components. The green and blue curves show the neuronal responses elicited by the individual stimulus components when presented alone. The tuning curves of the component responses are arranged such that at each VA direction, the data points on the green and blue curves correspond to the responses elicited by the individual stimulus components of that VA direction (note the color-coded abscissas for the component directions in Fig. 1A2).

The two example neurons responded more strongly to the "low contrast & high coherence" component (shown in green) than to the "high contrast & low coherence" component (shown in blue) when the motion direction is at the preferred direction (PD) of each neuron (Fig. 1). This is expected since MT neurons are sensitive to motion coherence within a large coherence range (Britten et al., 1993), whereas their contrast response function saturates at a low luminance contrast (Sclar et al., 1990). Consequently, the average of the response tuning curves to the two stimulus components (shown in gray) was biased toward the "low contrast & high coherence" component. Surprisingly, we found that when the two stimulus components were spatially overlapping within the RF, the neuronal responses elicited by the bi-directional stimuli were strongly biased toward the "high contrast & low coherence" component (Fig. 1A). This response bias was robust and occurred when we placed the spatially overlapping stimuli at different sites within the RF (Fig. 1B).

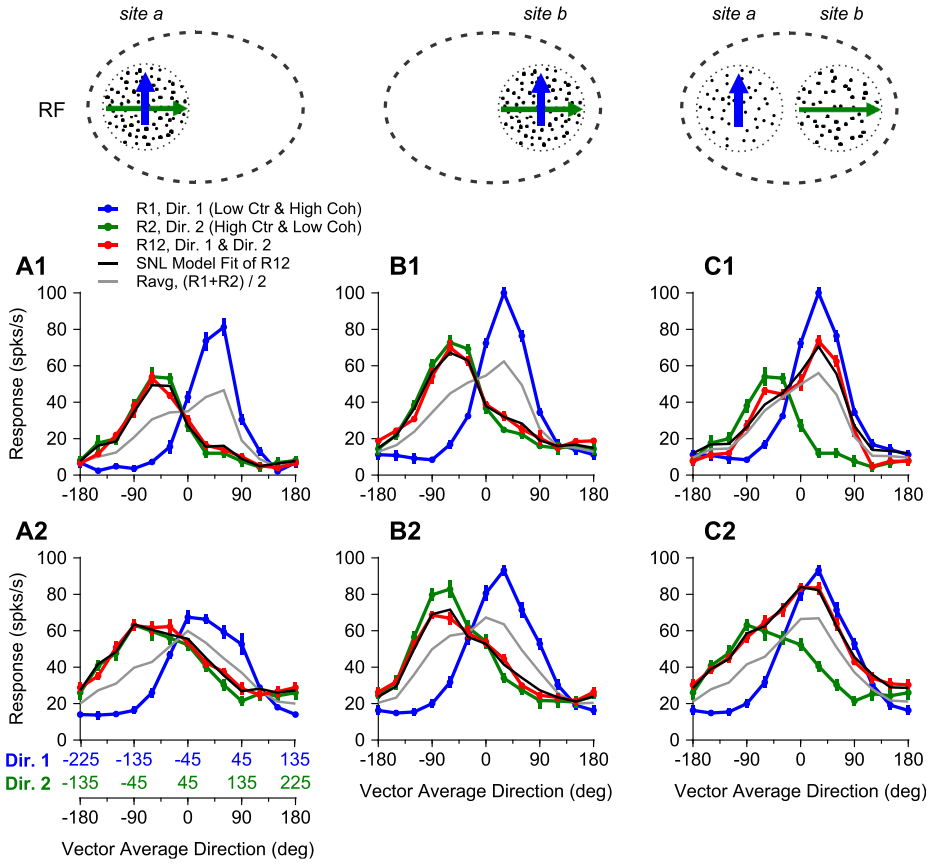

Figure 1. Visual stimuli and response tuning curves of 2 example neurons. ***A1, B1***: cartoon depictions of overlapping bi-directional stimuli. Visual stimuli could be placed on one side of the RF, *site a* (***A1***), or on the opposite side of the RF, *site b* (***B1***). ***C1****:* depiction of spatially separated bi-directional stimuli, with one random-dot patch placed on either side of the RF, with at least 1° of visual angle between them. Blue and green dots represent 2 separate achromatic random-dot patches. The component shown in blue (R1) moved at the clockwise side of the two components, and the dots had a low luminance contrast (36%) and a high motion coherence (100%). The green component (R2) moved at the counter-clockwise side, and the dots had a high luminance contrast (76%) and a low motion coherence (60%). ***A1-C1:*** Response tuning curves from an example neuron. The black curve is the summation plus non-linear interaction (SNL) model fit of the bi-directional response. ***A1:*** Response weight from the SNL model fit for the "low contrast & high coherence" component (w1) was 0.07, while the response weight for the "high contrast & low coherence" component (w2) was 0.91. ***C1:*** Response weights when spatially-separated were 0.69 (w1) and 0.58 (w2). ***A2-C2:*** Response tuning curves from another example neuron. ***A2:*** The response weights for this example neuron were 0.14 and 0.99 (w1 and w2 respectively). ***A2:*** Response weights when stimuli were spatially-separated were 0.56 (w1) and 0.75 (w2). The blue and green abscissas (***A2***) represent the unidirectional components for Dir. 1 and Dir. 2, respectively. The black abscissa corresponds to the vector averaged (VA) direction of the bi-directional stimuli. The blue and green axes are shifted 90° relative to each other.

Two overlapping visual stimuli could stimulate not only the RFs of single MT neurons, but also the RFs of single V1 neurons. The response bias toward the "high contrast & low coherence" component may be caused by the neural processes within area MT, or alternatively inherited from earlier visual areas such as V1. To determine the contribution of earlier visual areas to the response bias, we placed two stimulus components at different locations within the RF of a given MT neuron. The two stimulus components were separated by a gap of at least 1° of visual angle (illustrated in Fig. 1C). With this spatial arrangement, the RF of a single V1 neuron could only be stimulated by one of the two stimulus components, whereas the RF of an MT neuron could still be stimulated by both components. We found that the response tuning to the bi-directional stimuli changed drastically when the two stimulus components were spatially separated. MT responses elicited by the bi-directional stimuli no longer showed a bias toward the "high contrast & low coherence" component, but roughly followed a scaled average of the component responses (Fig. 1C).

Figure 2 shows the population averaged tuning curves across 70 MT neurons. The population averaged response elicited by the "low contrast & high coherence" component moving in the PD of each neuron, aligned to 0°, was significantly greater than that elicited by the "high contrast & low coherence" component moving in the PD (one-tailed paired t-test, $p = 4.1 \times 10^{-7}$). However, when the two stimuli were overlapping, the population response elicited by the bi-directional stimuli was strongly biased toward the weaker "high contrast & low coherence" component, regardless of the spatial location within the RF (Fig. 2A and 2B). The bias toward the "high contrast & low coherence" component at a given VA direction was in a manner of "higher-contrast-take-all". For example, at a VA direction of 45° where the "low contrast & high

coherence" component moved in the PD (0°) and the "high contrast & low coherence" component moved in 90°, the bi-directional response closely followed the much weaker response elicited by the "high contrast & low coherence" component. When the two stimulus components

