Neural mechanisms of visual segmentation using motion and depth cues in cortical area MT

By

Venkata Lakshmi Anjani Sreeprada Chakrala

A dissertation submitted in partial fulfillment of

the requirement for the degree of

Doctor of Philosophy

(Neuroscience)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 12/15/2022

This dissertation is approved by the following members of the Final Oral Committee:

Xin Huang, Associate Professor, Neuroscience

Ari Rosenberg, Associate Professor, Neuroscience

Bradley Postle, Professor, Psychology and Psychiatry

Yuri Saalamann, Associate Professor, Psychology

Meyer Jackson, Professor, Neuroscience

Acknowledgements

This dissertation was made possible by the contribution and support of many people.

First, I would like to thank my advisor, Dr. Xin Huang. Without your guidance and timely support, the studies in this dissertation could not have been completed. Through your mentorship, I learned how to be a well-rounded scientist from handling the nitty gritty of experiments to critically evaluating findings to asking important questions. I am grateful for your patience and kindness, and you have taught me by example how to be a compassionate mentor. Your dedication has provided me with a valuable skill set to carry forward and memories to cherish.

I am sincerely thankful to the members of my thesis committee – Dr. Ari Rosenberg, Dr. Yuri Saalmann, Dr. Brad Postle, and Dr. Meyer Jackson. Your guidance and constructive feedback over the years helped me improve the experimental design and analyses of the studies that contributed to this dissertation.

I am extremely grateful to the lab members and alumni. Dr. Steven Wiesner and Dr. Jianbo Xiao, you were amazing mentors who helped me develop the skills I needed to complete this dissertation. You made my transition into graduate school less challenging. I would like to thank Bryce Arseneau for his excellent help in training animals for the studies in this dissertation. Bikalpa Ghimire and Ying Cao, thank you for being the best co-workers. I will always remember your valuable inputs on the projects, and your camaraderie that made my time in lab more productive and enjoyable.

I would also like to thank the Neuroscience Training Program, the McPherson Eye Research Institute, the Department of Neuroscience, and the Wisconsin National Primate Research Center. These organizations provided me with valuable training, education, and financial support.

Importantly, I would like to thank my family- Amma, Nanna, and Pranav for always being there for me even if we are continents apart. Thank you for always believing in me and loving me unconditionally. I am immensely grateful to Siyang for his love and support all along. Thank you for all the feedback on my writing and presentations, you taught me how to be an effective communicator. I would also like to thank my friends Anil, Srihita, Prathysha, Ragavi, Mounika and Manasa for their support and lifelong friendship that keep me grounded.

Table of Contents

Acknowledgements i

Table of Contents ii

Abstract iii

Chapter 1: General Introduction 1

Chapter 2: Neural representation of overlapping motion surfaces separated in depth: role of binocular disparity and selective attention 11

Chapter 3: Neural representation of multiple spatially separated speeds in cortical area MT 73

Chapter 4: Discussion and summary 110

Abstract

Our dynamic world requires us to discern objects in complex environments to effectively interact with our surroundings. Our visual system has the remarkable ability to seamlessly segment objects from each other, unmatched by current machine vision systems, that enables us to understand visual scenes and guides actions. In natural vision, visual motion combined with other features like depth and cognitive abilities like attention is important to segment objects. Visual motion processing is widely studied at both perceptual and physiological levels. However, the neural mechanisms underlying the segmentation of multiple objects aided by motion combined with other features like depth and spatial location are not fully understood. The goal of this dissertation is to elucidate the neural mechanisms underlying the segmentation of multiple objects using motion direction, speed, depth (binocular disparity), and spatial location. Toward this, in chapter 2, I determined the neural representation of multiple objects, that differed in depth and motion direction and investigated the effect of selective attention on this representation. We designed a novel motion discrimination task, in which macaque monkeys selectively attended to one of two moving surfaces separated in depth. The task was paired with electrophysiological recordings from the middle temporal area (MT), known for its selectivity to visual motion and binocular disparity. Our results suggest that to represent multiple surfaces with different depths and motion direction, MT neurons leverage their preference to the binocular disparity of the constituent surfaces, so that the information about the two motion directions is distributed over neuronal subgroups defined by their disparity preferences (near and far-preferred neurons). Selective attention to one disparity enhanced the representation of the attended surface and its motion direction. In chapter 3, I investigated how MT neurons encoded two spatially separated stimuli moving at different speeds. Akin to chapter 2, we found that the MT neurons' spatial and speed preferences to single motion

were preserved in their responses to two spatially separated speeds, which might help in their segmentation. Together, our results elucidate prevailing neural strategies of encoding multiple visual stimuli, and how the visual system exploits spatial cues either two-dimensional or three-dimensional, together with motion cues to segment multiple objects.

Chapter 1: General Introduction

Vision is one of the most important senses of primates. Since the physical world is full of objects, often crowded together, visual systems have evolved to seamlessly segment one object from another and from the general surroundings. This segmentation is fundamental to visual perception and visually guided behavior. Primate visual systems rely on a variety of features or cues (e.g., contrast, shape, color, motion, or depth) to achieve visual segmentation. Some cues might be more powerful than others under different circumstances. For example, visual motion and depth might be more compelling cues than color to break camouflage and thus enable the detection of prey or predators. The visual system not only integrates (cues from the same object) and segregates (cues from different objects) these cues, but also makes accurate estimates of them for precisely representing the visual scene. For example, estimating the speed of the prey might be as crucial as detecting it, for successful foraging.

The study of visual motion perception has been instrumental in shaping our understanding of vision. The rules governing motion processing have been widely studied at both perceptual and physiological levels. A wealth of psychophysics has identified 2D motion as an important cue for the segmentation of objects and surfaces (Burr & Thompson 2011 for review). 2D motion combined with other cues like stereoscopic depth, as often available in natural vision, can better segment multiple objects (Hibbard & Bradshaw 1999, Snowden & Rossiter 1999, Greenwood & Edwards 2006, Goutcher 2016). In addition to visual features, the role of cognitive processes like selective attention in segmentation has been established (Desimone & Duncan 1995, Valdes-Sosa et al 1998, Valdes-Sosa et al 2000, Tse 2005, Harrison et al., 2019). However, the underlying neural mechanisms of motion-based segmentation of multiple objects are not fully understood,

especially when motion is combined with other visual features like depth. In my thesis, I aim to investigate the neural representation of multiple visual motion stimuli with a focus on finding the neural signatures of visual segmentation. Because the middle-temporal cortex (area MT) of primates is widely known for its importance in visual motion processing, we focus on area MT of macaque monkeys in our neurophysiological studies.

Motion and depth processing in area MT

Neurons in macaque area MT are tuned to both motion direction in the (2D) frontoparallel plane and the horizontal binocular disparity[1] (Maunsell & Van Essen 1983a, Maunsell & Van Essen 1983b, Bradley et al. 1995, DeAngelis & Newsome 1999, DeAngelis & Uka 2003). Perturbation studies involving lesions and microstimulation of area MT revealed its direct contribution to the perception of motion and depth features (Salzman et al 1992, DeAngelis & Newsome 1999, Kim et al 2015). Is there a functional significance of these two visual features being coded by the same brain area? My first guess would be area MT might contribute to the perception of motion in depth. But most of the evidence suggests otherwise that MT is largely specialized for 2D motion processing (Hejja-Brichard et al 2020, Lagae et al 1994, Maunsell & Van Essen 1983b). What is the role of disparity in the processing of 2D motion in area MT? One neurophysiology study (Bradley et al 1995) has established that disparity plays an intrinsic role in MT motion computation by providing a means of distinguishing motion signals from different surfaces. This study focused on the role of disparity cues in segregating overlapping motion signals moving in opposite directions. However, using a smaller angular separation between the motion directions might place a stricter constraint on the use of disparity cues for segmenting the motion

[1] Horizontal disparity is the positional difference between the left and right eye retinal images of objects. It provides valuable information about the depth of an object relative to the fixation plane (Howard & Rogers 1995).

signals. Moreover, this study used a simple fixation task and hence there was no measure of how well the animal segregated the motion signals using disparity cues. Hence, it is not fully understood how disparity helps in segmenting two motion signals spanning a range of directions and what neural mechanisms might help achieve this motion segmentation aided by disparity cues. In chapter 2, I address this gap by using a behavioral task that requires the animals to combine motion and disparity cues to achieve visual segmentation and I examine the underlying neural mechanisms.

Visual motion information comprises both the direction of an object's movement and its speed. Besides their established role in coding for motion direction, MT neurons are selective to a range of speeds (Maunsell & van Essen 1983a, Mikami, Newsome, & Wurtz, 1986, Krekelberg & van Wezel 2013) and there is evidence that MT is directly involved in speed perception (Orban et al. 1995, Pasternak & Merigan 1994, Liu & Newsome 2005). Most of these studies focused on understanding how MT neurons encode a single speed (except for Krekelberg & van Wezel 2013). However, realistically, visual scenes contain regions encompassing the motions of multiple objects. This relative motion is a salient cue for segmenting the objects from each other (Lamme et al 1993, Grossberg, 1994, Braddick, 1993). Most of the studies that characterized the neural responses to relative motion focused on placing one motion in the classical receptive fields (CRF) and the other in the suppressive surrounds of MT (Allman et al. 1985, Born, 2000, Huang et al. 2007) determining the influence of motion beyond the CRF. MT neurons have larger receptive fields (RF) than those in earlier visual areas, making it more likely to encompass multiple objects within their RFs under natural vision. It remains unclear how different speeds of objects influence MT neuronal responses in a manner to aid perceptual segmentation of multiple objects. In chapter

3, I examine this by characterizing MT neurons' tuning to two speeds presented next to each other in the RFs and determining the neural code that supports the segmentation of two speeds.

Neural encoding of multiple motion stimuli in area MT

Dynamic visual scenes often have one or more motion signals. A difficult computational task faced by visual motion processing is to either integrate or segment local motion signals to encode behaviorally relevant information. For example, motion signals from different regions of the same object/surface need to be integrated and signals from the same (or nearby) region but belonging to different objects need to be segregated. The visual system, of course, can combine other cues (like color, texture, or depth of the objects) with motion to make this task less challenging in a natural setting. But even when motion is the only differentiating cue between two overlapping surfaces, humans and non-human primates can not only segment multiple surfaces, but also accurately estimate their motion directions or speeds (when there is a big enough difference between the motion directions or speeds). This phenomenon, well studied in motion psychophysics, is called transparent motion, where two groups of dots (random dot patches) moving in different directions (or speeds) produce a perception of two separate surfaces sliding on top of each other. Neuroscientists have used transparently moving stimuli to reveal the neural mechanisms of motion-based segmentation, controlling for other cues (Snowden et al. 1991, Qian & Andersen, 1994, Treue et al. 2000; Rosenberg et al. 2008, Krekelberg & van Wezel, 2013, McDonald et al. 2014, Xiao et al. 2014, Xiao et al. 2015, Wiesner et al. 2020).

Early studies of neural mechanisms of transparent motion stimuli reported that adding a second motion direction to a neuron's preferred motion direction (PD) suppressed the neuronal responses in area MT i.e., the response to two motion surfaces is less than that to the preferred

motion surface alone (Snowden et al. 1991; Qian and Andersen 1994b). Snowden and others (1991) also found that this response suppression increased as the second motion direction was made more and more different from the neuron's preferred direction, with the maximum suppression occurring when the second direction was opposite to the PD (motion opponency). This means that not only MT neurons differentially respond to single and two overlapping motion directions, but also this differential activation depends on the component directions of the overlapping motion stimuli.

To determine how much of an influence each component motion direction has on the responses to overlapping directions, previous studies in our laboratory expressed MT neuronal responses to overlapping motion as a weighted sum of the responses to component directions alone (Xiao et al. 2014; Xiao and Huang, 2015; Wiesner et al 2020). This method is powerful because it allows for capturing the heterogeneity of weighting of each stimulus component in the responses to overlapping motions across a population of neurons. The weights in turn might reveal the computations underlying the segmentation and representation of the two overlapping motion directions. For example, one group of neurons showing a significantly higher weight to one stimulus component than the other might imply a biased representation of that component over the other. One can imagine a segmentation strategy where the representations of the components of multiple motion stimuli are distributed over subgroups of neurons (via the response weights described above), as though the information about the components is segregated at a population level. In chapter 2, I examine if such a strategy is implemented by MT neurons to represent two overlapping motion directions separated in depth, focusing on how disparity cues shape the response weights to individual motion directions of the overlapping stimuli.

In chapter 3, I examine how MT neuronal responses are affected by splitting different motion components, differing in their speeds but not directions, into spatially distinct stimuli. In other words, the two speeds are not overlapping, but form a border that is only defined by the motion (motion discontinuity border). In natural vision, this border could be an important cue for segmenting objects and even for perceiving their forms. A study in area V4 found that a significant portion of neurons show selectivity for the orientation of the motion discontinuity border (defined by motion direction alone) suggesting that area V4 could play an important role in form-from-motion (FfM) perception, where the shape of an object is derived by motion direction cues alone (Mysore et al. 2006). An ongoing study in our laboratory tested how neurons in area MT encode two adjacent stimuli with a motion direction discontinuity border placed within their classical RFs (Wiesner & Huang SfN abstract 2019). They found that responses show a strong spatial location bias towards the component that elicits a stronger response when presented alone, regardless of the orientation of the border. They concluded that a spatial bias would be beneficial to the visual system by making the population response heterogeneous to better differentiate the two components. Much less is understood about how MT neurons encode two speeds that are non-overlapping in a way that helps distinguish them, and I explore this in chapter 3.

Selective attention and object-based representations

As stated above, in chapter 2, I aim to study the role of binocular disparity in segmenting two motion surfaces. It is well known that selective attention can modulate neuronal responses to single and multiple stimuli (Treue & Maunsell 1996, Ferrera & Lisberger 1997, Reynolds et al. 1999, Treue & Martínez Trujillo1999, Recanzone & Wurtz 2000, Li & Basso 2005, Lee & Maunsell 2010, Ni et al. 2012). Therefore, when studying the role of binocular disparity in representing multiple surfaces, it is important to control attention. We are also interested in the

role of attention in segmentation itself. One possibility is that attention may be directly involved in and required for the segmentation process. Another possibility is that the initial process of segmenting two different surfaces does not require attention. However, after the initial representation of individual surfaces is established through pre-attentive sensory processing, attention may exert its selective influence on the initial neural representation and further enhance the representation of the attended surface and therefore facilitate segmentation. If that is the case, we would expect the neural signature of segmentation aided by disparity cues to occur earlier than the effects of selective attention. I will evaluate these possibilities in chapter 2 by studying the temporal development of responses to overlapping motion stimuli with disparity cues. Furthermore, although spatial attention and feature-based attention are well studied, surface- and object-based attention is not well understood. Wannig and colleagues (2007) showed that when two surfaces in different colors are spatially overlapping within the "spotlight" of attention (ruling out the space-based attention), attention can select surface representations to modulate area MT neuronal responses to motion. However, as the surfaces were defined by a combination of color and motion, given MT neurons are not usually selective to color, it was not clear where in the brain these surface representations formed initially. This conveys a need that our study in chapter 2 addresses by examining the neural mechanisms that represent surfaces, defined by disparity and motion direction (two features conjointly coded by MT neurons), and the attentional modulation on this representation.