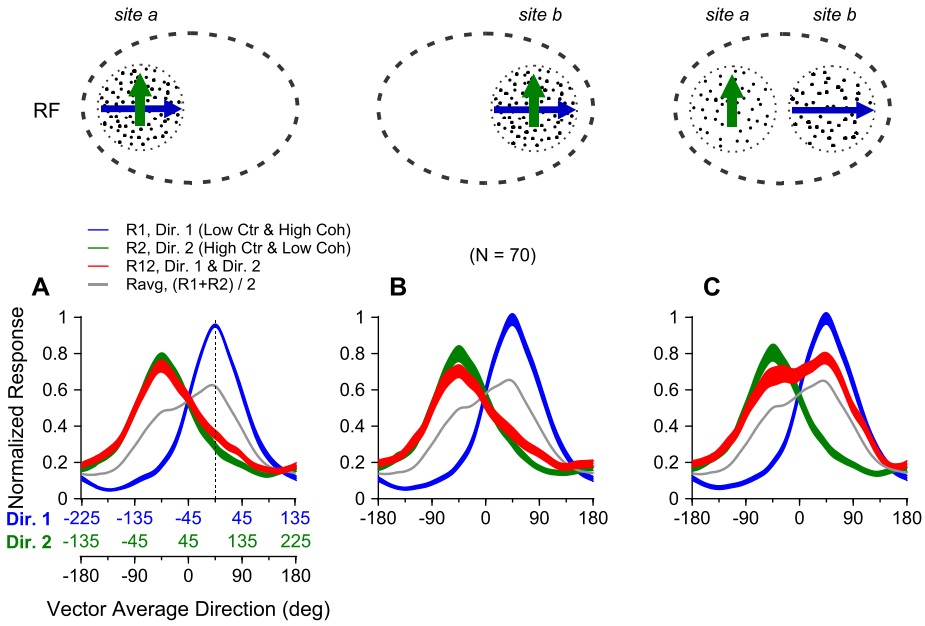

Figure 2. Population-averaged tuning curves from the bi-directional (red curves) and unidirectional component (blue and green curves) stimuli. The abscissa convention is the same as in **Figure 1.** The VA direction of 0° is aligned with the PD of each neuron before spline fitting and averaging. ***A:*** Averaged responses when stimuli were placed at *site a* within the RF. The vertical dashed line indicates an example of the response of a neuron when the "low contrast & high coherence" component moves in the PD of the neuron (blue curve), and the "high contrast & low coherence' component (green curve) moves 90° apart in the counter-clockwise direction (see results). ***B:*** Averaged responses when stimuli were placed at *site b* within the RF. ***C:*** Averaged responses when stimuli were spatially separated. The width of each curve indicates the standard error (SE).

were spatially separated within the RF, the strong bias toward the "high contrast & low coherence" component was abolished (Fig. 2C). The population response to the bi-directional stimuli now showed roughly equal weighting of the responses elicited by the individual stimulus components.

The SNL model provided an excellent fit of the MT responses elicited by the bi-directional stimuli, illustrated by the black curves in Figure 1. Across our population of neurons the model accounted for, on average, 83% of the response variance (see Methods). Figure 3 compares the response weights for the two stimulus components obtained from the SNL model fits. In the overlapping condition, the mean response weight w_2 for the "high contrast & low coherence" component was significantly greater than the weight w_1 for the "low contrast & high coherence" component (one-tailed paired t-test, $p = 1.9 \times 10^{-45}$ for *site a*, $p = 2.5 \times 10^{-28}$ for *site b*) (Fig. 3A). Nearly all data points, each representing the result from one neuron, were below the unity line. The mean response weight for the "high contrast & low coherence" component was 0.97 (std = 0.24), whereas the mean weight for the "low contrast & high coherence" component was 0.23 (std = 0.25), indicating a dominant effect of the "high contrast & low coherence" component in determining the neuronal response to the bi-directional stimuli.

When the two stimulus components were spatially separated within the RF, the response weights changed significantly, becoming symmetrically distributed relative to the unity line (Fig. 3B). The spread of response weights in the spatially-separated condition (Fig. 3B) is larger than that in the overlapping condition (Fig. 3A). The mean response weight for the "high contrast & low coherence" component decreased to 0.66 (std = 0.32), whereas the mean weight for the "low contrast & high coherence" component increased to 0.68 (std = 0.43). The mean weights for the two stimulus components were no longer different (paired t-test, $p = 0.8$) but both were significantly greater than 0.5 of response averaging (t-test, $p < 0.001$).

To quantify the response bias toward an individual stimulus component, we created a bias index (BI) :

$$BI = (w_2 - w_1)/(w_2 + w_1). \qquad (4)$$

A positive value of the index indicates a bias toward the "high contrast & low coherence" component, whereas a negative value indicates a bias toward the "low contrast & high coherence" component. Figure 3C shows how this bias index changes with the spatial arrangement of the visual stimuli. In the overlapping condition, the mean bias index is 0.73 (std = 0.23), which is significantly greater than 0 (one-tailed t-test, $p = 7.45 \times 10^{-35}$). In the spatially-separated condition, the mean bias index is -0.01 (std = 0.95), which is not significantly different from 0 (one-tailed t-test, p = 0.7). The mean bias index obtained in the overlapping condition is significantly greater than that in the spatially-separated condition (one-tailed paired t-test, $p = 4.7 \times 10^{-9}$), indicating a change of the response bias when the spatial arrangement of the visual stimuli is altered.

We previously found that the tuning curves of some MT neurons in response to overlapping bi-directional stimuli can show a directional "side-bias" toward one of the two direction components (Xiao and Huang, 2015). A subgroup of neurons prefers the stimulus component at the clockwise side of two motion directions, whereas another group of neurons prefers the component direction at the counter-clockwise side. These response biases can occur even when both stimulus components have the same luminance contrast and motion coherence. In the experiment shown in Figures 1-3, the "high contrast & low coherence" component always moved at the counter-clockwise side of the two component directions (illustrated in Fig. 2A, 2B).

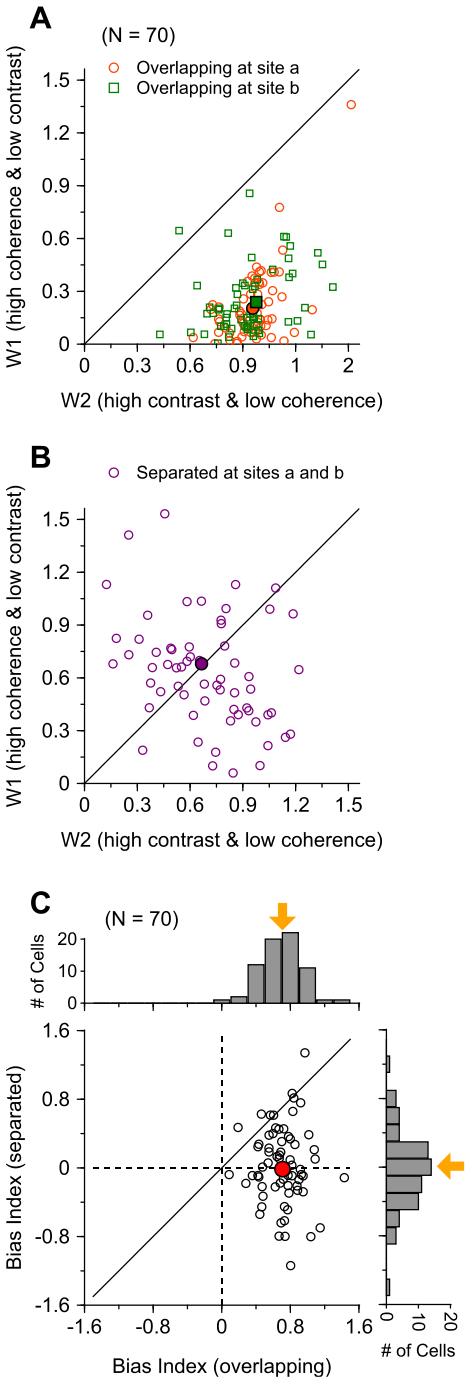

Figure 3. Comparing model-fitted response weights (*w1* and *w2*) for individual stimulus components. ***A:*** Response weights for each neuron from overlapping bi-directional stimulus conditions. Weights for the "high contrast & low coherence" (*w2*) component are on the abscissa, while weights for the "low contrast & high coherence" (*w1*) component are on the ordinate. Orange circles are response weights from overlapping stimuli at *site a*, and the green squares are response weights from overlapping stimuli at *site b*. Each dot is an individual neuron. Almost all of the neurons fall under the unity line. Solid circle and square represent the population average for each overlapping site respectively. ***B:*** Response weights for each neuron from the spatially separated stimuli. The abscissa and ordinate are the same as in ***A***. The weights are evenly distributed around the unity line, with the population average not significantly different from the unity line (see results). ***C:*** Comparing Bias Index from overlapping conditions (abscissa) vs. The Bias Index from the spatially separated conditions (ordinate) (See methods). The top histogram shows the distribution of the Bias Index for the overlapping conditions, and the histogram along the side shows the distribution of the Bias Index for the spatially separated conditions. The overlapping Bias Index for each neuron was greater than 0, while the spatially separated Bias Index was equally distributed around 0.

Could the strong bias that we found toward the “high contrast & low coherence” component in the overlapping condition be due to a biased neuron sample that happened to have a strong side-bias toward the direction component at the counter-clockwise side? To address this concern, we used different arrangements of the direction components.

In Figure 4, panels A and B show the averaged direction tuning curves of 15 MT neurons when the “high contrast & low coherence” component was placed at the counter-clockwise side, as in Figure 2. When the “high contrast & low coherence” component was placed at the clockwise side of the two component directions, the responses of the same 15 neurons elicited by the bi-directional stimuli still showed a strong bias toward the “high contrast & low coherence” component under the overlapping condition (Fig. 4C), and showed roughly equal weighting of the two stimulus components under the spatially-separated condition (Fig. 4D). Placing the “high contrast & low coherence” component at the clockwise or counter-clockwise side of the two component directions had no effect on the response bias, as measured by the bias index under the overlapping and spatially-separated conditions (Wilcoxon rank sum test, $p = 0.6$).

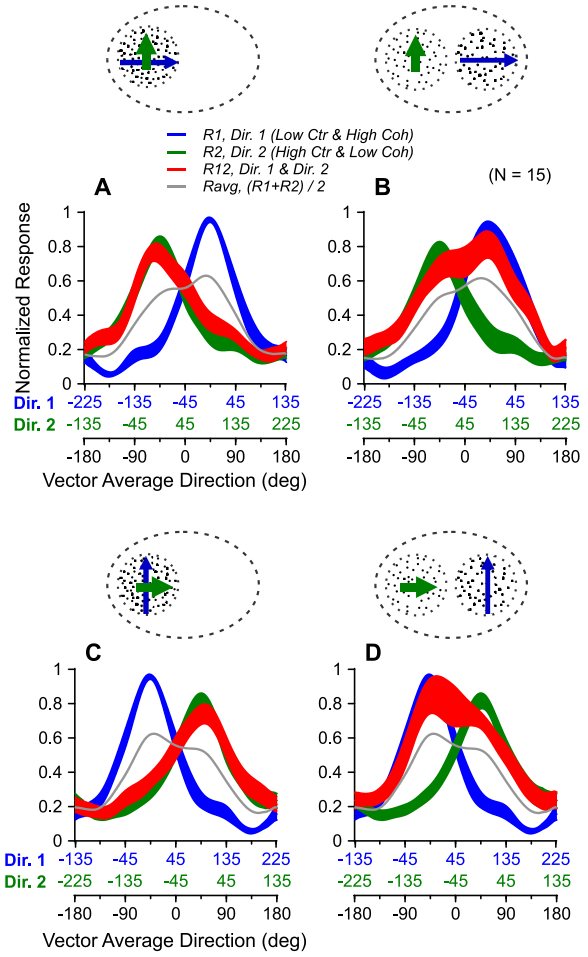

Figure 4. Switching clockwise/counter-clockwise orientation of individual stimulus components. ***A, C:*** Population-averaged response tuning curves from overlapping stimulus conditions. ***B, D:*** Population-averaged response tuning curves from spatially-separated conditions. ***A, B:*** Same orientation as the main paradigm, with the "low contrast & high coherence" (Dir. 1) component moving at the clockwise side and the "high contrast & low coherence" (Dir. 2) component moving at the counter-clockwise side. ***C, D:*** Orientation of stimulus components switched. The "low contrast & high coherence" component now moving at the counter-clockwise side, while the "high contrast & low coherence" component is now on the clockwise side. Notice the switch in the blue and green abscissas in ***A*** and ***C***. Bias toward the "high contrast & low coherence" component for the overlapping conditions was consistent regardless of component orientation (***A*** and ***C***). For the spatially-separated conditions, the equal weighting of both components was consistent when the orientation was switched (***D***).

We examined the timecourse of the neuronal responses under the overlapping and spatially-separated conditions. Figure 5 shows the PSTHs calculated using a bin width of 10 ms when either the "high contrast & low coherence" component or the "low contrast & high

coherence" component moved in the PD. When MT neurons first responded to the onset of the static stimuli, even before the onset of motion, the response to the bi-directional stimuli already closely followed the response elicited by the "high contrast & low coherence" component. This response bias was present regardless of whether the stimuli were overlapping or spatially separated.

When the stimuli were overlapping, the neuronal response to the bi-directional stimuli throughout the motion period continued to follow the response elicited by the "high contrast & low coherence" component, regardless of whether the it moved in the PD, eliciting a strong component response (Fig. 5A), or 90° away from the PD, which elicited a weak component response (Fig. 5B).

When the visual stimuli were spatially separated, the bi-directional response continued to follow the response elicited by the "high contrast & low coherence" component for about 50-70 ms, before MT neurons started to respond to the visual motion. During the initial 20-30 ms of the motion response, the response elicited by the bi-directional stimuli closely followed the response elicited by the stimulus component that moved in the PD. The response then started showing an effect of response normalization - the stimulus component moving in the non-PD slightly "pulled" down the response to the bi-directional stimuli (Fig 5C, D). Since the strong response bias towards the "high contrast & low coherence" component occurred at the very beginning of the stimulus onset, it is unlikely that the bias toward the "high contrast & low coherence" component when the stimuli were overlapping was due to attention.

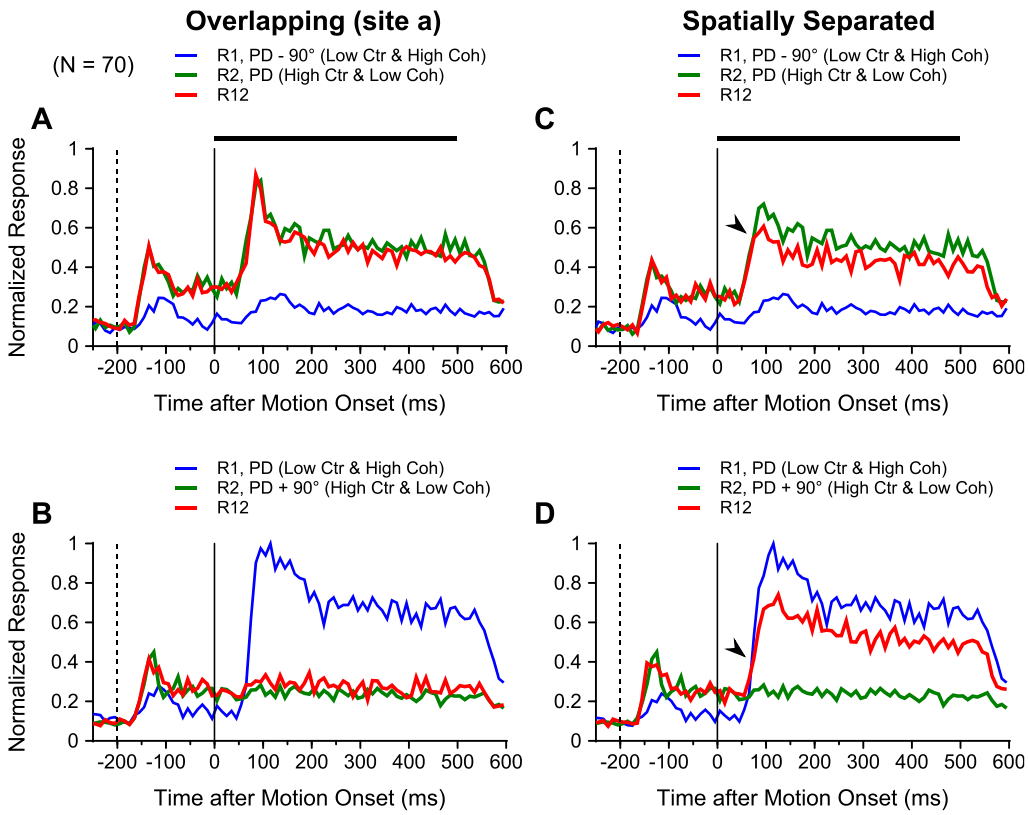

Figure 5. Population time course analysis for overlapping and spatially separated conditions. ***A*, *B*:** Time course of neuronal response when stimuli overlapping at site *a*. Both stimulus components turn on at -200ms (vertical dashed line), and remain motionless. Motion begins at 0ms and continues for 500ms (black horizontal line). ***A*:** Neuronal responses when "high contrast & low coherence" component moved in the PD, while the "low contrast & high coherence" component moved 90° apart on the clockwise side. The bi-directional response (red curve) follows the "high contrast & low coherence" component from the time that the stimulus is turned on, and continues to follow that component throughout the motion period. ***B*:** Neuronal responses when the "low contrast & high coherence" component moved in the PD, while the "high contrast & low coherence" component moved 90° apart on the counter-clockwise side. ***C*:** Neuronal response to spatially separated stimuli. Arrangement is the same as in ***A***. The bi-directional response again follows the "high contrast & low coherence" component closely when both stimuli turn on, and continues to follow the "high contrast & low coherence" component for the first ~75ms of the motion period, when normalization begins to "pull down" the bi-directional response towards the "low contrast & high coherence" component (arrow). A similar result can also be seen in ***D***.