Divisive normalization model

From the description above, neuronal responses to multiple stimuli can be described as a weighted sum of responses to the individual components presented alone (weighted summation model). What are the factors that define/influence these weights? In chapter 2, we turned to the

divisive normalization model (DNM) to answer this question. In this descriptive model, the response to a stimulus is determined by the summed activity generated by the stimulus drive, along with pooled responses (the normalization signal) of nearby neurons (Heeger 1992, Carandini & Heeger 1994, Simoncelli & Heeger 1998, Carandini & Heeger 2011). DNM is widely used to explain the responses to multiple stimuli/inputs in different visual areas (area MT: Britten & Heuer 1999, Heuer & Britten 2002; area MST for multisensory integration: Ohshiro et al. 2011). Importantly, DNM enhances a weighted summation model by enabling us to define the response weights of the individual components in terms of the component stimulus properties (e.g., intensity or contrast etc.,) and their interaction. For example, several studies have used the normalization model to describe visual neurons' responses to two stimuli in their RFs that had different contrasts (Heuer & Britten 2002, Busse et al. 2009). The influence of adding a second stimulus to the first, on the neuronal responses, depended on the contrast values of the stimuli. To capture this contrast dependence using DNM, these studies used contrast values as gain factors/weights for the excitatory drive from the stimulus components and for the normalization signal. DNM is powerful in predicting a range of neural computations from equal summation to winner-take-all competition of the components of multiple stimuli (Britten & Heuer 1999, Heuer & Britten 2002, Busse et al. 2009, Xiao et al. 2014, Wiesner et al. 2020). Other studies have modified the divisive normalization model to explain the effects of attention across a wide range of behavioral and stimulus conditions (Boynton 2009, Lee & Maunsell 2009, Reynolds & Heeger 2009, Lee & Maunsell 2010, Ni et al. 2012). These studies propose that attention can alter the balance between the stimulus drive of the attended stimuli and the summed activity of the neighboring responses (the normalization signal), weakening/strengthening the gain control, and thereby resulting in changes of neural activity. In chapter 2, I used a modified version of the DNM to describe MT

neuron responses to two overlapping motion directions separated in depth and to capture the effects of selective attention on the responses.

Overview and Objectives

The neural representation of multiple moving stimuli in area MT has been well documented. However, most of these studies focused on stimuli differing in motion cues alone. It is not completely understood how multiple cues or features interact with each other when present inside a neuron's receptive field (RF), and how they influence neuronal responses to form the representation of multiple objects. The goal of this dissertation is to elucidate the neural mechanisms underlying the segmentation of multiple surfaces and objects using 2D motion direction, speed, and disparity cues. Toward this goal, in chapter 2, I examine the neural mechanisms that represent multiple surfaces, defined by disparity and motion direction (two features conjointly coded by MT neurons), and probed the effect of selective attention on this representation. These findings have interesting implications for neural surface-based representations resulting from the combination of two visual cues and pertinent attentional mechanisms. In chapter 3, I investigate how two spatially separated stimuli moving at different speeds within the RF are encoded and elucidate how the interactions between the spatial and speed preferences of MT neurons might aid in perceptual discrimination of the two speeds. In addition to contributing to our understanding of speed processing, which is currently limited, these findings will provide new insights into how different encoding strategies might be at play when multiple stimuli are overlapping in space vs when they are separated. Finally, this work reduces the gaps in our understanding of how the brain achieves segmentation of objects in natural vision, and how different visual features and cognitive processes like attention contribute to this process. The

findings of this thesis will likely stimulate new ideas to be tested toward understanding vision in the natural environment.

Chapter 2: Neural representation of overlapping motion surfaces separated in depth: role of binocular disparity and selective attention

Introduction

Natural scenes often contain multiple objects and surfaces situated closely in space. A fundamental function of visual processing is to segment a visual scene into distinct objects. Various objects commonly have different visual features (for example depth, motion, or color) and the visual system can exploit these features to aid in their segmentation. However, it is not fully understood how these different features interact with each other when present inside a neuron's receptive field (RF), and how they influence neuronal responses to form the representation of multiple objects.

Motion and stereoscopic depth play important roles in visual segmentation. While motion and depth cues alone can aid in segmentation independently (e.g., motion transparency: Wallace & Mamassian 2003; stereo transparency: Wallace & Mamassian 2004), together they can provide strong cues for segmenting a visual scene into distinct objects and surfaces (Hibbard & Bradshaw 1999, Snowden & Rossiter 1999, Greenwood & Edwards 2006, Goutcher 2016, Kohler et al. 2019, Nawrot & Blake 1989, Bradshaw & Rogers 1996, Ban et al 2012). For example, it becomes easier to segment two overlapping surfaces from each other when they are moving at different stereoscopic depths than when they are moving at the same depth. In addition to visual features, cognitive processes like selective attention can also facilitate segmentation (Desimone & Duncan 1995, Valdes-Sosa et al 1998, Valdes-Sosa et al 2000, Wannig et al. 2007, Tse 2005, Harrison et al., 2019). Selectively attending to one of the two overlapping surfaces can make it easier to segment the surfaces and help in extracting information about the attended surface. In this study,

we aimed to understand the roles of stereoscopic depth and selective attention in segmenting overlapping surfaces moving in different directions. For this, we designed a novel motion discrimination task, in which using binocular disparity cues, macaque monkeys had to selectively attend to one of the two moving surfaces, overlapping in 2D space but separated in depth. Successful segmentation of these stimuli gives rise to the perception of transparent motion. To elucidate the role of disparity cues on the neural encoding of transparent motion, we paired the behavioral task with recordings from the middle temporal cortex (area MT), which plays a key role in both motion and depth processing (Maunsell & Van Essen 1983, Albright 1984, Bradley et al., 1995, DeAngelis & Uka 2003, Salzman et al 1992, DeAngelis & Newsome 1999, Kim et al 2015).

Neurons in macaque MT are tuned to both motion direction in the (2D) frontoparallel plane and horizontal binocular disparity, an important depth cue (Maunsell & Van Essen 1983; Bradley et al., 1995; DeAngelis and Newsome,1999; DeAngelis & Uka 2003). A wealth of psychophysics studies suggests that depth cues can aid in 2D motion integration and segmentation (Hibbard & Bradshaw 1999, Snowden & Rossiter 1999, Greenwood & Edwards 2006, Goutcher 2016). What is the role of disparity in 2D motion processing in area MT? Smolyanskaya and colleagues (2013) showed that MT neurons encode motion direction and disparity features independently such that their tuning for a single motion direction does not change as the binocular disparity of the direction is varied. In contrast, when two motion directions are overlapping in a MT neuron's RF, Bradley and others (1995) showed that the neuronal responses increased as the disparity of one of the two motion direction was made more and more different from that of the other. This suggests that binocular disparity plays an intrinsic role in MT motion processing by providing a means of distinguishing motion signals from different surfaces. However, this study profiled the interaction

between the two features only for one set of motion directions for each neuron i.e., its preferred and null directions. It remains to be tested how a population of MT neurons, with a range of direction and disparity preferences, encode overlapping surfaces moving in different directions presented at two disparities.

Previous studies in our laboratory have shown that the responses of MT neurons to overlapping motion directions (without disparity cues) can be described as a weighted sum of the responses elicited by the individual component directions (Xiao et al. 2014; Xiao & Huang, 2015; Wiesner et al 2020). Xiao and others (2014) showed that when one of the two overlapping motion directions had a higher contrast than the other, MT neurons showed a greater response weight for the component with higher contrast which also elicited the stronger direction tuning response when presented alone. Even when the components elicited identical tuning responses, Xiao and Huang (2015) reported that for many MT neurons, the response weight for one stimulus component was significantly greater than the other. This unequal/biased pooling of the component responses allows different neuronal subgroups to selectively represent one motion direction over the other of the overlapping stimuli. Such an encoding rule might be beneficial for segmenting the two motion surfaces. We asked if such a strategy is implemented by MT neurons to represent two overlapping motion directions separated in depth, focusing on how disparity cues shape the response weights to individual motion directions of overlapping stimuli.

One purpose of selective attention task in this study was to ensure that the monkeys explicitly used disparity cues to represent the overlapping motion directions. The other purpose was to study how attention modulates the neural representation of the overlapping motion surfaces with disparity cues. Many past studies suggest that attention selectively enhances the neural representation of attended stimuli and reduces the influence of unattended stimuli (Treue &

Maunsell 1996, Ferrera & Lisberger 1997, Reynolds et al. 1999, Treue & Martínez Trujillo1999, Recanzone & Wurtz 2000, Li & Basso 2005, Lee & Maunsell 2010, Wannig et al 2007). Attention can select a spatial location (spatial attention), a feature of an object (feature-based attention), or a whole object/surface (i.e., all features of an object are selected; object/surface-based attention). Less is known about the neural mechanisms of object/surface-based attention compared to those of spatial and feature-based attention. Wannig and colleagues (2007) showed that when two surfaces in different colors are spatially overlapping within the "spotlight" of attention (ruling out only space-based attention), attention can select surface representations to modulate MT neuronal responses to motion. However, as the surfaces were defined by a combination of color and motion, given MT neurons are not usually selective to color, it was not clear where in the brain these surface representations formed initially. This conveys a need that our study addresses by examining the neural mechanisms that represent surfaces, defined by disparity and motion direction (two features conjointly coded by MT neurons), and the attentional modulation on this representation.

Methods

Three male adult rhesus monkeys (Macaca mulatta) were the subjects in the neurophysiological experiments. We followed experimental protocols approved by the Institutional Animal Care and Use Committee of the University of Wisconsin–Madison and conformed to U.S. Department of Agriculture regulations and to the National Institutes of Health guidelines for the care and use of laboratory animals. Procedures for surgical preparation and electrophysiological recordings were akin to those described previously (Xiao et al. 2015). We implanted a head post and a recording cylinder during sterile surgery with the animal under

isoflurane anesthesia. For electrophysiological recordings from neurons in area MT, we took a vertical approach and used tungsten electrodes (1–3 M; FHC). We identified area MT by its characteristically large portion of direction selective neurons, small RFs relative to those of neighboring medial superior temporal cortex (area MST), its location at the posterior bank of the superior temporal sulcus, and visual topography of the RFs (Gattass & Gross, 1981). We amplified electrical signals and identified single units with a real-time template-matching system and an offline spike sorter (Plexon). We monitored animal's eye position using a video-based eye tracker (EyeLink, SR Research) at a rate of 1000 Hz. For some of recording sessions from monkey B, we tracked both the left and right eye positions which were then used to calculate the horizontal vergence values (left eye X position-right eye X position).

Experimental procedures

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

After we isolated the spike waveforms of a single neuron, we first characterized the neuron's direction selectivity by interleaving trials of a 30° x 27° random-dot patch (RDP), covering the entire monitor, moving in different directions at a step of 45° and at a speed of 10°/s. The direction selectivity and preferred direction (PD) were determined using MATLAB (MathWorks). We then characterized the speed tuning of the neuron using a random dot patch

moving at different speeds (1, 2, 4, 8, 16, 32, or 64°/s) in the PD of the neuron. Using a cubic spline, we took the preferred speed (PS) of the neuron as the speed that evoked the highest firing rate in the fitted speed-tuning curve. Next, we mapped the receptive field (RF) location of the neuron by recording responses to a series of 5° × 5° patches of random dots that moved in the PD and at the PS of the neuron whose location we varied randomly to tile the screen (35×25°) in 5° steps without overlap. We then interpolated the raw map of these responses at an interval of 0.5° to obtain the location with the highest firing rate response as the center of the RF.

We used green and red anaglyph filters (Kodak Wratten filters, nos. 25 and 58) to create stereoscopic depth on the monitor using horizontal disparity. To achieve a disparity of + n° or - n°, we used two identical dot patterns, one in green and the other in red with the center of the red dot pattern offset horizontally by +n° or -n° relative to that of the green dot pattern, respectively. As the animals viewed the stimuli through a green filter on the left eye and red filter on the right eye, only one set of dots was visible to each eye creating a binocular disparity that resulted in depth perception. Crosstalk between the two eyes, as viewed through the filters, was 14.5% (the luminance of red dots was 13.85 and 2.32 cd/m2 through red and green filters; the luminance of red dots was 14.5 and 1.8 cd/m2 through red and green filters; the background luminance was 0.68 cd/m2 and 1.01 cd/m2 through red and green filters, respectively). We determined the selectivity of a MT neuron to binocular disparity by measuring its responses to a series of RDPs moving in its preferred direction and at its preferred speed (often within 5 to 20°/s) at fifteen different disparities from -1.5° to +1.5°.

Visual stimuli

Transparent motion stimuli constituted two overlapping circular RDPs of size 15° moving at two different depths: one at -0.1° binocular disparity (near) and the other at 0.1° binocular

disparity (far) relative to the fixation point on the monitor (Fig 1A). We refer to these constituent RDPs as 'stimulus components': the one at near disparity as 'near component' and the one at far disparity as 'far component'. Each RDP's dot density was 0.89 dots/deg^2 with infinite dot lifetime and had 100% coherence. The stimuli were always centered at (0°,0°) in the visual space covering the neurons' RFs but not necessarily centered on the RFs. The speed of the motion stimuli was 10°/s in most sessions and 12°/s in a few sessions.

Behavioral tasks

For the attention task (Fig 1B), we trained Monkey B and G to report the correct motion direction of the stimulus component at cued disparity (+ 0.1° or - 0.1°). Every trial began with the appearance of a fixation dot at the center of the screen and the reference stationary dots. The animal had to maintain fixation within 1°x 1° window of the dot to continue the trial (see Results for control of vergence). As a cue for attention, a static circular dot-patch with a diameter of 15° appeared for 500 ms (long enough to establish depth perception) at either near plane I.e., -0.1° disparity or far plane I.e., 0.1° disparity ("Cue Near/Attend Near" or "Cue Far/Attend Far conditions" respectively) with the reference stationary dots at zero disparity covering the rest of the monitor, followed by a 300 ms of blank screen with a fixation dot. Then, one stimulus component at a disparity consistent with that of the cue (Fig 1A "Single components") or the overlapping motion stimuli at 0.1° and -0.1° disparities (Fig 1A "Overlapping stimuli") appeared. The single stimulus or overlapping stimuli remained stationary for 500ms, and then started moving for 600ms in one of the 12 directions: 0 to 360° at 30° step (four example directions shown by green and blue arrows in Fig 1A). After the motion period, the stimuli were off, and 12 reporting targets arranged from 0 to 360° at 30° step along a circle of 10° diameter, appeared for 450 ms for monkey B and 600 ms for monkey G. During this time window, the animals had to report the

motion direction of the cued component by making a saccade to the corresponding target and holding fixation within an invisible 1.5°x 1.5° target window for 350ms to receive a juice reward. To make the task demanding of attention, we used two fixed angle separations between the overlapping motion directions (60°/DS60 or 120°/DS120). This also reduced the chances of the animals predicting the direction of the component at cued disparity just by attending to the other/un-cued component. To control for any effects of clockwise (C) or counter-clockwise (CC) arrangement of the overlapping stimuli on MT neuronal responses (Xiao et al 2015), we included two different configurations for each angle separation of transparent motion stimuli in our experiments: 1. near component motion direction was at clockwise (C) side of that of the far component (i.e., Near C-Far CC) and 2. near component motion direction was at the counter-clockwise (CC) side of that of the far component (i.e., Near CC-Far C). In each block of trials, we randomly interleaved trials of single stimulus located either at near or far disparity ("Single") with trials of overlapping motion stimuli with components at near and far disparities ("Overlapping").

For the attention-away task, the stimuli were centered on MT neurons RFs. To direct the animal's attention away from the RF, we trained Monkey R to perform a two alternative forced choice (2-AFC) task of discriminating whether the motion direction of an RDP, presented in the opposite visual hemifield, was clockwise (C) or counter-clockwise (CC) to vertically upward direction. Every trial began with the appearance of a fixation dot at the center of the screen and the reference stationary dots at zero disparity. The animal fixated within 1°x 1° window of the dot for 200 ms. Then the stimuli in the RFs ("Single" or "Overlapping" disparity motion stimuli in Fig 1A) appeared along with an "attention stimulus": a random-dot patch with a diameter of 5°, placed at 10° eccentricity in the opposite hemifield. Both remained stationary for 250ms. Then the "attention stimulus" moved for 500 ms either C or CC direction from the vertically upward

direction by 10°, 15°, or 20° in randomly interleaved trials. Then, all the stimuli were turned off, and two reporting targets appeared on the left and right side of the FP. The animal made a saccade to the left (or the right) side if the motion of the "attention stimulus" was counterclockwise (or clockwise) of vertical and maintained fixation on the chosen target for 200ms to receive a small juice reward.