Our results show that luminance contrast has a dominant effect on MT responses elicited by overlapping stimuli when luminance contrast and motion coherence compete with each other. To evaluate the generality of this finding, we conducted a second experiment using visual stimuli that differ in luminance contrast and motion speed. We have previously found that when two

overlapping random-dot patches moved in the same direction at different speeds, within a range of low to intermediate speeds, the responses of MT neurons elicited by the bi-speed stimuli is biased toward the faster speed component (X. Huang et al., unpublished data). Motivated by this finding, we used motion speed to compete with luminance contrast. As in the main experiment, the visual stimuli contained two random-dot patches moving in two directions separated by 90° and we varied the VA direction to measure the direction tuning curves. One stimulus component had a high luminance contrast of 77.5% and moved at a slower speed of 2.5°/s. The other stimulus component had a low luminance contrast of 37.5% and moved at a faster speed of 10°/s. Both stimulus components moved at 100% coherence and were either overlapping or spatially-separated within the RF of a given MT neuron as in the first experiment. We also measured the direction tuning curves when the two stimulus components both had high luminance contrast (77.5%) and moved at 2.5°/s and 10°/s, respectively, at 100% coherence.

We recorded from 13 MT neurons using these visual stimuli. Figure 6 shows the population averaged tuning curves. When both stimulus components had a high contrast, the population averaged response elicited by the faster (10°/s) stimulus component moving in the PD (0°) was greater than that elicited by the slower (2.5°/s) component moving in the PD (indicated by the purple and blue curves, respectively in Fig. 6A). When the two stimulus components were overlapping, the tuning curve elicited by the bi-directional stimuli (shown in red) is biased toward the faster stimulus component, more than what is predicted by the average of the component responses (shown in gray) (Fig. 6A). We fitted the direction tuning curves using the SNL model for each neuron (Eq. 1). The median response weight obtained by the model fit for the faster stimulus component (0.88) was significantly greater than the median weight (0.41) for the slower component (Wilcoxon signed rank test, $p = 7.3 \times 10^{-4}$). This result extended our

previous finding of the response bias toward the faster stimulus component for stimuli moving in the PD of neurons (unpublished results), to stimuli moving in different directions.

When the overlapping stimuli moving at different speeds had different luminance contrasts, the responses elicited by the bi-directional stimuli showed a strong bias toward the "high contrast & slower speed" component, even though the response to the "high contrast & slower speed" component alone was significantly weaker than the response to the "low contrast & faster speed" component at the PD (Fig. 6B). We found the same result when the two stimulus components overlapped at a different site within the RF (Fig. 6C). Under the overlapping condition, the median response weight for the "high contrast & slower speed" component was 0.81, which was significantly greater than the median weight (0.17) for the "low contrast & faster speed" component (Wilcoxon signed rank test, $p = 2.4 \times 10^{-4}$). Separating the two stimulus components within the RF abolished the bias toward the "high contrast & slower speed" component (Fig. 6D). As the spatial arrangement of the stimulus components changed from overlapping to spatially-separated, the median bias index (Eq. 4) decreased significantly from 0.65 to -0.08 (Wilcoxon signed rank test, $p = 0.0012$). These results confirmed that luminance contrast has a dominant effect on MT responses elicited by overlapping stimuli that are competing in more than one feature domain. Together, our results suggest that the impact of spatial arrangement on the competition between multiple visual stimuli within the RF may be general.

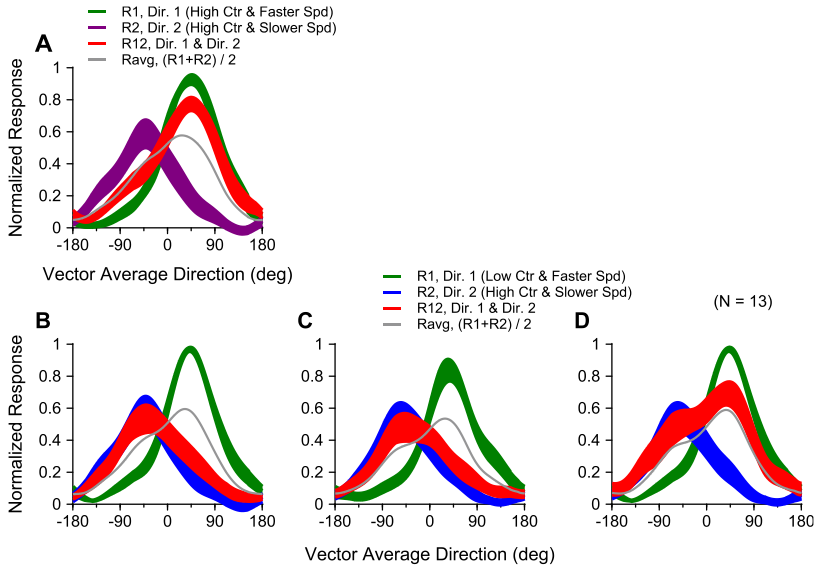

Figure 6. Population-Averaged tuning curves from speed and luminance contrast experiment. Both dot patches moved at 100% coherence. ***A:*** Both random-dot patches had a high luminance contrast. Only the motion speed was different between components. Dir. 1 (purple curve) component moved at a faster speed (10°/second) and the Dir. 2 (green curve) component moved at a slower speed (2.5°/second). Dots moved in different directions, separated by 90°. (***B, C, D***): One stimulus component (Dir. 1, green curve) had a low luminance contrast but moved at a faster speed. The other stimulus component (Dir. 2, blue curve) had a high luminance contrast and moved at a slower speed. ***B:*** Response tuning curves when stimuli were overlapping at *site a* in RF. ***C:*** Response tuning curves when stimuli were overlapping at *site b* in RF. ***D:*** Response tuning curves when stimuli were spatially separated. Bi-directional responses for both overlapping and spatially-separated conditions are consistent with motion coherence and luminance contrast experiment (Fig 1, 2).

Computer simulations

Our spatially separated visual stimuli fall inside the RFs of single MT neurons, whereas only one of the stimulus components fall inside the RFs of single V1 neurons. Hence, our spatially-separated visual stimuli can interact within the RFs of MT neurons but not V1 neurons. In contrast, the overlapping stimuli can interact within the RFs of both MT and V1 neurons. To explore the neural mechanisms underlying our physiological findings, we conducted computer simulations using a hierarchical feedforward model adapted from Simoncelli and Heeger (1998).

Simoncelli and Heeger's seminal model (1998) consists of two processing stages corresponding to areas V1 and MT. Each stage carries out a series of computations including spatiotemporal filtering, spatial pooling, rectification, and divisive normalization. At the V1 stage, simple cells receive input directly from the visual stimulus and complex cells pool inputs from rectified and divisively normalized responses of V1 simple cells. At the MT stage, MT neurons pool inputs from V1 complex cells, followed by rectification and divisive normalization (Simoncelli and Heeger, 1998).