Experimental design and statistical analysis

Neuronal response (firing rate) was calculated during the stimulus motion period (600 ms for the attention task and 500 ms for attention away task) averaged across repeated trials. We constructed the response tuning curves to unidirectional stimuli (Single component) and to bidirectional stimuli (Overlapping) and fitted the raw direction tuning curves using cubic splines at a resolution of 1°. For each vector averaged (VA) direction of the overlapping motion directions, we determined the responses elicited by the bidirectional stimuli and the constituent unidirectional stimulus components. To average the direction tuning curves across neurons, we rotated the spline-fitted tuning curve elicited by the bidirectional stimuli such that the VA direction of 0° was aligned with the PD of each neuron. We then normalized neuronal responses by each neuron's maximum of response to near and far components and averaged the aligned, normalized tuning curves across neurons. For brevity, we pooled the responses of Near C-Far CC and Near CC-Far C conditions for population averaged tuning curves in Fig 3 and Fig 14. The resulting tuning curves always had near component at C side and far component at CC side for visualization. To do so, we first horizontally flipped the tuning curves to the bidirectional and unidirectional stimuli of neurons of Near CC-Far C conditions, along the axis of VA direction. We then averaged the flipped tuning curves together with the tuning curves of Near C-Far CC condition.

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

SNL and LWS models

To quantify the relationship between the responses elicited by the bidirectional stimuli and those elicited by the individual stimulus components, we fitted the direction tuning curves using a linear weighted summation (LWS) model (Eq. 1) and a summation plus nonlinear interaction (SNL) model (Eq. 2) (Xiao et al. 2014) by minimizing the sum of squared error:

$$R_{nfpred}(\theta_n, \theta_f) = w_n R_n(\theta_n) + w_f R_f(\theta_f) + c \quad (1)$$

$$R_{nfpred}(\theta_n, \theta_f) = w_n R_n(\theta_n) + w_f R_f(\theta_f) + b R_n(\theta_n) R_f(\theta_f) \quad (2)$$

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

We classified MT neurons based on their responses to individual near and far components. "Near neurons" are those that had a higher response to preferred direction at near disparity (-0.1°)

than at far disparity (+0.1°) i.e., max ($R_n(\theta)$)>max ($R_f(\theta)$). "Far" neurons had max ($R_f(\theta)$)>max ($R_n(\theta)$). To determine whether the response elicited by the bidirectional stimuli showed a significant bias toward one of the two stimulus components, we compared the response weights, w_n and w_f (Eq. 2) using the weight bias index (wbi) as defined by Eq. 3.

$$wbi = (w_n - w_f)/(w_n + w_f) \quad (3)$$

We fitted the response tuning curves to the bidirectional stimuli using a couple of variants of a divisive normalization model (Carandini & Heeger 2011; see Results). All model fits were obtained using the constrained minimization tool "*fmincon*" (MATLAB) to minimize the sum of squared error. To evaluate the goodness of fit of a model for the response tuning curve to the bidirectional stimuli, we calculated the percentage of variance (PV) the model accounts for as follows:

$$PV = 100 * (1 - SSE)/SST \quad (4)$$

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).

Center of mass (CoM) of bidirectional tuning curves

We computed the center of mass (CoM) of the bidirectional tuning curves in different time windows after the motion onset as follows:

$$CoM(R_{nf}) = \frac{\sum R_{nf}(\theta).\vec{\theta}}{\sum R_{nf}(\theta)} \quad (5)$$

where θ is the vector averaged (VA) direction of bidirectional stimuli and $R_{nf}(\theta)$ is the neuronal response to the stimuli in each time window.

Results

To understand the neural mechanisms by which stereoscopic depth cues aid in 2D motion segmentation, we asked how neurons in the middle-temporal (MT) cortex represent overlapping moving surfaces separated in depth. Ensuring the use of depth in segmenting the moving surfaces (random-dot patches), we trained two macaque monkeys to perform direction discrimination at a cued binocular disparity (near or far) using selective attention. We paired the task with recordings from MT neurons while presenting the overlapping random-dot motion stimuli at two binocular disparities in their receptive fields.

We aimed to examine if MT neurons' representation of the overlapping moving surfaces at different depths is biased to either surface. These biases of individual neurons can reflect the preferential encoding of one component over the other. We hypothesized that as the overlapping motion stimuli are separated in depth, area MT can distribute the encoding of each component stimulus over subgroups of neurons defined by their disparity preferences. We examined the effects of selective attention to either disparity on this encoding distribution. We also characterized if the response bias to a surface at either disparity has temporal dynamics distinct from the response modulation caused by selective attention. We next examined how directing attention away from the stimuli affected MT neurons' responses to the stimuli. For this, we recorded responses of MT neurons to the overlapping stimuli in their RFs from a third monkey while he directed his attention to the opposite hemifield. Finally, we formulated a unified normalization model that can describe the data from the three animals.

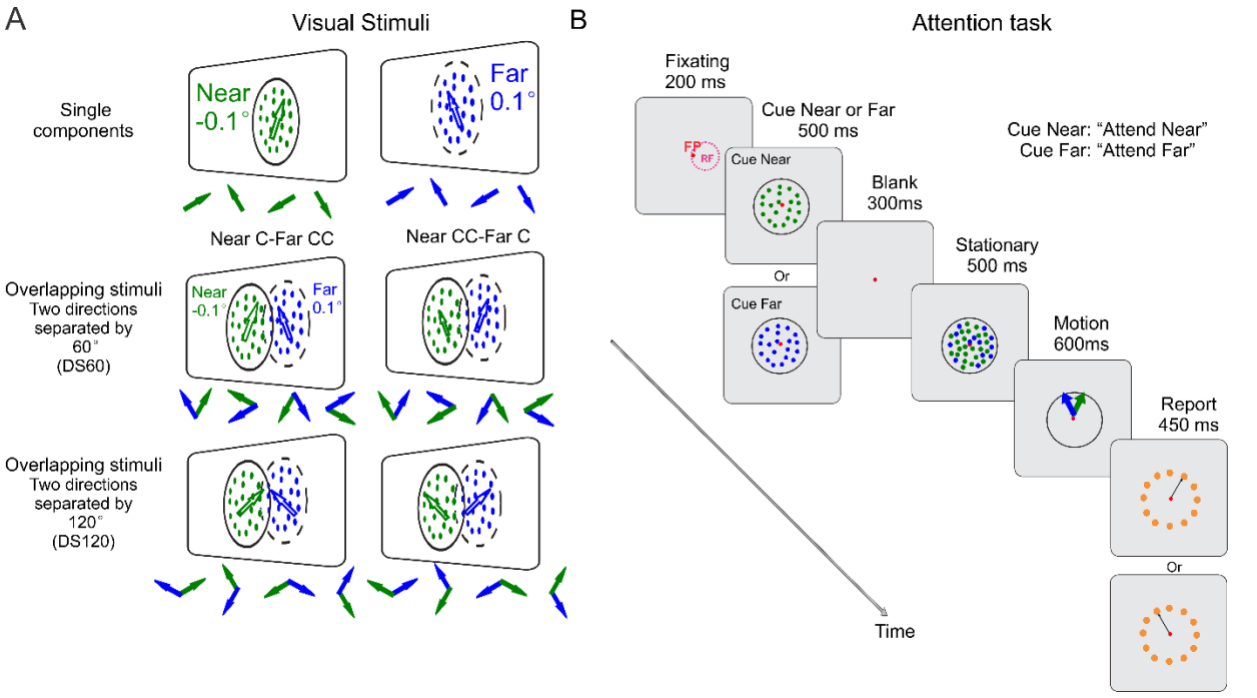

Figure 1. Visual stimuli and attention task.

(A) Visual Stimuli: Two random dot patches (RDP), one at -0.1° (near) disparity and the other at 0.1° (far) disparity, were presented individually or overlapping in space. The two RDPs moved in different directions from 0° to 360° such that their angle separation is either 60° or 120° with different angular arrangements (Near clockwise C-Far counterclockwise CC; Near counterclockwise CC-Far clockwise C). The dots were achromatic, and the colors shown are only for illustration purpose. (B) Attention task: Each trial was initiated after the fixation was achieved, then a cue was presented on either the near plane (as shown in the figure) or the far plane randomly. After 300ms blank period, either a single motion or overlapping motion stimuli (as shown in the figure) appeared stationary for 500ms, and then moved for 600ms. The animal had to maintain fixation all along. At the end of each trial, the monkey was required to make an eye movement to report the motion direction of the stimulus at the cued plane and hold the fixation within a 2°x2° window of the correct target for 450ms.

Behavior

To ensure a good quality of selective attention and the use of depth cues in the direction discrimination task (see Fig 1B and Methods), we excluded recording sessions during which the percentage of correct choice trials of all trials was below 50%. Our final dataset includes recordings from 61 MT neurons from monkey B from 49 sessions (mean % correct trials=75) and 56 neurons from monkey G from 44 sessions (mean % correct trials=72). Both monkeys performed better at

discriminating motion direction when there is only one motion surface (Fig 2A and C: "Single" condition) compared to when there are two overlapping motion surfaces separated in depth (Fig 2A and C: "DS60" and "DS120" conditions) (monkey B: mean % Single=87, mean % Overlapping=72, one-sided signed rank p= 3.8917e-10; monkey G: mean % Single=86, mean % Overlapping=68, one-sided signed rank p= 5.7995e-09). We were interested in examining any differences in performance of motion discrimination at near and far disparities of the overlapping stimuli ("cue Near"/ "Attend Near" and "cue Far"/ "Attend Far" respectively). When monkeys attended to a single motion surface presented alone, the performance was not different when the single surface was at the near and far disparity (monkey B: mean % correct "Single Near"=88, mean % correct "Single Far"=86, one-sided signed rank test p= 0.9815; monkey G:mean % correct "Single Near"=86, mean % correct "Single Far"=86, one-sided signed rank test p= 0.6735). When two motion directions were overlapping separated in depth, monkey G was significantly better at discriminating motion direction at near disparity than far (Fig 2C, D; mean % correct "Attend Near"=72, mean % correct "Attend Far"=65, one-sided signed rank test p= 4.2858e-13), whereas monkey B performed equally well at discriminating the motion direction at near and far disparities (Fig 2A, B, mean % correct "Attend Near"=72, mean % correct "Attend Near"=72, one-sided signed rank test p= 0.5612).

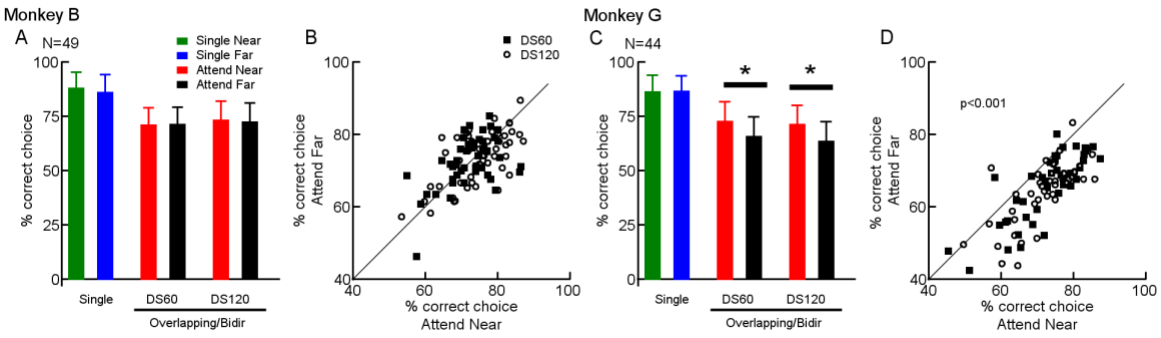

Figure 2. Attention task performance.

(A) Monkey B: Average % of correct choice trials across 49 recording sessions of Monkey B for Single (green: Near, blue: Far) and Overlapping motion directions (red: Attend Near, black: Attend Far). DS60 and DS120 indicate 60° and 120°separations between overlapping motion directions. The error bars show standard deviations. (B) Monkey B: % of correct choice trials of Attend Near conditions (or Cue Near) and Attend Far (or Cue Near) conditions are plotted on x and y axes, respectively. Each symbol represents one recording session. Solid squares and open circles are for DS60 and DS120 conditions. p value is from one-sided signed rank test comparing x and y values. (C) Same as (A) for Monkey G across 44 sessions. (*) indicates $p<0.001$, paired signed rank test. (D) Same as (B) for Monkey G.

Heterogeneity of MT responses to overlapping motion stimuli separated in depth: Near and Far neuronal subpopulations

Fig 3 shows the responses of four representative neurons (Fig 3A and B from monkey B and Fig 3C and D from monkey G). Rows correspond to two different stimulus arrangements (Near C-Far CC and Near CC-Far C) for each angle separation between the overlapping motion directions (DS60 and DS120) shown in the cartoon on the left. In each recording session, as the monkeys performed the attention task, we measured the responses of MT neurons to a range of motion directions (0 to 360° at a 30° step) characterizing the direction tuning curves for single components and for overlapping motion stimuli with an attention cue to near or far surface (Fig 3). For each panel, we arranged the tuning curves such that, the responses to overlapping motion stimuli (red/attend near and black/attend far curves) are aligned with the responses to the single near and far components that constitute the overlapping stimuli at each VA direction (green and blue curves, respectively). Note the color-coded abscissae for the component directions in Fig. 3A4. To capture the preferences of individual neurons to single components, we classified MT neurons based on their responses to single near and far components. “Near neurons” are those that had a higher response to preferred direction at near disparity (-0.1°) than at far disparity (+0.1°) i.e., max $(R_n(\theta))$>max $(R_f(\theta))$. “Far” neurons had max $(R_f(\theta))$>max $(R_n(\theta))$.

When the stimulus components were overlapping, for the two example near neurons (Fig 3A1 to A4 from monkey B, Fig 3C1 to C4 from monkey G), the bidirectional tuning curves (red and black curves) were biased to the preferred component by definition i.e., the single component at near disparity (green curve). The bidirectional tuning curves of the example far neuron from monkey B (Fig 3B1 to B4) were biased to the motion at far disparity (far component), which is also its more preferred disparity when presented alone. Deviating from this trend, the bidirectional tuning curves of the example far neuron from monkey G (Fig 3D1 to D4) were biased to the motion at near disparity (near component), which is its less preferred disparity. For the example neurons from monkey B, selective attention to the preferred disparity had a noticeable enhancement of this bias to the motion direction at that disparity (comparing red/Attend Near and black/Attend Far curves in Fig3A1 to B4), whereas we did not see a prominent effect of attention on the bidirectional responses of the example neurons from monkey G (Fig 3C1 to D4). Notably, the biases of all the bidirectional curves in Fig 3 are quite different from those predicted by the average of the responses (shown by grey dashed lines) to the single components at two disparities. This suggests that a weighting scheme different from averaging (equal weighting) of component responses might describe the bidirectional responses better.

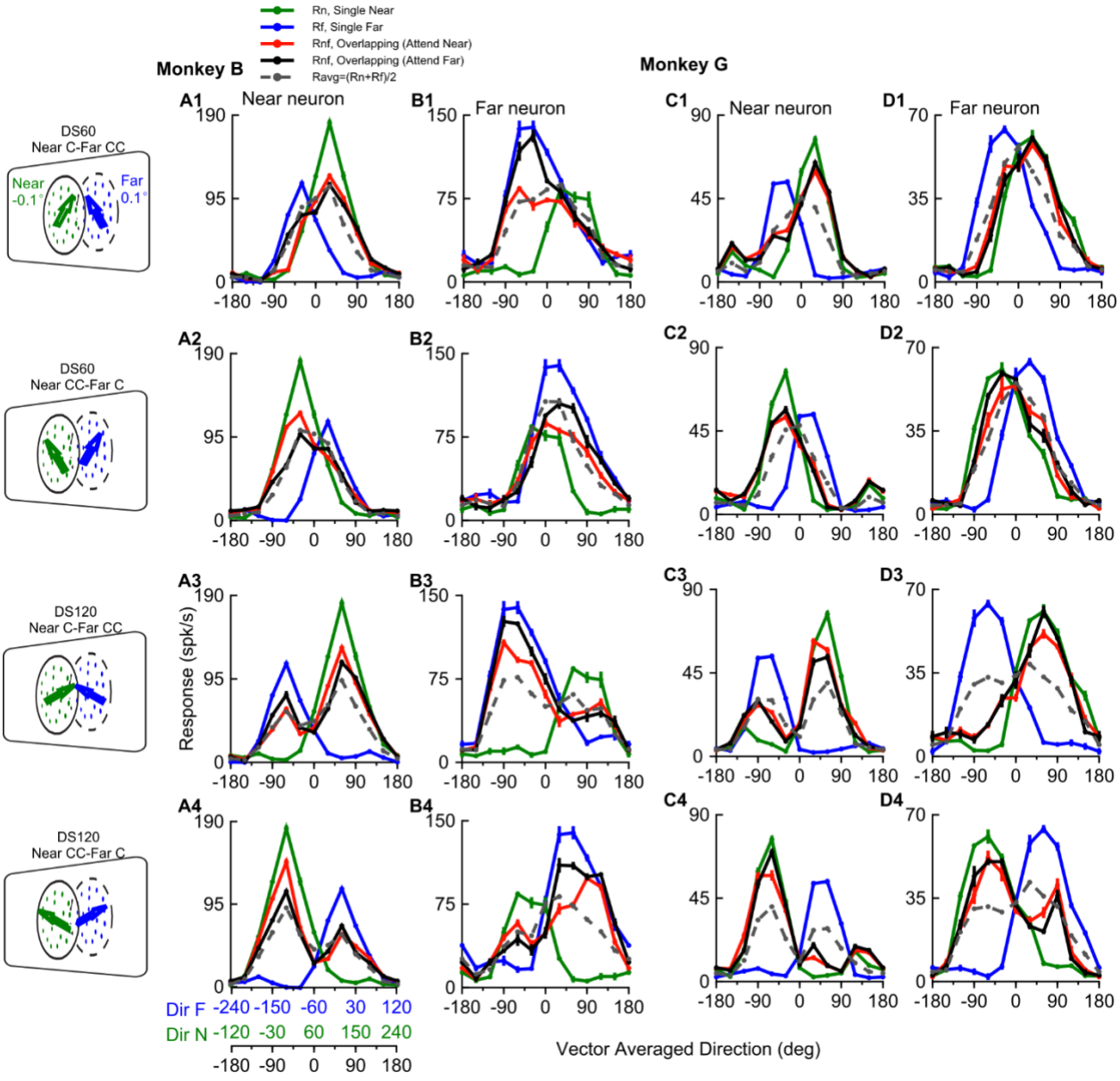

Figure 3. Example MT neurons' tuning curves to single and overlapping directions separated in depth.