We generated random-dot visual stimuli that are similar to those used in our physiology experiments and simulated the neuronal responses in areas MT and V1 to the visual stimuli. The visual stimuli and a simplified architecture of the model are illustrated in Figure 7. The diameter of each random-dot patch was 3°, extending 63 pixels. The RF sizes of model V1 and MT neurons, set by the sizes of the convolution filters, were 1.2° and 10° in diameter, respectively (see Methods for details). The population of model neurons in both V1 and MT stages completely tiled a sphere in the spatiotemoporal frequency domain, as in Simoncelli and Heeger's model (1998). The RFs of V1 and MT neuron populations covered a region of the visual field that was 17.3° x 17.3°. In the overlapping condition, the apertures of two random-dot patches overlapped within the RFs of single MT neurons (Fig. 7A). In the spatially-separated condition, the two random-dot patches were placed side by side, separated by a blank gap that was 1.2° wide, within the RFs of single MT neurons (Fig. 7B). In the overlapping condition, the V1 neurons whose RFs covered *site a* were activated by both stimulus components (Fig. 7A). In the spatially-separated condition, V1 neurons were activated by only one stimulus component, either at *site a* or *site b* (Fig. 7B).

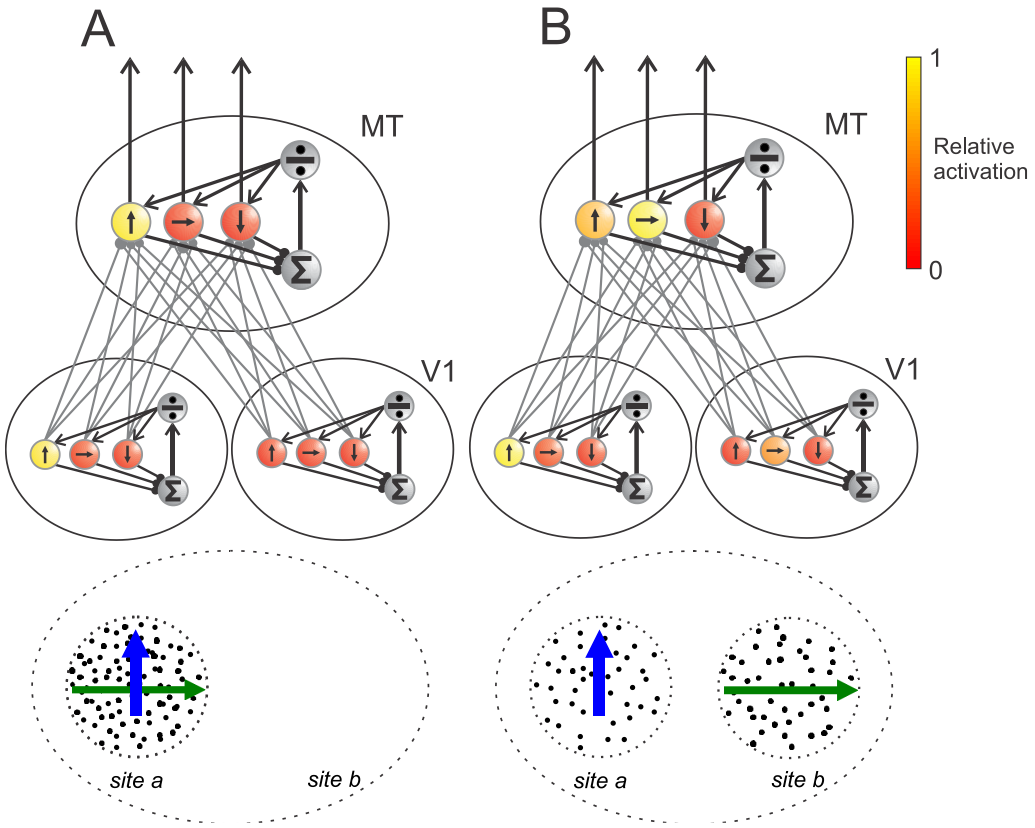

Figure 7. Simplified architecture for the feed-forward normalization model. The MT layer receives projections from multiple RFs in the V1 layer. Linear responses are divisively normalized by the sum of local activity at both cell layers. Blue arrows represents the "high contrast & low coherence" component, moving in the neuron's PD. The green arrows represent the "low contrast & high coherence" component, moving 90° away on the clockwise side. The activity in overlapping component condition (***A***) demonstrates how the "low contrast & high coherence" signal is eliminated in the V1 layer, while both signals are preserved in the projections to MT in the separated component condition (***B***).

We tuned the model parameters (see Methods) to match the experimentally measured contrast response functions of V1 and MT neurons (Sclar et al., 1990) and the coherence response function of MT neurons (Britten et al., 1998). The simulated contrast response functions of V1 and MT neurons fitted the experimental data almost perfectly and the simulated coherence response function of MT neurons also matched the data well (Fig. 8A-C). As far as we know, an experimentally measured coherence response function of V1 neurons has not been described previously. Our simulations show that V1 responses increased monotonically with the

coherence level of moving random-dot stimuli. Notably, the model V1 neurons had relatively higher firing rates in response to low coherence stimuli and more trial-to-trial variability in comparison with the model MT neurons (Fig. 8D).

The neuronal responses in area MT to our visual stimuli used in experiment 1, were well captured by the model simulations. Consistent with the experimental data (Fig. 2), the tuning curve of model MT neurons to the "low contrast & high coherence" component had a higher peak response than that of the "high contrast & low coherence" component (Fig. 9A, B). In the overlapping condition, the simulated MT response elicited by the bi-directional stimuli was strongly biased toward the weaker "high contrast & low coherence" component (Fig. 9A), as found in the neural data. The model also captured the change of MT response tuning with the separated spatial arrangement of the visual stimuli. In the spatially-separated condition, the MT tuning curve elicited by the bi-directional stimuli was no longer dominated by the "high contrast & low coherence" component (Fig. 9B).

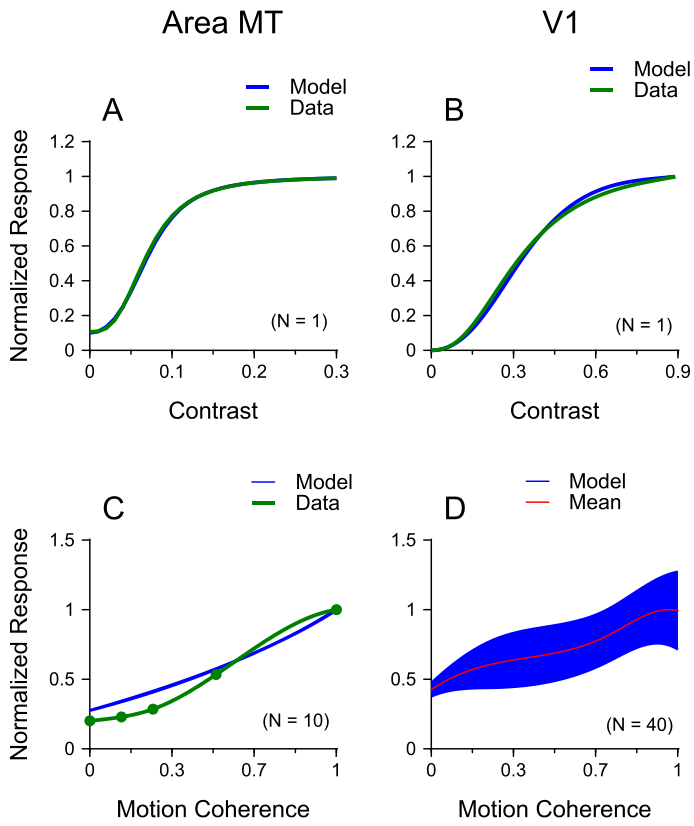

Figure 8. (***A, B***) Fitted contrast response functions to sinusoidal gratings for simulated MT (***A***) and V1 (***B***) cells plotted beside experimentally derived equations from Sclar, Maunsell, and Lennie (1990). Simulated cell responses are deterministic when using sinusoidal grating stimuli. ***C:*** Fitted coherence response to high contrast random dots for simulated MT neuron plotted beside spline-fit data from Britten (1998). ***D***: Coherence response to high contrast random dots for a simulated V1 complex cell.

At the V1 stage of the model, the tuning curves of V1 complex cells showed a slightly higher peak response to the "high contrast & low coherence" component than to the "low contrast & high coherence" component (Fig. 9C). In the overlapping condition, the simulated V1 response elicited by the bi-directional stimuli was biased toward the "high contrast & low coherence" component (Fig. 9C), to the extent similar to that found for the model MT neuron (Fig. 9A), as measured by the weights for the component responses using the SNL model fits. The bias index (Eq. 3) for the V1 model neuron was 0.90 and that for the MT model neuron was 0.93. Both showed a strong bias toward the "high contrast & low coherence" component. These

simulation results suggest that the strong response bias toward the “high contrast & low coherence” component found in area MT is inherited from V1.