(A1-A4) Single component direction (green: Near and blue: Far) and overlapping/bi-directional (red: Attend Near and black: Attend Far conditions) tuning curves of an example near neuron of Monkey B. Grey dashed lines show the average of the single direction tuning curves (green and blue). The x-axis labeled in black indicates the vector average direction of the bidirectional stimuli. The x-axes labeled in blue and green (A4) indicate the direction of the far (Dir F) and near component (Dir N), respectively. The three x-axes are aligned such that the component directions shown in blue and green correspond to the directions of the two stimulus components at each vector average direction. (B1-B4) Tuning curves of an example far neuron of Monkey B. (C1-C4) and (D1-D4) are tuning curves of example near and far neurons of Monkey G, respectively. (A1-D1), (A2-D2), (A3-D3), and (A4-D4) are response tuning of example neurons to different direction separations and arrangements of the overlapping motion directions shown in the leftmost cartoons.

Population responses

Fig 4 shows the tuning curves averaged across near and far neurons for DS60 and DS120 conditions. As the averaged tuning curves of the two stimulus arrangements Near C-Far CC and Near CC-Far C had consistent trends, for simplicity, we flipped the latter tuning curves to align with the former curves and averaged them together in each panel. For both monkeys, for near neurons (Fig 4A, C, E, G), when the two overlapping motion directions are separated in depth, the averaged tuning curves are biased to the tuning curves elicited by single near component which is also the preferred component of these neurons. While monkey B far neurons' bidirectional tuning curves showed a bias to the preferred (far) component like the near neurons (Fig 4B, D), monkey G far neurons' tuning curves showed a bias to the near component (Fig 4F, H). Only for monkey B and not for monkey G, the population averaged bidirectional tuning curves showed differences (subtle) between the two attention conditions, "Attend Near" and "Attend Far". Attention to near component moving close to neurons' preferred direction (PD) increased the response to the overlapping stimuli averaged across near neurons compared to that of attention to far (Fig 4A, C: red data points higher than black ones at the peaks of the tuning curves). Similarly, attention to far component moving close to PD increased the bidirectional response averaged across far neurons compared to attention to near (Fig 4B, D: black data points higher than the red ones at the peaks of the tuning curves).

Neuronal biases to near and far components of the overlapping stimuli

To quantify the biases of bidirectional tuning curves to the single component tuning curves, for each neuron, we fitted the bidirectional responses with the SNL model (see Eq 1 in Methods) which provided an excellent fit (92% variance explained for monkey B and 83% for monkey G averaged across neurons; see Eq. 4 in Methods). The weights for near and far single components

(w_n and w_f) obtained from the model denote the biases of bidirectional tuning curves to the near and far component tuning curves, respectively. Weight bias index (WBI) defined as $wbi = (w_n - w_f)/(w_n + w_f)$ quantifies the net bias to a single component for any given bidirectional tuning curve, positive values indicating a bias to the near component ("near bias") and negative values indicating a bias to the far component ("far bias").

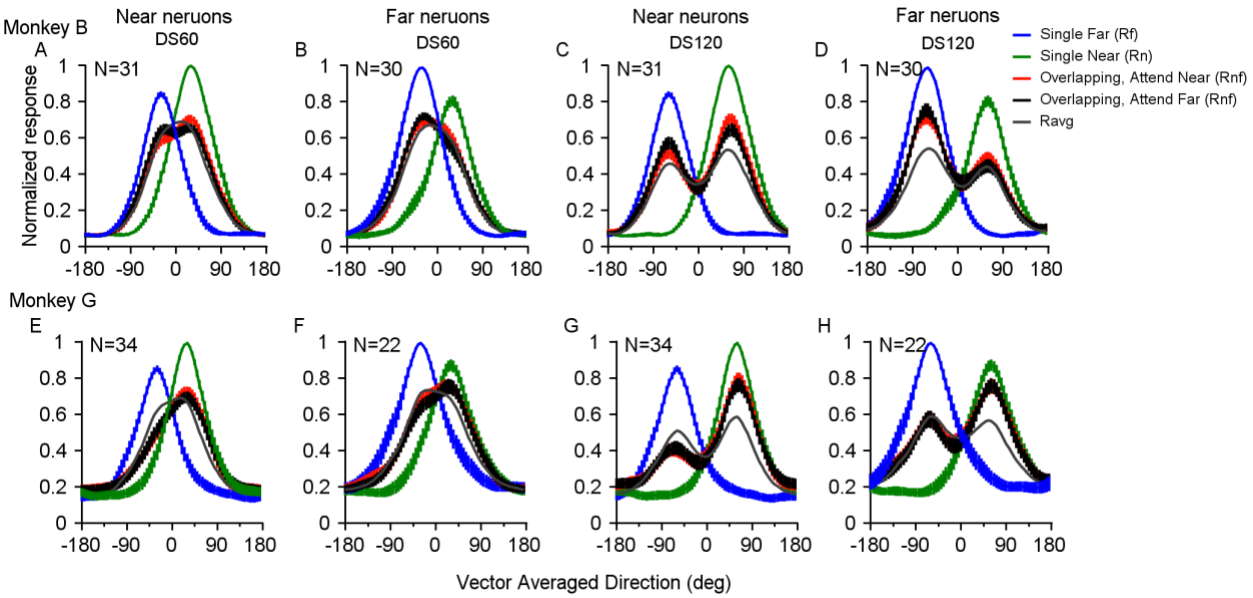

Figure 4. Population-averaged tuning curves to the single motion components and the bidirectional stimuli. Green, blue curves indicate the tuning to the near and far components: single near, single far. Red and black indicate the tuning to overlapping near and far components (bi-directional tuning) in two attention conditions: Attend Near and Attend far, respectively. The x-axis is the vector averaged (VA) direction of the bidirectional stimuli. The direction of 0° was aligned with the PD of each neuron before the tuning curves were averaged across neurons. (A, B) Near and Far neurons of monkey B, respectively, for DS60 condition. (C, D) Near and far neurons of monkey B, respectively, for DS120 condition. (E, F) Near and far neurons of monkey G, respectively, for DS60 condition. (G, H) Near and far neurons of monkey G, respectively, for DS120 condition. Note for each panel Near CC-Far C tuning curves are flipped and averaged with Near C-Far CC tuning curves.

Figure 5 compares the WBIs between "Attend Near" and "Attend Far" conditions for near and far subpopulations (green and blue dots respectively). For monkey B, across near neurons, we saw a stronger bias to near component (near bias) when the animal attended the near surface than

the far surface (mean WBI (Attend Near) =0.12, mean WBI (Attend Far) =-0.01, one-sided signed rank test p=5.4289e-09). In contrast, far neurons showed a stronger bias to far component (far bias) when the animal attended the far surface than the near surface (mean WBI (Attend Near) =-0.17, mean WBI (Attend Far) =-0.30, one-sided signed rank test p=0.0067). Across attention conditions, for monkey B, near neurons showed a significant near bias (mean WBI = 0.06, one-sided signed rank test p=0.0024), and far neurons showed a significant far bias (mean WBI = -0.24, one-sided signed rank test p=5.5994e-20). Distinctly, for monkey G, both near and far neurons showed a significant near bias in both attention conditions and attention did not change the magnitude of this bias significantly (near neurons: mean WBI (Attend Near) =0.35, mean WBI (Attend Far) =0.34, one-sided signed rank test p= 0.2932; far neurons: mean WBI (Attend Near) =0.24, mean WBI (Attend Far) =0.25, one-sided signed rank test p= 0.5916). Notably, regardless of the attention condition, for both monkeys, near neurons showed stronger bias to the near surface than the far neurons (across attention conditions, mean WBI (near neurons) =0.3410, one-sided signed rank test p=1.1198e-35mean WBI (far neurons) =0.2436, one-sided signed rank test p= 3.1188e-22). This suggests that the representation of overlapping motion surfaces with disparity cues in area MT might be distributed such that the neurons that fire more (or "prefer") to a surface at one disparity than the other, also selectively represent the more "preferred" surface when the two surfaces are overlapping in their receptive fields. Importantly, this selective representation of one component over the other is more than what's expected from simply averaging the component responses, as shown by heterogenous weights. This selective representation by subgroups of neurons defined by their disparity preferences might be one way area MT exploits the disparity cues of the overlapping surfaces to segregate them.

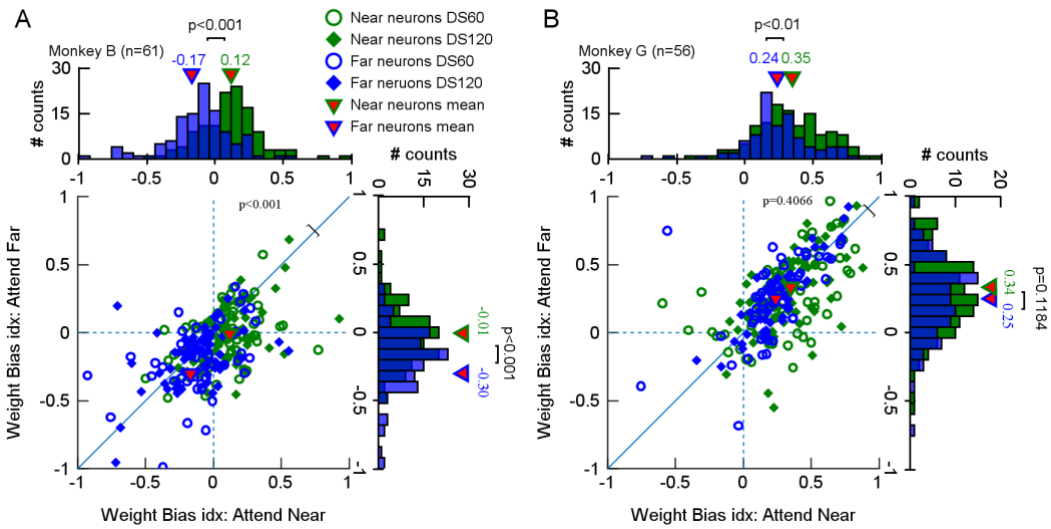

Figure 5. Effects of attention on weight bias indices (WBI) of near and far neurons.

(A) Monkey B: On the x-axis and y-axis are the WBIs ($wbi = (w_n - w_f)/(w_n + w_f)$) of Attend Near and Attend Far conditions, respectively. Each dot represents the result from one bi-directional tuning curve of a neuron. Open circles and solid diamonds correspond to DS60 and DS120 conditions, respectively. Each neuron contributes to two of each of those symbols from Near C-Far CC and Near CC-Far C conditions. Green and blue dots denote near and far neurons. The triangles show the mean values for each group. p values are from one-sided signed rank test. (B) Same as (A) for monkey G.

This led us to evaluate if the preferences of MT neurons to a single component (defined in the disparity domain: near or far) correlated with the representational biases of the two components overlapping in their receptive fields. To capture the variation in the extent of preferences to near and far components of "near" and "far" neurons, we defined a component preference index (CPI):

$$\text{CPI} = (\max(R_n) - \max(R_f))/(\max(R_n) + \max(R_f)) \quad (6)$$

where R_n and R_f are the tuning responses to the near and far component motion directions (green and blue curves in Fig 3). Positive and negative values denote a preference to near and far

disparities, respectively.

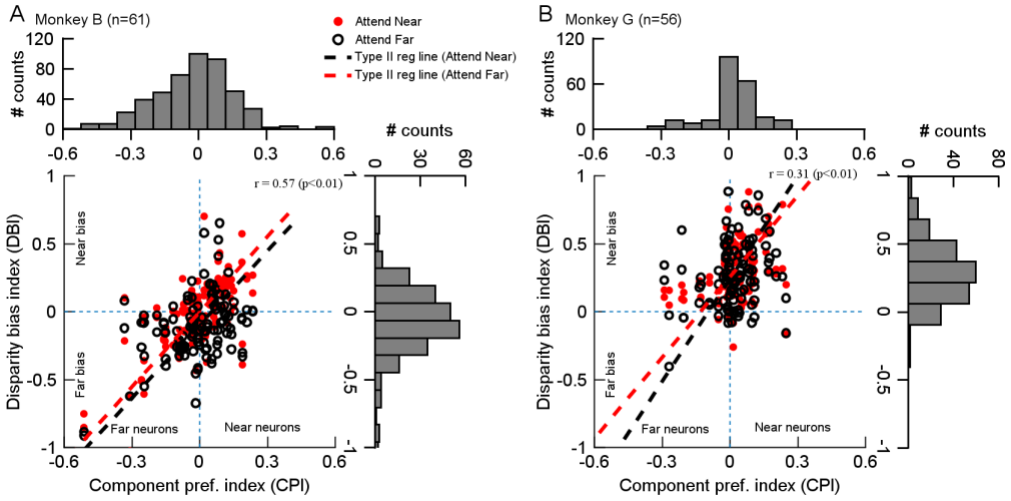

Figure 6. Disparity bias index (DBI) vs Component preference index (CPI)

(A) Monkey B: On the x and y axes are the CPIs and DBIs of Attend Near and Attend Far conditions, in red solid circles and black open circles, respectively. Each neuron contributes to two of each of those symbols from DS60 and DS120 conditions. The dotted lines are the type 2 regression fits. r is the Pearson correlation coefficient between x and y values of Attend Near and Attend Far (red and black symbols) together with the corresponding p values in brackets. (B) Same as (A) for monkey G.

Previous work in the laboratory (Xiao and Huang 2015) showed that MT neurons response to two overlapping motion stimuli (with no disparity cues) could be biased to either clockwise (C) or counterclockwise (CC) components. The purpose of including the stimulus arrangements in our study i.e., Near C-Far CC (NC) and Near CC-Far C (NCC) was to control for any possible C-CC biases that might be overlayed with disparity biases. Distinguishing the disparity biases (near or far) from the C-CC biases, we quantified the net bidirectional bias to near or far component by combining the weights from the responses to the two stimulus arrangements NC and NCC together. For this, we defined the disparity bias index (DBI) for each attention condition as:

$$\mathrm{DBI} = ((w_n - w_f)_{NC} + (w_n - w_f)_{NCC})/(w_n + w_f)_{NC} + (w_n + w_f)_{NCC}) \qquad (7)$$

DBI is analogous to WBI in Fig 5, with positive and negative values indicating net near and far biases of bidirectional tuning curves. Fig 6 examines the relationship between DBI and CPI values for all neurons for "Attend Near" (red dots) and "Attend Far" conditions (black dots) (note: each neuron contributes to two data points: one from DS60 and the other from DS120 conditions for each attention condition). Data from both monkeys indicated that there was a significant positive correlation between CPI and DBI measures, regardless of the attention condition (Monkey B: Pearson correlation coefficient value r=0.57; p<0.01 and type II regression line slope=1.84: Attend Near; slope=1.80: Attend Far; Monkey G: r=0.31, p<0.01 type II regression line slope=2.52: Attend Near; slope=1.98: Attend Far). This further confirms that the responses of MT neurons to two overlapping moving surfaces at different stereoscopic depths are significantly biased toward the surface that was located closer to the neurons' preferred disparity. And the magnitude of this bias is proportional to the extent of the neuron's preference to one disparity over the other. While there was a consistent positive correlation of DBIs with CPIs for both monkeys, the main difference is that for monkey G, the DBI values are shifted upward to the positive side for both negative and positive CPI values indicating both near and far neurons exhibited a near bias.

Effects of attention on model weights and neuronal responses

Response weights

We asked if selective attention changed MT neurons' bi-directional response weights of either component in a way to facilitate the representation of the attended component. To obtain the weights, we used LWS model fits described by the following equations:

$$R_{(An,f)pred} = w_{nAn}R_n + w_{fAn}R_f + c \quad (8)$$

$$R_{(n,Af)pred} = w_{nAf}R_n + w_{fAf}R_f + c \quad (9)$$

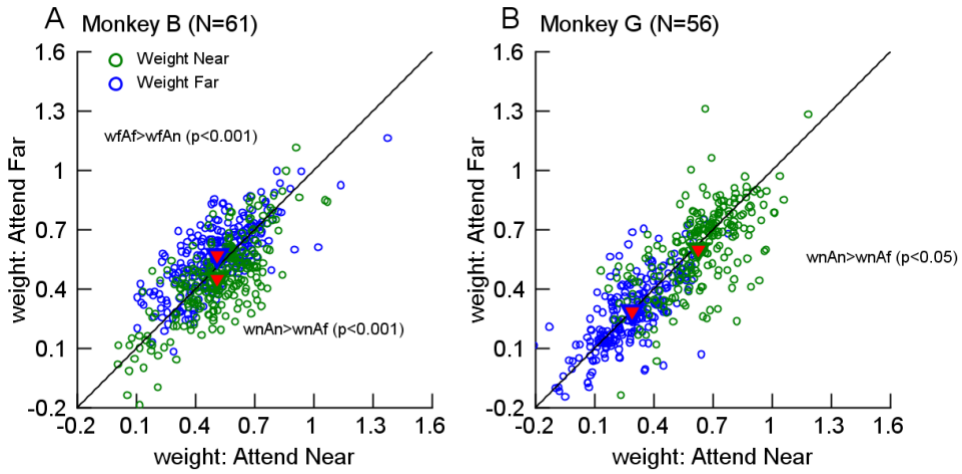

Figure 7. Effects of attention on response weights of near and far components

(A) Monkey B: On x and y axes are the weights from Attend Near and Attend Far conditions, respectively. Weight of near (x: w_{nAn}, y: w_{nAf})and far (x: w_{fAn}, y: w_{fAf}) components are shown in green and blue open circles, respectively. Each dot represents weights from a bi-directional tuning curve. Each neuron contributes to four points from DS60, DS120 conditions with two arrangements: Near C-Far CC and Near CC-Far C. p-values are from one-sided signed rank test. (B) Same as (A) for monkey G.

where $R_{(An,f)pred}$ and $R_{(n,Af)pred}$ are the bidirectional response in "Attend Near" and "Attend Far" conditions. w_{nAn}, w_{fAn} and w_{nAf}, w_{fAf} are the weights of near and far components in the two attention conditions, respectively.

Fig 7 compares the weights of the near and far motion components between Attend Near (w_{nAn} and w_{fAn}) and Attend Far (w_{nAf} and w_{fAf}) conditions, obtained from Eqs 5A, 5B. Representation of the attended component could be facilitated either by simply enhancing the response weight of the attended component or by simply suppressing the response weight to the other component ($w_{nAn} > w_{nAf}$ & $w_{fAn} = w_{fAf}$ or $w_{fAn} < w_{fAf}$ & $w_{nAn} = w_{nAf}$) or a combination of the two ($w_{nAn} > w_{nAf}$ & $w_{fAn} < w_{fAf}$). We found that MT neurons in monkey B used a combination of the two to achieve the enhanced representation of the attended component (mean w_{nAn}=0.51 , mean w_{nAf}=0.45 , one-sided signed rank test p= 5.1171e-10; mean w_{fAn}=0.51, mean w_{fAf}=0.57,

one-sided signed rank test p= 3.7623e-10). On the other hand, across MT neurons in monkey G, attention to near surface only enhanced the weight of near surface (mean w_{nAn}=0.65 , mean w_{nAf}=0.60, one-sided signed rank test p= 0.023) compared to attention to the far surface, while the weight of far surface remained unchanged between the attention conditions (mean w_{fAn}=0.29, mean w_{fAf}=0.28, one-sided signed rank test p= 0.104).

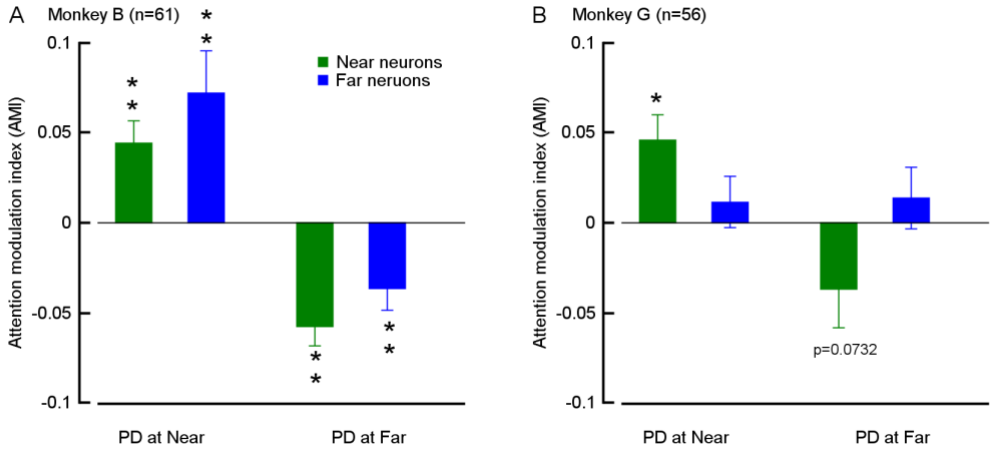

Figure 8. Effects of attention on neuronal responses to bidirectional stimuli with preferred direction (PD) at near or far: Attention modulation indices (AMI).

AMI>0 and AMI<0 imply R_{nf} (Attend Near)> R_{nf} (Attend Far) and vice-versa. (A) Monkey B: Green and blue are the mean AMI values across near and far neurons. PD at Near and PD at Far indicate the responses to bidirectional stimuli with preferred direction (PD) at near/less preferred direction at far and preferred direction (PD) at far /less preferred direction at near. (B) Monkey G. Stars indicate values significantly above or below zero: one-sided signed rank test ** p<0.001, *p<0.05. Note that the responses used for this analysis are from 200 ms after the motion onset through the end of motion (600 ms).

Attention modulation indices

We aimed to measure the sizes of attention effects on neuronal responses to overlapping stimuli constituting neurons' preferred direction (PD) at one disparity and non-preferred direction (non-PD: 60 or 120 deg away from PD) at the other disparity. The purpose of this was to determine if there are any asymmetries in the effects attention to a preferred combination of direction and disparity features vs a less preferred combination of the two features. Fig 8 quantifies the effects

of attention on neuronal responses using an attention modulation index (AMI) for each neuron defined as:

$$AMI=(R_{nf} (\text{Attend Near})- R_{nf} (\text{Attend Far}))/ (R_{nf} (\text{Attend Near})+ R_{nf} (\text{Attend Far})) \quad (10)$$

where R_{nf} is the bidirectional response to overlapping stimuli constituting neurons' preferred direction (PD) at one disparity and non-preferred direction (non-PD: 60 or 120 deg away from PD) at the other disparity. Positive or negative AMI values indicate response facilitation when attending near or far, respectively, and AMI values close to zero indicate that neuronal responses are unmodulated by attention. For monkey B, across both near and far neurons, when the preferred direction (PD) and a non-preferred direction (non-PD) were overlapping at different disparities, attention to the preferred direction selectively enhanced MT neuron's response than attention to the less preferred direction (Fig 8A mean AMI>0 for PD Near $p<0.01$ and mean AMI<0 for PD Far, $p<0.01$ one-sided signed rank test). We saw this selective enhancement of response when attending preferred direction compared to less preferred direction component of overlapping stimuli, regardless of whether PD was at near or far disparities. This suggests little influence of disparity preference on the response modulation of attention to preferred direction vs a less preferred direction.

For monkey G, like monkey B, across near neurons, when PD and non-PD were overlapping, attention to attention to PD selectively enhanced neuronal response than attention to non-PD, although only reaching near significance for "PD Far" (Fig 8B mean AMI>0 for PD Near, $p<0.05$ and mean AMI<0 for PD Far, $p=0.0732$ one-sided signed rank test). Surprisingly, across far neurons, attention to neither "PD Near" nor "PD Far" yielded significant response modulation.

Overall, for monkey B, when preferred direction (PD) and a non-preferred direction (non-PD) are overlapping at two disparities, attention to a surface with preferred direction yielded a

higher response across all neurons than attention to non-PD. In contrast, for monkey G, for near neurons, only when PD was at the preferred disparity (i.e., PD at Near), attention had a significant response modulation.

Normalization strengths

We asked to what extent neuronal responses to a preferred stimulus (resulting from a combination of direction and disparity) are influenced/normalized by adding a non-preferred stimulus in the receptive fields for the two attention conditions. We quantified the strength of normalization of bidirectional responses in relation to single component responses of the overlapping stimuli constituting neurons' preferred direction (PD) at one disparity and non-preferred direction (non-PD: 60 or 120 deg away from PD) at the other disparity.

$$NI = (R_{nf} - R_{nonPD})/(R_{PD} - R_{nonPD}) \quad (10)$$

where R_{nf} is the bidirectional response, R_{PD} is the response to component with motion in PD and R_{nonPD} is the other component. NI= 1 when $R_{nf} = R_{PD}$; NI=0 when $R_{nf} = R_{nonPD}$; NI=0.5 when $R_{nf} = (R_{PD} + R_{nonPD})/2$. Mean NI values significantly closer to 1 imply a bias to R_{PD} suggesting a max or soft-max type of operations (weak normalization effect) and values significantly below 0.5 and closer to 0 imply a bias to R_{nonPD} suggesting a loser-take-all type of behavior (strong normalization effect), and values equal to 0.5 imply component response averaging.

We found that the amount of response normalization was dependent on whether the non-PD was at a preferred disparity or a less preferred disparity relative to the PD. Fig 9 shows the NI values for two overlapping stimulus conditions: 1. preferred direction at near disparity (PD at Near) and 2. preferred direction at far disparity (PD at Far). For monkey B, for near neurons, adding a

a non-PD at a far disparity to PD at a near disparity showed a weaker normalization than predicted by response averaging i.e., NI = 0.5 (Fig 9A1 for Attend Near and Attend Far, mean NI>0.5 for PD Near $p<0.01$ one-sided signed rank test). For far neurons, adding a non-PD at a far disparity to PD at a near disparity showed a weaker normalization than the component responses averaging only for Attend Far condition (Fig 9A1 for Attend Far, mean NI>0.5 for PD Near $p<0.05$ one-sided signed rank test).

Together, for all neurons, adding a non-PD stimulus at preferred disparity to PD at less preferred disparity had a higher normalization effect on MT neurons responses than adding a non-PD stimulus at a less preferred disparity to PD at preferred disparity for different attention conditions. (Fig 9A2; "Attend pref. dsp": mean NI (PD at pref dsp) = 0.63; "Attend less pref. dsp" mean NI (PD at less pref dsp) = 0.53, one-sided signed rank test p= 2.3706e-04; "Attend less pref dsp": mean NI (PD at pref dsp) = 0.56; "Attend pref. dsp" mean NI (PD at less pref dsp) = 0.52, one-sided signed rank test p = 0.0030). This suggests a higher response normalization, in monkey B, of component responses when preferences in direction and disparity domains are pitted against each other vs not, in transparent motion stimuli. In contrast, for monkey G, across near and far neurons, adding a non-PD at near disparity to PD at far disparity had a higher normalization effect on the neuronal responses than adding a non-PD at far disparity to PD at near disparity (Fig 9B2; "Attend Near": mean NI (PD at Near) = 0.57; mean NI (PD at Far) = 0.42, one-sided signed rank test p= 6.3813e-05; "Attend Far": mean NI (PD at Near) = 0.69; mean NI (PD at Far) = 0.41, one-sided signed rank test p= 1.1107e-08). This suggests a possible "dominance" of the near surface in determining the response normalization in monkey G.

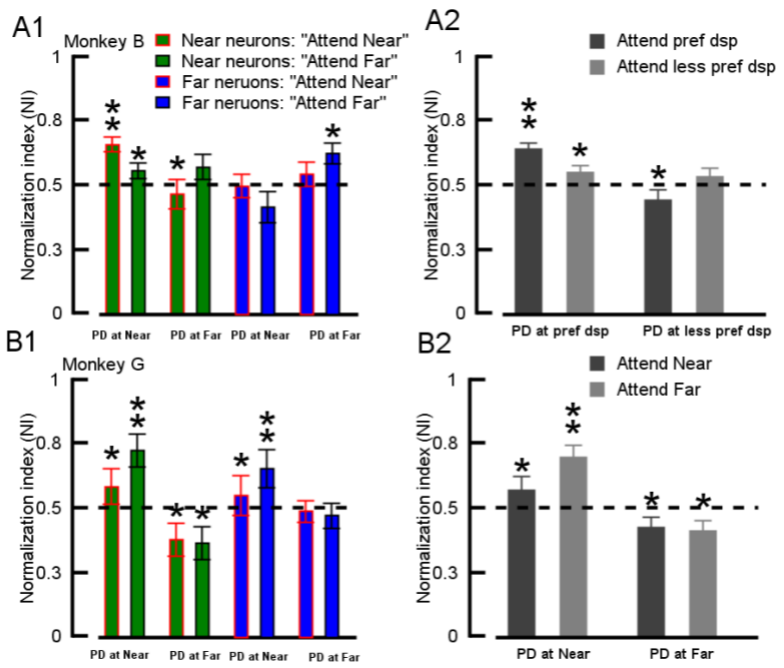

Figure 9. Normalization indices for PD at Near/non- PD at Far and PD at Far/non-PD at Near overlapping stimuli

(A1) Monkey B: Green and blue are the mean NI values across near and far neurons. Red and black borders indicate values for Attend Near and Attend Far conditions. PD at Near and PD at Far indicate the responses to bidirectional stimuli with preferred direction (PD) at near/non-preferred direction at far and preferred direction (PD) at far /non-preferred direction at near. (A2) Monkey B: Mean values across all neurons pooled and rearranged from A1. Dark and light grey colors indicate Attend preferred disparity (Attend pref dsp) and attend less preferred disparity (Attend less pref dsp) conditions. (B1) Monkey G: Same as A1. (B2) Monkey G: Mean values across all neurons pooled and rearranged from B1. Dark and light grey colors indicate Attend Near and Attend Far conditions. Stars indicate values significantly above or below 0.5 (component response averaging): one-sided signed rank test ** $p<0.001$, *$p<0.05$.

Time course of response tuning to overlapping motion directions separated in depth

We asked if the response bias to a surface at either disparity has temporal dynamics distinct from the response modulation caused by selective attention. To measure the time course of the response tuning, we characterized the population averaged bidirectional tuning curves (shown in Fig 4) using a time window of 50 ms and sliding at a step of 10 ms.