In the spatially-separated condition, the V1 response elicited by the bi-directional stimuli was the same as that elicited by the single stimulus component that was placed within the RFs of V1 neurons (Fig. 9D, E). Although the V1 peak response elicited by the “high contrast & low coherence” component at *site a* was slightly stronger than that elicited by the “low contrast & high coherence” component at *site b*, the MT response elicited by the bi-directional stimuli was skewed toward the “low contrast & high coherence” component (Fig. 9B). These simulation results suggest that the MT response elicited by the bi-directional stimuli in the spatially-separated condition (Fig. 9B) is mainly the result of feature competition within area MT.

The response tuning curves of single MT neurons measured by varying the VA direction of the bi-directional stimuli can be mapped to the responses of a population of neurons that have different PDs elicited by the bi-directional stimuli moving in a given VA direction. Figure 7 summarizes the changes of the response distributions across neuron populations at V1 and MT stages under the overlapping and spatially separated conditions. These results reveal the importance of neural processing at different stages of the visual hierarchy on determining how multiple visual stimuli compete with each other within neurons’ RFs.

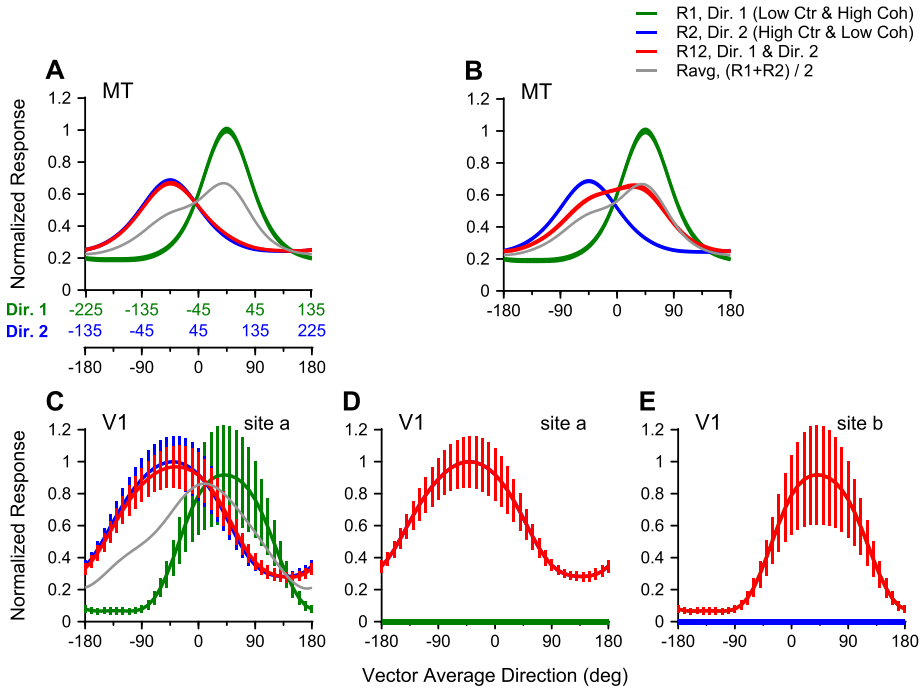

Figure 9. (***A, B***) Simulated MT neuron tuning curves show trends similar to recordings for stimulus components presented in the overlapping (***A***) and spatially-separated (***B***) conditions. SNL fitted weights were w_1=0.09 and w_2=0.94 for the overlapping condition and w_1=0.36 and w_2=0.64 for the separated condition. ***C, D, E:*** Simulated V1 complex cells show a greater peak response to the "high contrast & low coherence" stimulus component and an increased baseline response. The simulated V1 complex cell response to the overlapping stimulus condition (***C***) closely follows the "high contrast & low coherence" component, nearly disregarding the "low contrast & high coherence" component. Simulations of V1 neuron responses in the spatially separated condition (***D, E***) show an identical response to the component that is at the respective site centered on the cell's RF, and no response to the component centered at the opposite site. This confirms that the simulated V1 neurons only receive input from one component in the spatially separated condition.

Discussion

We have shown that how MT neurons represent multiple moving stimuli competing in more than one feature domain depends on the spatial arrangement of the visual stimuli. When two visual stimuli were spatially overlapping, neuronal responses in area MT were dominated by the stimulus component that had higher luminance contrast. This occurred even when the peak response of the direction tuning curve in MT elicited by the higher-contrast component alone

was significantly weaker than the response elicited by a competing stimulus that had lower contrast but moved at either higher motion coherence or at a faster speed. When two visual stimuli were spatially separated within the RFs of MT neurons, the strong bias toward the higher-contrast component was abolished and MT responses approximated a scaled average of the responses elicited by the individual stimulus components used in this study. The results obtained using the overlapping stimuli cannot be explained by a descriptive model of normalization based on competition between the responses of populations of MT neurons elicited by the individual stimulus components. Analysis of the response timecourse and computer simulations using a V1-MT model suggest that the contrast dominance found with overlapping stimuli is due to normalization occurring at an input stage being fed into MT and MT neurons cannot overturn the contrast dominance based on their own feature selectivity.

Consideration of the effect of attention

Attention can bias neuronal responses elicited by multiple stimuli in the RF in favor of the attended stimulus (Reynolds et al. 1999; Li and Basso 2005; Treue and Maunsell 1996; Ferrera and Lisberger 1997; Treue and Martinez-Trujillo 1999; Recanzone and Wurtz 2000; Wannig et al., 2007; Lee and Maunsell, 2010). Could the strong response bias toward the higher-contrast component found with the overlapping stimuli be due to an attentional bias? Although in this study the animals performed a simple fixation task without the need to engage goal-directed (i.e. endogenous) attention. Could the higher-contrast component capture stimulus-driven (i.e. exogeneous) attention (Corbetta and Shulman, 2002) and bias the neuronal response toward the higher-contrast component? Several considerations argue against this possibility.

While an abrupt stimulus onset captures attention (Yantis and Jonides, 1984), a visual stimulus that is brighter than other distractors does not automatically capture attention (Jonides and Yantis, 1988). The two stimulus components of our overlapping stimuli were turned on at the same time. The stimulus onset may automatically draw attention toward the spatial location of the overlapping stimuli, but it is unlikely to draw attention toward only the higher-contrast component alone. In fact, the simultaneous onset of multiple visual stimuli makes the discrimination of a target stimulus more difficult than when the target is turned on by itself (Huang et al., 2005; Huang and Paradiso, 2005). Furthermore, the stimulus-driven attention occurs with a time delay (Nakayama and Mackeben, 1989) and the effect on neuronal responses in MT is transient, lasting for about 70 ms (Busse, Katzner and Treue, 2008). In contrast, we found that the response bias toward the higher-contrast component is present in the very beginning of the neuronal responses following the onset of the static stimuli, and the bias is persistent throughout the whole 500-ms motion period (Fig. 5). In addition, Wannig and colleagues (2007) have shown that attention directed to one of two overlapping surfaces can alter the responses of MT neurons. However, the effect of surface-based attention led to a firing rate chang of about 20% in MT between conditions when attention was directed to two different surfaces (Wannig et al., 2007). Even if, for some reason, the animals were consistently attending to the higher-contrast component throughout the experiment, the effect of surface-based attention would be insufficient to account for the nearly complete dominance by the higher-contrast component in our results.

Consideration of motion coherence

When two stimulus components overlap, the random dots from each component constitute only half of the total dots of the two moving surfaces. Could the strong response bias toward the "high contrast & low coherence" component be due to a reduction of the motion coherence of the "low contrast & high coherence" component when the stimuli overlapped? We think this is an unlikely explanation because overlapping both stimuli reduces the percentage of the signal dots relative to the total number of dots in the visual stimuli. This is true not just for the "low contrast & high coherence" component, but also for the "high contrast & low coherence" component. In addition, our stimuli moved in two directions separated by 90º. Human observers can reliably segregate the two stimulus components at this angle separation and the low contrast & high coherence component still appears to move coherently. Overlapping does not change the relative coherence levels of the two stimulus components and does not appear to interfere with the perceptual coherence of the stimulus components. When overlapping random-dot stimuli have the same luminance contrast but move at different motion coherence, MT responses to both stimulus components are biased toward the high coherence component (Xiao et al., 2014).