Figures 10 and 11 show the time course of the bidirectional tuning curves (red and black curves in Fig 4) averaged over near and far neuronal subgroups of monkey B and monkey G,

respectively. Row A1 to D1 shows the response in “Attend Near” and row A2 to D2 showss the response in “Attend Far” conditions. Row A3 to D3 show the time course of the response obtained by equal weighting (averaging) of the responses elicited by the individual component directions presented alone (Ravg; shown in grey in Fig 4). In the right half of the tuning curves (vector averaged direction VA>0), the motion direction at near disparity is closer to the preferred direction than the motion direction at far disparity and vice versa is true for the left half (VA<0). We refer to the right half as more preferred direction at near disparity: “mPD near” stimuli and the left half as more preferred direction at near disparity: “mPD far” stimuli.

There are notable differences in the responses between early and late time epochs across neuron subgroups and attention conditions. For example, in early time epochs, monkey B’s near neurons showed a higher response to mPD near than mPD far (Fig 10 A1, A2, C1, C2). While this trend held up in “Attend Near” condition (Fig 10 A1, C1) in the later epochs, the response gradually shifted to the far surface (mPD far) with time when attention was towards the far surface (more obvious in Fig 10 A2). For far neurons, when attention was to near surface, with time, we saw a similar shift of activity from preferred disparity (far surface) to the attended disparity (near surface) (in Fig 10 B1, initial high on mPD far and brief higher response on mPD near in the later time epochs). For monkey G (Fig 11), across time epochs, both near and far neurons, there was a higher activation to mPD near than to mPD far. Overall, temporal changes with attention were less pronounced in monkey G’s data (Fig 11). Notably, far neurons of both monkeys showed a more tonic response than near neurons.

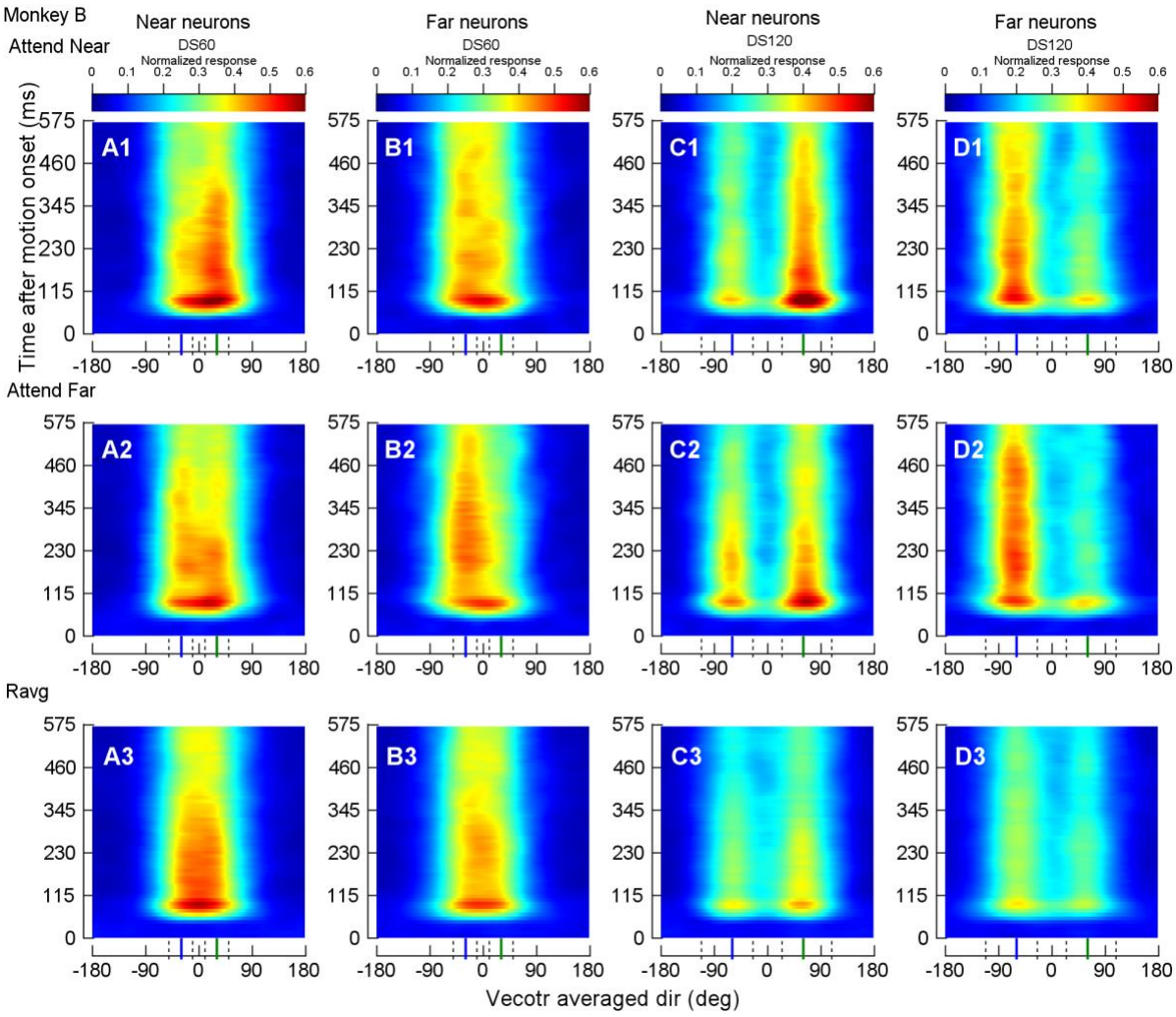

Figure 10. Time course of disparity bias in the bi-directional tuning curves of Attend Near and Attend Far conditions of monkey B.

Vector averaged (VA) direction of two motion directions and time after motion onset are plotted on x and y axes, respectively. Hotter colors indicate higher response and cooler colors indicate lower response. (A1 to A3) Responses averaged over Near neurons for DS60. (B1 to B3) Responses averaged over Far neurons for DS60. (C1 to C3) Responses averaged over Near neurons for DS120. (B1 to B3) Responses averaged over Far neurons for DS120. (A1 to D1) Responses of Attend Near condition. (A2 to D2) Responses of Attend Far condition. (A3 to D3) Average of responses of the components of each group. The green and blue lines in the bottom represent the preferred direction of the near (clockwise side) and far (counterclockwise side) component. The surrounding perforated lines indicate the direction window used in Fig 13.

Fig 12 shows the temporal evolution of three different measures of bidirectional tuning responses from Fig 10 and 11: the peak location over different VA directions, the peak amplitude,

and the center of mass (CoM) location (represents the spread of activity across VA directions, see Methods) for monkey B (Fig 12 A1-A4 and B1-B4) and monkey G (Fig 12 C1-C4 and D1-D4) from the motion onset.

For monkey B, for near neurons, the peak locations of the bidirectional tuning curves of both attention conditions were initially close to VA direction 0° and progressively shifted to 30° over a period of 80–100 ms (Fig. 12A1, red and black curves). At VA direction 30° of overlapping directions separated by 60° (DS60), the two component directions were 0° at near disparity (i.e., the preferred direction PD of the neuron) and 60° at far disparity, respectively. In other words, these neurons responded most strongly when the near component direction was close to the PD (marked by the top perforated lines in Fig 12 A1 as "PD at Near"). In contrast, far neurons responded the most when the far component direction was close to the PD (marked by the bottom perforated lines in Fig 12 B1 as "PD at Far"). This suggests an initial bias to the preferred direction at preferred disparity, across monkey B's MT neurons.

After this initial progressive shift to PD at preferred disparity, the peak locations diverged between the bidirectional curves of the two attention conditions in the later time epochs. For near neurons, when the attention was toward the far surface ("Attend Far"), in the later time epochs (300 to 400 ms after the motion onset), the peak location deviated from "PD at Near" shifting towards "PD at Far" i.e., when the preferred direction was at the attended surface (shown by the black arrow in Fig 12 A1). For far neurons, attending to near surface shifted the peak location toward "PD at Near" between 200 and 300 ms (shown by the black arrow in Fig 12 B1). In other words, attention to less preferred disparity "pulled" the peak location of population averaged responses away from the preferred disparity towards the less preferred disparity, even though for brief periods of time.

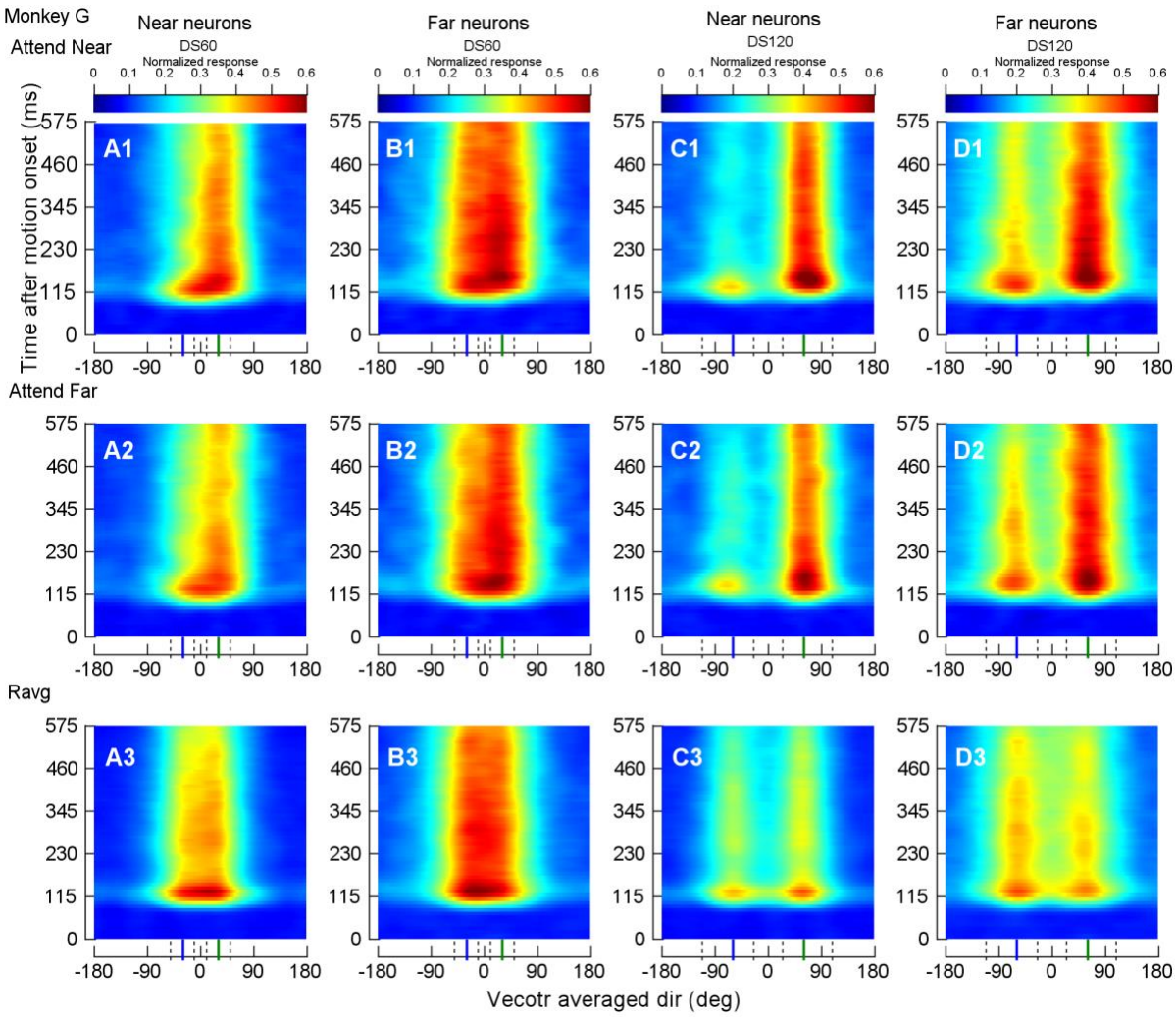

Figure 11. Time course of disparity bias in the bi-directional tuning curves of Attend Near and Attend Far conditions of monkey G.

Vector averaged direction of two motion directions and time after motion onset are plotted on x and y axes, respectively. Hotter colors indicate higher response and cooler colors indicate lower response.
(A1 to A3) Responses averaged over Near neurons for DS60. (B1 to B3) Responses averaged over Far neurons for DS60. (C1 to C3) Responses averaged over Near neurons for DS120. (B1 to B3) Responses averaged over Far neurons for DS120. (A1 to D1) Responses of Attend Near condition. (A2 to D2) Responses of Attend Far condition. (A3 to D3) Average of responses of the components of each group. The green and blue lines in the bottom represent the preferred direction of the near (clockwise side) and far (counterclockwise side) component. The surrounding perforated lines indicate the direction window used in Fig 13.

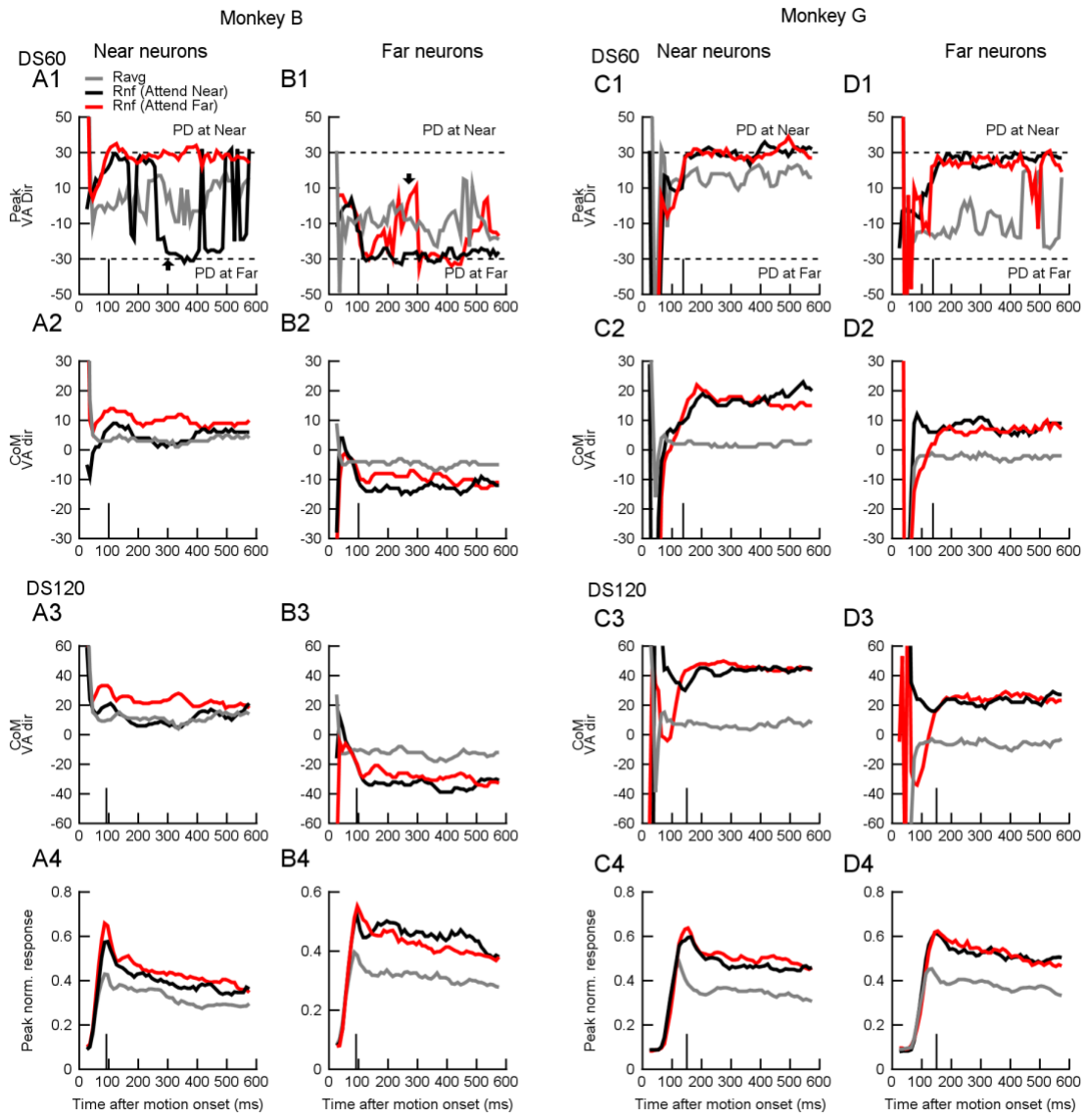

Figure 12. Time course of peak and center of mass (CoM) locations of population averaged bidirectional tuning curves

On x-axis of each panel is the time after motion onset in ms (center of the sliding time window). Red and black lines indicate the bidirectional responses to Attend Near and Attend Far, respectively. Grey lines indicate the averaged responses to single near and far components (Ravg). (A1-D1) Peak location DS60 i.e., VA direction of the motion directions that elicited the peak response of bidirectional tuning curves in each time window. Dashed lines indicate the VA directions corresponding to preferred direction (PD) at Near and PD at Far disparities. (A2-D2) CoM location i.e., the VA direction that corresponds to the CoM of the tuning curves in each time window. (A3-D3) CoM location DS120. (A4-D4) Peak response of the bidirectional tuning curves DS120. Small black vertical lines on every x-axis corresponds to the max of the peak responses over all time windows. (A1-A4) Monkey B Near neurons. (B1-B4) Monkey B Far neurons. (C1-C4) Monkey G Near neurons. (D1-D4) Monkey G Far neurons. Note that the peak locations for DS120 and the peak responses for DS60 are not shown in the figure for brevity.