It is also unlikely that the contrast dominance was due to the noise dots in the high contrast & low coherence component interfering with the overlapping low contrast & high coherence component. We controlled the coherence of a stimulus component by randomly assigning a percentage of dots of that stimulus component as noise or signal dots at every monitor frame (Newsome and Pare, 1988). A signal (or noise) dot may turn into a noise (or signal) dot of that stimulus component at the next monitor frame. At each monitor frame, the stimulus component loses some signal dots to noise dots, and gains the same number of signal

dots from noise dots. As a result, the noise dots and signal dots of the same stimulus component are bound together. Perceptually, it is difficult to segregate the noise dots from the signal dots of the same stimulus component. Because of this bounding, the noise dots of the high contrast & low coherence component are not an independent entity and do not appear to interfere with the motion coherence of the low contrast & high coherence component perceptually. Our stimuli are different from the scenario when two stimulus components both move coherently in different directions and random dots moving at 0% coherence, bound with neither stimulus component, are added and overlap with the two stimulus components. In the latter scenario, the independent noise dots may interfere with both stimulus components.

Finally, our finding that the strong bias toward the high-contrast component also occurred when both stimulus components moved at 100% coherence at different speeds also suggests that the contrast dominance is not due to the confounds of motion coherence when stimuli are overlapping.

Mechanisms underlying the neural representation of multiple visual stimuli

The primate visual system is hierarchically organized (Felleman and Van Essen, 1991). The response properties of neurons in extrastriate visual area are shaped by inputs fed into the area, as well as by intra-areal and feedback processes. To understand the mechanisms underlying neural encoding of visual stimuli, it is important to determine how these processes contribute to the RF properties in a given visual area. However, it is often difficult to disentangle the contribution of feedforward inputs from intra-areal processes. We have previously shown that, in

response to overlapping stimuli, MT neurons show a bias toward the stimulus component that has a higher signal strength, defined by either luminance contrast or motion coherence. The response bias can be described by a model of divisive normalization (Xiao et al., 2014). Because neurons in V1 also show a bias toward the stimulus component that has a higher contrast (Busse et al. 2009; MacEvoy et al. 2009) and divisive normalization may occur in both V1 and MT (Simoncelli and Heeger, 1998; Heuer and Britten, 2002), it was unclear how the feedforward input from V1 and normalization within MT contributed to the response bias found in MT (Xiao et al., 2014).

In this study, we are able to differentiate the impacts of feedforward input and neural processing within MT on the response properties of MT neurons by using two visual stimuli that compete in more than one feature domain, and by manipulating the spatial arrangement of the visual stimuli. Neurons in V1 may respond more strongly to the individual "high contrast & low coherence" component than to the "low contrast & high coherence" component used in our experiment due to V1 neurons' sensitivities to contrast and coherence. When two stimuli overlap, the responses of V1 neurons elicited by both stimulus components may show a strong bias toward the "high contrast & low coherence" component due to divisive normalization in V1 (Fig 9C). The response input to MT from both stimulus components has already been mixed and is strongly biased toward the "high contrast & low coherence" component. MT neurons are no longer able to remix the two stimulus components according to their own sensitivities to contrast and coherence. In other words, MT neurons inherit the response bias toward the high contrast component from their input.

When two visual stimuli are sufficiently separated, MT neurons receive inputs from two different pools of V1 neurons and each neuron pool responds to only one stimulus component (Fig. 7B). The neuronal responses elicited by the two stimulus components remain separated in V1. MT neurons mix the responses elicited by the two stimulus components via spatial and directional pooling and divisive normalization within MT. The result of mixing reflects the sensitivities of MT neurons to different stimulus features. Under this circumstance, we found that MT response weight for a stimulus component is stronger when that stimulus component elicits a stronger population neural response in MT (Fig. 2). Given the parameters of our visual stimuli, the population MT responses to the "high contrast & low coherence" component alone and the "low contrast & high coherence" component alone are roughly the same, on average, MT neurons weight the two stimulus components equally.

Here we considered only the input directly from direction selective neurons in V1 to MT (Maunsell and Van Essen, 1983; Movshon and Newsome, 1996; Simoncelli and Heeger, 1998; Rust et al., 2006), but not the indirect pathway from cortical areas V2 and V3 to MT (Felleman and Van Essen, 1991; Ungerleider and Desimone, 1986; Maunsell and Van Essen, 1983) since the indirect input to MT contributes mainly to the processing of binocular disparity and stimuli moving at fast speeds (Ponce et al., 2008; Ponce et al., 2011).

Implications on normalization and neural encoding of multiple stimuli

Our finding that the response weights for multiple competing stimuli depend on the spatial arrangement of the visual stimuli provides a new perspective on the well-established normalization model (Carandini and Heeger, 2011). Previous work shows that neuronal

responses elicited by multiple stimuli in many brain areas can be described by response normalization (Morgan et al., 2008; Ohshiro et al., 2011; Busse et al., 2009; Xiao et al., 2014; Bao and Tsao, 2018), which is usually represented in a form $R = \frac{S_1^n}{f(S_1,\ S_2) + \sigma} \cdot R_1 + \frac{S_2^n}{f(S_1,\ S_2) + \sigma} \cdot R_2$. In the equation, R is the neuronal response elicited by both stimulus 1 and 2; R_1 and R_2 are the responses elicited by stimulus 1 and 2 alone, respectively; S_1 and S_2 represent the signal strengths of stimulus 1 and 2 respectively, and n and σ are model parameters. The signal strength is in the numerator of the equation, normalized by the combined signal strength of all stimulus components. The equation predicts that the response weight for a stimulus component increases with its signal strength, but does not differentiate the spatial arrangement of the visual stimuli. If S is a measure of signal strength in terms of the physical property of a visual stimulus, e.g. luminance contrast or speed, the equation predicts no difference in terms of the response weights when the spatial arrangement of the visual stimuli is changed, which is inconsistent with our results. Because the brain has no direct access to the physical property of a visual stimulus and has to make an inference based on the neural responses, it is reasonable to assume that the signal strength is represented by the summed activity of a population of neurons. We made a surprising new finding that, under some circumstances, the neuronal response to multiple stimuli in a given brain area cannot be predicted by the population neural responses elicited by the individual stimulus components in the same brain area. One must consider the neural computations occurring along the hierarchical visual pathway.

Majaj, Carandini and Movshon (2007) showed that pattern-direction selective neurons in MT characterized by overlapping drifting gratings (i.e. plaid) do not integrate the directions of the component gratings when they were spatially separated within the RF, suggesting that the

computations underlying pattern-direction selectivity in MT is local. Different from the plaid, the overlapping random-dot stimuli used in our study elicit the percept of the individual stimulus components moving transparently. We showed that changing the spatial arrangement of visual stimuli can have a substantial impact not only on motion integration but also the competition between multiple stimuli. Our results revealed that luminance contrast has a dominant effect in determining stimulus competition within a local spatial region, when multiple stimuli differ in more than one feature domain. When visual stimuli are spatially separated, the effect of contrast is substantially reduced.