In summary, for monkey B, initially, the peak location was biased to PD at preferred disparity, regardless of the attention cue. The effects of attention kicked in later (>200 ms after motion onset) where the peak location remained around PD at preferred disparity when attending the preferred disparity, and it was shifted for brief periods of time toward PD at less preferred disparity when attending the less preferred disparity. Note that, overall, the time course of peak locations of the bidirectional tuning curves is distinct from that of the tuning curves obtained by simply averaging the tuning to the individual component motions. This suggests that the bias to PD at either disparity (observed from the peaks of the bidirectional tuning curves) is systematic and functionally important for preferential and dynamic representation of the components of the overlapping stimuli.

For monkey G, across both near and far neurons, the averaged bidirectional tuning curves peaked when the PD was at near disparity for both attention conditions (red and black lines reach "PD at Near" for near neurons in Fig 12 C1 and D1) from about 150 ms through the end of motion period. Different from monkey B's data, monkey G's data suggests a persistent preferential representation of the near disparity when two disparities are overlapping in MT neurons' receptive fields. And attention to far surface did not shift the peak location toward the PD at Far. Although we did not see any differences in the time course of peak locations for near and far neurons (Fig 12 C1 vs D1), the CoM location of near neurons was closer to "PD at Near" than that of far neurons (Fig 12 C2 vs D1 and C3 vs D3). This suggests a lower extent of near bias in far neurons than in near neurons for monkey G across time epochs which is in line with the observations from monkey B's data (also see Fig 3B) throughout the motion period.

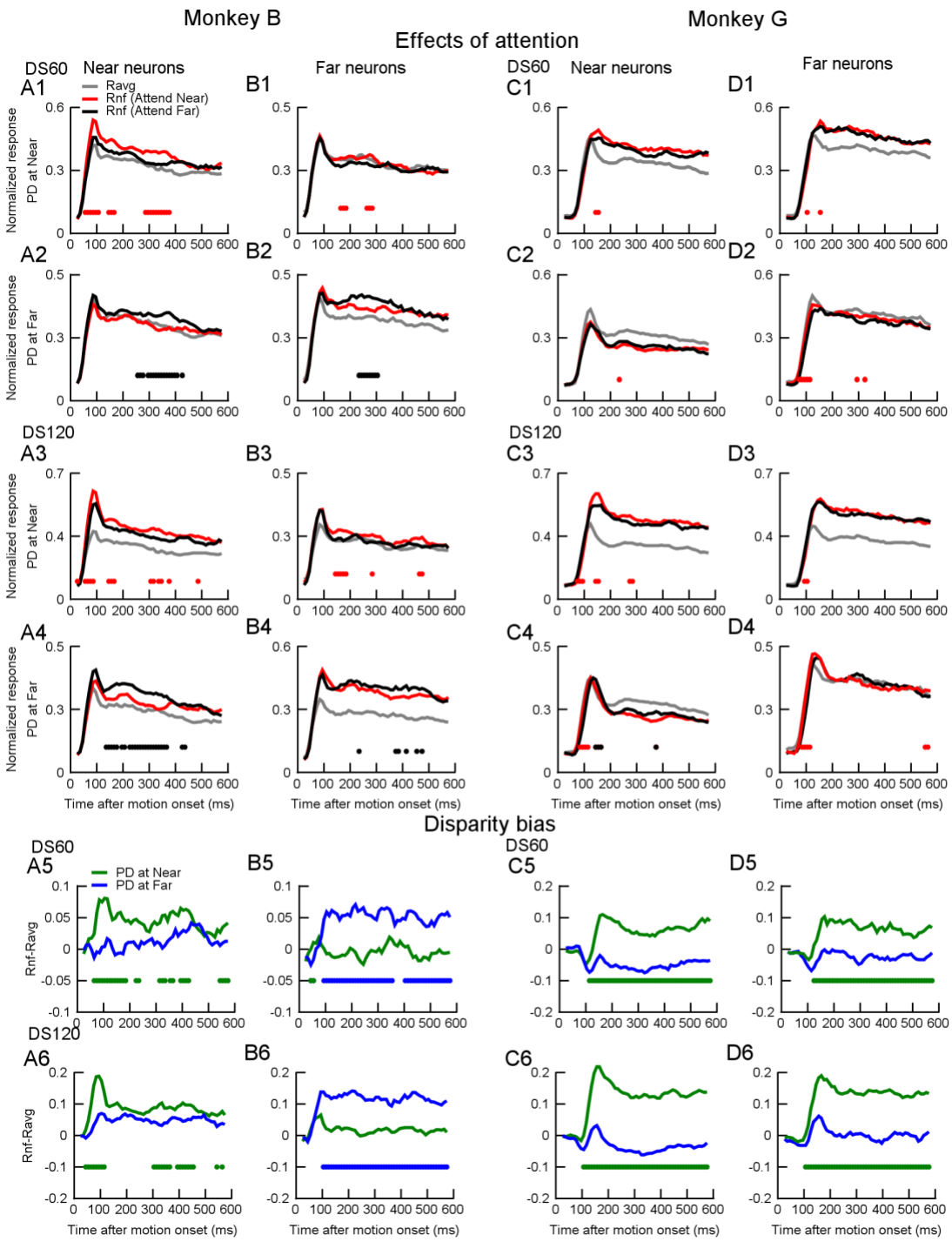

Figure 13. Time course of effects of attention on neuronal responses and disparity bias.

The responses are obtained by averaging the normalized bidirectional responses from VA=25 to 70 deg for PD at Near and VA=-70 to -25 deg for PD at Far for DS60, and from VA=55 to 100 deg for PD at Near and VA=-100 to -55 deg for PD at Far for DS120 in each time window. (A1-D1) Responses to PD at Near bidirectional stimuli (DS60). (A2-D2) Responses to PD at Far bidirectional stimuli (DS60). (A3-D3) Responses to PD at Near bidirectional stimuli (DS120). (A4-D4) Responses to PD at Far bidirectional

stimuli (DS120). Red and black curves indicate the bidirectional responses to Attend Near and Attend Far, respectively (Rnf (Attend Near) and Rnf (Attend Far)). Grey lines indicate the averaged responses to single near and far components (Ravg). Red and black dots (look like flat lines) indicate the time windows where Rnf (Attend Near) significantly higher than Rnf (Attend Far) and vice-versa, respectively showing the differential effects of attention. (A5-D5) Rnf-Ravg DS60 where Rnf= (Rnf(Attend Near)+ Rnf(Attend Far))/2. (A6-D6) Rnf-Ravg DS120. Green and blue curves indicate the values obtained from PD at Near and PD at Far stimuli as defined above, respectively. Green dots indicate the time windows where Rnf-Ravg PD at Near is significantly higher than Rnf-Ravg PD at Far i.e., near bias and blue dots indicate vice-versa i.e., far bias. Significance based on $p<0.05$ of a one-sided signed rank test.

We next asked how the responses close to the peaks of the bidirectional tuning curves evolved over time across attention conditions and angle separations. To obtain the responses close to the bidirectional tuning curve peaks, we averaged the neuronal responses over a fixed VA direction window covering the peaks: PD at Near/non-PD at Far (i.e., covering VA=30 deg for DS60 and VA=60 deg for DS120) and PD at Far/ non-PD at Near (VA=-30 deg for DS60 and VA=-60 deg for DS120) stimuli. Fig 13 shows the time course of population averaged bidirectional responses to PD at Near and PD at Far stimuli for the two attention conditions and angle separations.

Across monkey B's near neurons (Fig 13 A1 for DS60 and 13 A3 for DS120), attending PD at Near (also the preferred disparity) elicited significantly higher response than attending non-PD at far both in early (starting at 20 ms peaking at ~90 ms) and late time epochs (after 200 ms). Attending PD at Far (less preferred disparity) also elicited significantly higher response than attending non-PD at near, but this response facilitation occurred only in the later time windows (Fig 13 A2, A4, note the black significance lines are after 200 ms) for the same neurons. This indicates that for near neurons, the effects of attention occurred earlier when attending PD at near (i.e., preferred disparity) than PD at far. This suggests that when two different motion directions are overlapping with disparity cues, for near neurons, there might be a temporal advantage to the representation of PD at preferred disparity over that of PD at a less preferred disparity under

selective attention. For far neurons, like for near neurons, both attention to PD at far and PD at near resulted in response facilitation compared to attention to non-PD at the other disparity (Fig 13 B1, B2 for DS60 and B3, B4 for DS120), although they occurred later compared to near neurons. Also distinct from near neurons' trend, we did not notice any major temporal differences between the attentional effects on neuronal responses to PD at far (i.e., preferred disparity) and PD at near.

For monkey G, the effects of attention on the averaged responses were briefer and weaker compared to those of monkey B. Across near neurons, attending PD at Near (also the preferred disparity) elicited significantly higher response than attending non-PD at Far mainly only in the initial period after the motion onset (Fig 13 C1, C3). Attending PD at Far had a higher response than non-PD at Far for only DS120 for a brief time window later (300 to 400 ms) during the motion period for these neurons (Fig 13 C4). For far neurons, effects of attention were not prominent when PD was at Near (no sig effect in Fig 13 D1, very brief intermittent sig effect in Fig 13 D3). Surprisingly, for these neurons, attending non-PD at near had a significantly higher response than attending PD at far (early and late windows in 13 D2, and early window in Fig 13 D4).

Next, we evaluated the time course of disparity bias of bidirectional responses across neuronal subpopulations. In each time window, we compared the extent to which bidirectional responses (averaged over the two attention conditions indicated as Rnf) differed from the average of the responses elicited by single components alone (Ravg gray curves) for both PD at Near and PD at Near, by plotting the differences Rnf-Ravg (PD at Near) and Rnf-Ravg (PD at Far) with time after the motion onset. For monkey B's near neurons, Rnf-Ravg was significantly higher for PD at Near than PD at Far (Fig 13A5, A6) right from the early time windows (~ 70 ms) after the motion onset suggesting an early development of bias to near disparity. Across far neurons, there

was a near bias (Rnf-Ravg PD at Near > PD at Far) for a brief period after motion onset (20 to 70 ms) and the bias flipped to the far component, also the preferred component (Rnf-Ravg PD at Far > PD at Near) around 70 ms and remains the same throughout the motion period (Fig 13B5, B6). For money G, both near and far neurons showed a bias to PD at Near from 100 ms after motion onset (Fig 13E3, G3, F3 and H3) suggesting a sustained near bias regardless of neurons' disparity preference throughout the motion period.

MT neurons' responses to overlapping motion directions separated in depth with attention outside the RF: control for attentional effects

As we observed different effects of attention on the responses from monkeys G and B, to tease apart attentional effects from neuronal representational biases, we asked how MT neurons responded to the overlapping motion stimuli at two disparities when attention is directed away from their RFs. For this, we trained monkey R to perform a demanding fine direction discrimination task in the visual field opposite to that of the RFs of the recorded MT neurons (defined as "Attend away" task; see Methods). Fig 14A1 to A4 show the tuning curves averaged across near and far neuronal populations for DS60 and DS120 of monkey R. Cohesive with monkey B's data (Fig 4A to D), for both near and far neurons, when the overlapping motion directions are separated in depth, the bidirectional tuning curves are biased to the tuning curves elicited by single component at neurons' preferred disparity. Weight bias indices shown in Fig 14B confirm this bidirectional bias to preferred disparity (mean WBI near neurons=0.07 significantly positive one-sided signed rank p= 4.2989e-04 and mean WBI far neurons=-0.03 significantly negative one-sided signed rank p= 0.0030; however, WBIs of monkey R are lower than B's: see Discussion). Consistent with the trends of monkey B and G (Fig 4), for monkey R even with attention away, the CPIs are significantly positively correlated with the DBIs (Pearson's

correlation coeff. r=0.24, p= 0.0072; Type II regression line slope=1.5). This suggests that even in the "absence" of selective attention, MT neurons exhibit a representational bias (measured by DBI) to component motion at more preferred disparity when two directions are overlapping separated in stereoscopic depth, and the magnitude of this bias is proportional to the extent of their preference to one component disparity over the other (measured by CPI). Fig 14 D shows the population averaged normalization index values for monkey R. Like monkey B's NI data (Fig 8), across both near and far neurons, over different stimulus configurations, mean NI values were significantly above 0.5 suggesting a bias to the PD component, regardless of whether it was at near or far disparities.

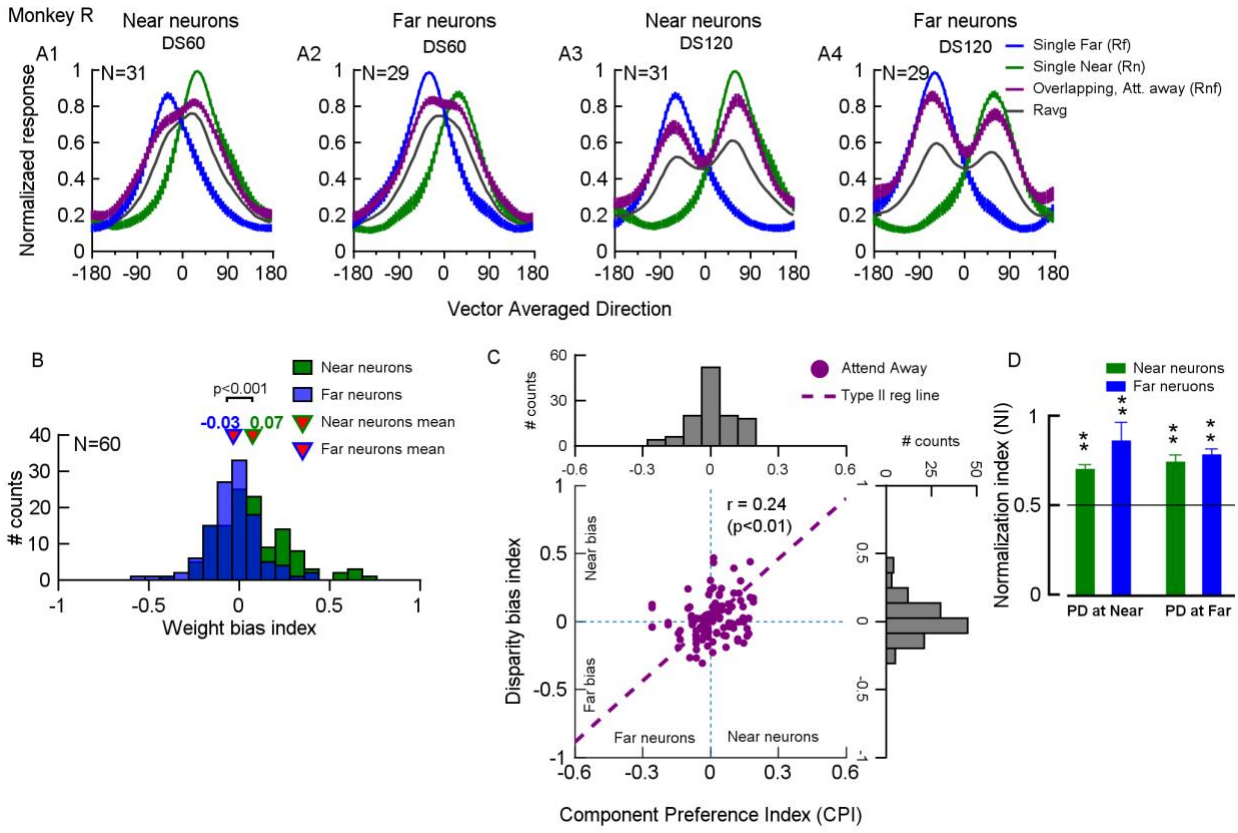

Figure 14. Attention away results from Monkey R.