A seminal computational model involving MT neurons pooling inputs from V1 and divisive normalization occurring in both V1 and MT stages has been successful in explaining a range of experimental results of MT responses, in particular, the pattern- and component-direction selectivity of MT neurons to a moving plaid (Simoncelli and Heeger, 1998; Rust et al., 2006). The model in its original form does not specify how features are spatially integrated and it does not differentiate spatially overlapping and separated stimuli (cf. Majaj et al., 2007). In this study, we adapted this model to simulate both spatially overlapping and separated conditions and showed that the framework of this model can explain our neurophysiological finding that stimulus competition can be altered by the spatial arrangement of visual stimuli. Also using this model, Busse, Wade and Carandini (2009) previously demonstrated the impact of response normalization in V1 on neuronal response in MT. They showed that, by making the contrasts of two drifting gratings of a plaid to be unequal, the response of a model MT neuron changed from representing the pattern motion of the plaid to mostly representing the higher-contrast grating component, likely due to contrast normalization in V1 (Busse et al., 2009). However, the MT

response elicited by the higher-contrast grating alone would also be greater than that elicited by the lower-contrast grating. The model-predicted response bias toward the higher-contrast component in MT, akin to our experimental result obtained using random-dot stimuli with unequal contrasts (Xiao et al., 2014), may also be contributed by response normalization within MT. Our current study provides unequivocal new evidence on how neuronal responses in MT are shaped by the hierarchical network. By using two stimuli competing in more than one feature domain, we demonstrated neurophysiologically and computationally the substantial impact of stimulus competition in the input stage (presumably in V1) on the neuronal responses in MT and how that impact changes with the spatial arrangement of the visual stimuli. We also made a surprising finding that under certain circumstances, MT neurons are unable to overturn the response bias inherited from their input according to their own feature selectivity.

References:

Albright, T. D. and R. Desimone (1987). "Local precision of visuotopic organization in the middle temporal area (MT) of the macaque." Exp Brain Res 65(3): 582-592.

Bao, P. and D. Y. Tsao (2018). "Representation of multiple objects in macaque category-selective areas." Nat Commun 9(1): 1774.

Born, R. T. and D. C. Bradley (2005). "Structure and function of visual area MT." Annu Rev Neurosci 28: 157-189.

Braddick, O. (1993). "Segmentation versus integration in visual motion processing." Trends Neurosci 16(7): 263-268.

Britten KH. (2003). The middle temporal area: motion processing and the link to perception In *The visual neurosciences*, ed. LM Chalupa, JS Werner, pp. 1203-16. London: The MIT press

Britten, K. H. and H. W. Heuer (1999). "Spatial summation in the receptive fields of MT neurons." J Neurosci 19(12): 5074-5084.

Britten, K. H. and W. T. Newsome (1998). "Tuning bandwidths for near-threshold stimuli in area MT." J Neurophysiol 80(2): 762-770.

Britten, K. H., M. N. Shadlen, W. T. Newsome and J. A. Movshon (1992). "The analysis of visual motion: a comparison of neuronal and psychophysical performance." J Neurosci 12(12): 4745-4765.

Britten, K. H., M. N. Shadlen, W. T. Newsome and J. A. Movshon (1993). "Responses of neurons in macaque MT to stochastic motion signals." Vis Neurosci 10(6): 1157-1169.

Busse, L., A. R. Wade and M. Carandini (2009). "Representation of concurrent stimuli by population activity in visual cortex." Neuron 64(6): 931-942.

Carandini, M. and D. J. Heeger (2011). "Normalization as a canonical neural computation." Nat Rev Neurosci 13(1): 51-62.

Carandini, M., D. J. Heeger and J. A. Movshon (1997). "Linearity and normalization in simple cells of the macaque primary visual cortex." J Neurosci 17(21): 8621-8644.

Ferrera, V. P. and S. G. Lisberger (1997). "Neuronal responses in visual areas MT and MST during smooth pursuit target selection." J Neurophysiol 78(3): 1433-1446.

Gattass, R. and C. G. Gross (1981). "Visual topography of striate projection zone (MT) in posterior superior temporal sulcus of the macaque." J Neurophysiol 46(3): 621-638.

Heuer, H. W. and K. H. Britten (2002). "Contrast dependence of response normalization in area MT of the rhesus macaque." J Neurophysiol 88(6): 3398-3408.

Heuer, H. W. and K. H. Britten (2002). "Contrast dependence of response normalization in area MT of the rhesus macaque." J Neurophysiol 88(6): 3398-3408.

Lee, J. and J. H. Maunsell (2010). "Attentional modulation of MT neurons with single or multiple stimuli in their receptive fields." J Neurosci 30(8): 3058-3066.

Li, X. and M. A. Basso (2005). "Competitive stimulus interactions within single response fields of superior colliculus neurons." J Neurosci 25(49): 11357-11373.

MacEvoy, S. P., T. R. Tucker and D. Fitzpatrick (2009). "A precise form of divisive suppression supports population coding in the primary visual cortex." Nat Neurosci 12(5): 637-645.

Majaj, N. J., M. Carandini and J. A. Movshon (2007). "Motion integration by neurons in macaque MT is local, not global." J Neurosci 27(2): 366-370.

Maunsell, J. H. and D. C. van Essen (1983). "The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey." J Neurosci 3(12): 2563-2586.

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 –378.

Movshon, J. A. and W. T. Newsome (1996). "Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys." J Neurosci 16(23): 7733-7741.

Movshon, J. A. and W. T. Newsome (1996). "Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys." J Neurosci 16(23): 7733-7741.

Newsome, W. T. and E. B. Pare (1988). "A selective impairment of motion perception following lesions of the middle temporal visual area (MT)." J Neurosci 8(6): 2201-2211.

Ohshiro, T., D. E. Angelaki and G. C. DeAngelis (2011). "A normalization model of multisensory integration." Nat Neurosci 14(6): 775-782.

Park, W. J. and Tadin, D. (2019). Motion Perception. In Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, J. T. Wixted (Ed.). doi:10.1002/9781119170174.epcn210

Ponce, C. R., J. N. Hunter, C. C. Pack, S. G. Lomber and R. T. Born (2011). "Contributions of indirect pathways to visual response properties in macaque middle temporal area MT." J Neurosci 31(10): 3894-3903.

Ponce, C. R., S. G. Lomber and R. T. Born (2008). "Integrating motion and depth via parallel pathways." Nat Neurosci 11(2): 216-223.

Qian, N. and R. A. Andersen (1994). "Transparent motion perception as detection of unbalanced motion signals. II. Physiology." J Neurosci 14(12): 7367-7380.

Qian, N., R. A. Andersen and E. H. Adelson (1994). "Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics." J Neurosci 14(12): 7357-7366.

Qian, N., R. A. Andersen and E. H. Adelson (1994). "Transparent motion perception as detection of unbalanced motion signals. III. Modeling." J Neurosci 14(12): 7381-7392.

Recanzone, G. H. and R. H. Wurtz (2000). "Effects of attention on MT and MST neuronal activity during pursuit initiation." J Neurophysiol 83(2): 777-790.

Recanzone, G. H., R. H. Wurtz and U. Schwarz (1997). "Responses of MT and MST neurons to one and two moving objects in the receptive field." J Neurophysiol 78(6): 2904-2915.

Reynolds, J. H., L. Chelazzi and R. Desimone (1999). "Competitive mechanisms subserve attention in macaque areas V2 and V4." J Neurosci 19(5): 1736-1753.

Rust, N. C., V. Mante, E. P. Simoncelli and J. A. Movshon (2006). "How MT cells analyze the motion of visual patterns." Nat Neurosci 9(11): 1421-1431.

Sclar, G., J. H. Maunsell and P. Lennie (1990). "Coding of image contrast in central visual pathways of the macaque monkey." Vision Res 30(1): 1-10.

Simoncelli, E. P. and D. J. Heeger (1998). "A model of neuronal responses in visual area MT." Vision Res 38(5): 743-761.

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

Treue, S. and J. C. Martinez Trujillo (1999). "Feature-based attention influences motion processing gain in macaque visual cortex." Nature 399(6736): 575-579.

Treue, S. and J. H. Maunsell (1996). "Attentional modulation of visual motion processing in cortical areas MT and MST." Nature 382(6591): 539-541.

Ungerleider, L. G. and R. Desimone (1986). "Projections to the superior temporal sulcus from the central and peripheral field representations of V1 and V2." J Comp Neurol 248(2): 147-163.

Wannig, A., V. Rodriguez and W. A. Freiwald (2007). "Attention to surfaces modulates motion processing in extrastriate area MT." Neuron 54(4): 639-651.

Xiao, J. and X. Huang (2015). "Distributed and Dynamic Neural Encoding of Multiple Motion Directions of Transparently Moving Stimuli in Cortical Area MT." J Neurosci 35(49): 16180-16198.

Xiao, J., Y. Q. Niu, S. Wiesner and X. Huang (2014). "Normalization of neuronal responses in cortical area MT across signal strengths and motion directions." J Neurophysiol 112(6): 1291-1306.