(A1-A4) Population-averaged tuning curves to the single motion components and the bidirectional stimuli of Near neurons (DS60), Far neurons (DS60), Near neurons (DS120) and Far neurons (DS120). Green, blue, purple curves indicate Single near, Single far, and Overlapping (Attention Away) bi-directional tuning, respectively. The x-axis is the vector average direction of the bidirectional stimuli. The direction of 0° was aligned with the PD of each neuron before the tuning curves were averaged across neurons. (B) Weight bias index (WBI) distribution for all neurons. Green and blue indicate Near and Far neurons. Note each neuron contributes to four WBI values for DS60, DS120, Near C-Far CC, and Near CC-Far C conditions. (C) On the x and y axes are the CPIs and DBIs. Each neuron contributes to two points from DS60 and DS120 conditions. The dotted lines are the type 2 regression fits. r is the Pearson correlation with p-values in the brackets. (D) Normalization indices for PD at Near and PD at Far stimulus conditions. Green and blue indicate Near and Far neurons. ** indicates significantly greater than 0.5 with one-sided signed rank test $p<0.001$.

Normalization model fitting: capturing bias variability across monkeys

From the results above, individual MT neurons' responses to overlapping motion stimuli separated in depth could well be approximated by weighted sum of their responses to the individual components presented alone. We saw that the relationship between the weights, at a population level, depended on neuronal disparity preferences (Fig 6) and attention conditions (Fig 7, at least for monkey B). To explicitly capture the role of disparity preferences (across different motion directions) and selective attention in determining the component weights, we turned to the divisive normalization model (DNM).

In the DNM, the response to a stimulus is modulated by the summed activity generated by the stimulus itself (the stimulus drive), along with pooled neighboring responses (the normalization pool) (Carandini & Heeger 2011). Previous studies have used this framework to describe neuronal responses elicited by multiple stimuli, predicting a range of neural computations from equal summation to winner-take-all competition of the components (Britten & Heuer 1999, Heuer & Britten 2002, Busse et al. 2009, Xiao et al. 2014; Xiao et al. 2015; Wiesner et al. 2020). Other studies have modified the divisive normalization model to explain the effects of attention across a wide range of behavioral and stimulus conditions (Boynton, 2009, Lee & Maunsell 2009, Reynolds & Heeger 2009, Lee & Maunsell 2010, Ni et al. 2012). These studies proposed that attention can

alter the balance between the stimulus drive of the attended stimuli and the summed activity of the normalization pool, weakening the gain control, and thereby resulting in an increased activity.

To apply the normalization model to individual neuronal responses across a range of directions of the overlapping stimuli separated in depth, we fitted the bi-directional response tuning curves under "Attend Near" and "Attend Far" conditions separately, using the following equation:

$$R_{nf_{pred}}(\theta_n, \theta_f) = \frac{\beta S_n^N R_n(\theta_n) + S_f^N R_f(\theta_f)}{\beta S_n^N + \alpha S_f^N + \sigma} + c \qquad (11)$$

R_{nfpred} is the response to the bidirectional stimuli predicted by the model; R_n and R_f are the measured component responses elicited by the near and far components moving in directions θ_n and θ_f, respectively, when presented alone. S_n and S_f represent the "signal strengths" of the near and far components (see below), respectively, representing the weights for the excitatory drive from the stimulus components. Previous studies suggested that response normalization can be tuned, such that individual stimulus components contribute differently to normalization (Ni et al. 2012; Rust et al., 2006; also see Carandini et al. 1997). To capture this tuned normalization, we used α, a positive parameter that scales the relative contributions of S_n and S_f to the normalization pool. β is an additional positive weighting parameter for the near component (see below for detailed interpretation). N, σ, and c are model parameters with the constraints of $0<N<100$, $\sigma>0$ and $c>0$.

Most past studies have used luminance contrast values of component stimuli as signal strengths (weights of the component stimulus drives: S_n and S_f) in DNMs of a similar form to capture the dependence of the response modulation on the contrast values of the stimulus components (Busse et al.2009, Xiao et al. 2014). Our stimuli had the same contrast but differed in

both direction and disparity features, and the neuronal responses to the overlapping stimulus across motion directions depended on neuronal disparity preferences. So, we took a novel approach to replace the contrast values of the stimulus components (referred to as "signal strength") in a typical DNM equation with population activity elicited by each component summed over a pool of neurons with varying direction and disparity preferences. We propose two different ways of defining the signal strengths of the component stimuli (S_n and S_f) contributing to the normalization pool. In model 1, we assume that the population activity that influences the weight of each component's drive may be sourced by a pool of neighboring neurons that have the same disparity preference as the given neuron but have different preferred directions. Although, realistically, the population activity that can influence a given neuron's response can be coming from a pool of neurons with a range of preferred disparities. Hence, in model 2, we freely varied this range/size of the neighboring neurons in terms of their disparity preferences and found the size values that best fit the bi-directional tuning responses in our data.

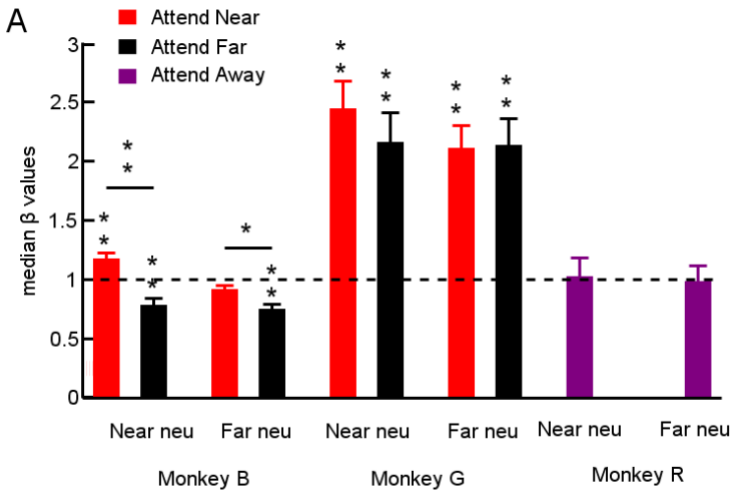

Figure 15. Normalization model parameter β for different animals.

Red and black indicate the median β values from Attend Near, Attend Far conditions, for monkeys B and G. Purple indicate the median β values from Attend Away condition for monkey R. "Near" and "Far"

indicate values across near and far neurons. * indicates significance of medians different from 1: p<0.001 of one-sided signed rank test.

Model 1. The signal strengths of the near or far component (S_n and S_f) in the "eye" of a given MT neuron are obtained by pooling the activity across neurons that have the same disparity preference as the given neuron but have different preferred directions evenly spanning 360°. This makes the signal strength of a stimulus component at a near or far disparity be invariant to the motion direction of the component, generalizing across all directions. The summed population response to a motion direction at a given disparity (i.e., one stimulus component) can be obtained by simply summing the responses of a single MT neuron to different directions evenly spanning 360° at the same disparity. Hence, we calculated S_n and S_f for each neuron using the following equations:

$$S_n = \sum_{\theta=0}^{360} R_n(\theta)\ ;\ S_f = \sum_{\theta=0}^{360} R_f(\theta), \qquad \text{(12A, 12B)}$$

where R_n and R_f are the measured component responses elicited by the near and far components moving in a direction θ, respectively, when presented alone.

Note that $S_n > S_f$ for near neurons and $S_n < S_f$ for far neurons owing to their disparity preferences. S_n/S_f quantifies the extent of this relative activation to near and far components for individual neurons. The divisive normalization model in Eq 11 enables us to define each neuron's component weights (w_n and w_f from Eq. 1 in Methods) in terms of its relative activation to the components shown by the equations below.

$$w_n = \frac{\beta S_n^N}{\beta {S_n}^N + \alpha {S_f}^N + \sigma} \qquad (13)$$

$$w_f = \frac{\beta S_f^N}{\beta {S_n}^N + \alpha {S_f}^N + \sigma} \qquad (14)$$

If we set the exponent N=1 in Eq. 13 and 14, $\beta = \frac{w_n/w_f}{s_n/s_f}$ i.e., $1/\beta = \frac{w_f/w_n}{s_f/s_n}$

β factors any additional scaling of near weight relative to far weight, besides the effect of one component having a higher signal strength than the other (S_n>S_f for near neurons and S_n<S_f for far neurons). β greater than 1 indicates relative favoring of near weight and β less than 1 indicates relative favoring of far weight.

Eq. 11 is akin to the tuned normalization model used by Ni and colleagues (Ni et al. 2012) to explain the effect of attention on the neuronal response to multiple stimuli within the RF. In equations 3A and 3B of Ni et al. (2012), there is also a term β in the numerator, which increases the weight of the attended stimulus component. On the other hand, a previous study from the lab (Wiesner et al. 2020) suggested that β can also represent biased stimulus drive (either via feedforward or lateral connections) that favors one stimulus component of overlapping stimuli under task conditions that require no selective attention (passive fixation). Here we interpret β as a scaling factor for one component over the other (here near over far as β is coefficient of Rn) resulting from attention to one component or a feedforward drive biased to the favored component or a combination of both.

This model fits the data from three monkeys reasonably well (mean PV values=89% for monkey B, 81% for monkey G, and 80% for monkey R). Fig 15 shows the median β values obtained from the model fits across near and far neurons for the three animals under different attention conditions (Attend Near, Attend Far and Attend away). For monkey B, across near neurons, attention to near showed a complete facilitative effect (i.e., β >1) on the near weight (median β >1 one-sided signed rank p<0.001). Across far neurons, attention to far showed a complete facilitative effect (i.e., β <1) on the far weight (median β <1 one-sided signed rank

$p<0.001$). This suggests that attention to a more preferred disparity facilitated the weight of the more preferred disparity component more than what would be expected by the ratio between the signal strengths of the components (S_n and S_f). For both near and far neurons, median β values in "Attend Near" condition were significantly greater than "Attend Far" condition (Near neurons: one-sided signed rank test p=2.0692e-08; Far neurons: one-sided signed rank test p=0.0016). This implies that attention to near component significantly scaled the weights more in favor of the near component compared to attention to far. In contrast, for monkey G, across both near and far neurons and attention conditions, there was a strong facilitation of the near weight (i.e., $\beta > 1$). As attention to far surface did not alter this strong facilitation of near weight, we interpret this as a biased stimulus drive of the near component (see Discussion). For monkey R, with attention away from the receptive fields, there was no additional scaling of near weight relative to far weight (β not sig. different from 1) besides the effect of one component having a higher signal strength than the other. This might suggest that, while attention is away from the RFs, differences in population activity elicited by the individual stimulus components could very well explain the scaling of near and far weights with respect to each other i.e., near neurons showing near bias and far neurons showing far bias in their bi-directional responses.

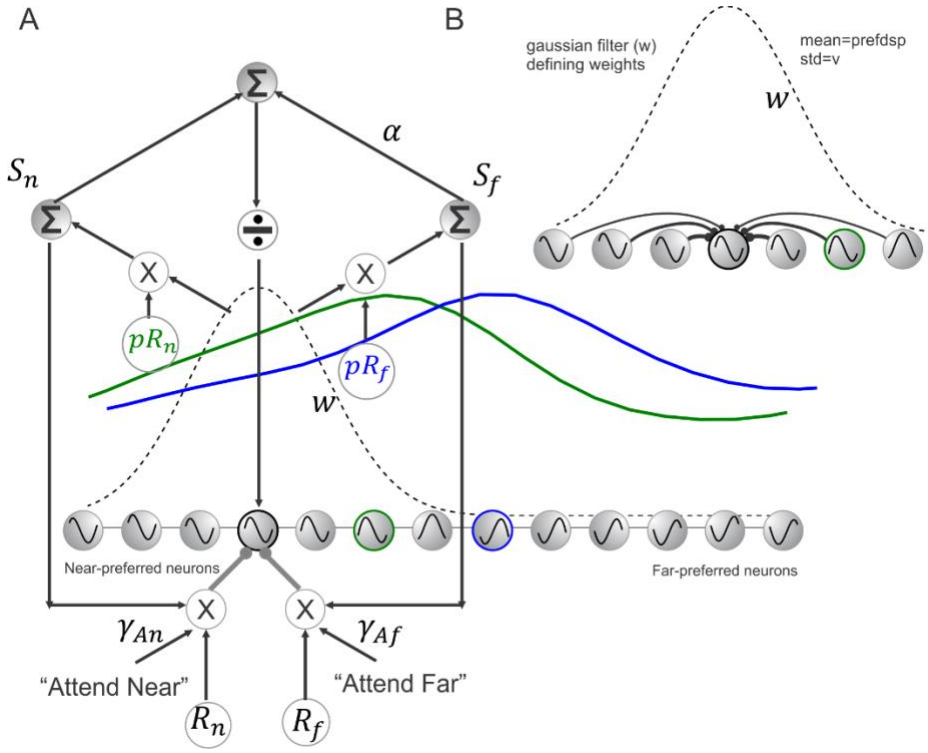

Figure 16. Cartoon depiction of implementation of Model 2.

(A) Each small grey circle represents a neuron and the black curve inside the circle indicates its disparity tuning. $R_n$and R_f are the inputs to a given neuron (shown in black circle) amplified by attentional gain factors γ_{An} and γ_{Af} in Attend Near and Attend Far conditions, respectively. S_n and S_f are additional gain factors coming from the summed activity of a pool of neurons. These summed activities are obtained by convoluting a weight matrix (w) with population responses to near and far stimulus components alone, given by pR_n (green curve) and pR_f (blue curve), respectively. S_n and S_f also contribute to the normalization signal, shown by the division symbol. (B) Weights (w) defined by a Gaussian with its mean equal to the preferred disparity of the given neuron, and standard deviation v denoting the size of the "normalization pool".

Model 2: For a given MT neuron, the bidirectional response weights to the stimulus components can depend on two factors: 1. how close its preferred disparity is from the stimulus component disparities and 2. how broad or narrow is the range of preferred disparities of the neurons whose summed activity influences a given neuron's responses. For example, a neuron with a preferred disparity closer to one stimulus component disparity than the other is likely to show a higher response bias to the former component than the latter. The extent of this bias to the component closer to the preferred disparity might be modulated by the activity of neighboring pool

of neurons with a range of disparity preferences: a neuronal pool with a narrow range of preferred disparities, which reflects the local preferences in signal strengths, might increase the bias to the preferred component (like in model 1). Contrastingly, a broader pool covering all preferred disparities reflecting the preference of the "whole" neuronal population to the stimulus components (which could be different from the local preference), might tip the bias towards the stimulus component that is preferred at a population level.

To test how these two factors might shape the bidirectional response weights in our data, in equation 4, we defined the signal strength of a stimulus component at a given disparity in the "eye" of a MT neuron as the sum of elicited responses of neighboring MT neurons with a range of preferred disparities. On one extreme, this pool of neurons entails the "whole" neuronal population covering a broad range of preferred disparities (i.e., $v = inf$), and we define the component signal strengths as $S_n = sum(pR_n)$ and $S_f = sum(pR_f)$ where pR_n and pR_f are the responses elicited by -0.1° disparity (near) component and 0.1° disparity (far) component, respectively, of a neuronal population with preferred disparities ranging from -1.6° to 1.6° (dataset from DeAngelis and Uka, 2003; Fig 17A). This is referred to as pure "global pooling" (orange lines in Fig 17 A for two example $prefdsp$ values). On the other extreme, the pool only covers the neurons with preferred disparities same as the given MT neuron (i.e., $v = 0$) i.e., $S_n = sum(pR_n(prefdsp))$; $S_f = sum(pR_f(prefdsp))$, referred to as pure "local pooling" (brown lines in Fig 17 A for two example preferred disparity $prefdsp$ values). We asked where in this spectrum are the neuronal pool sizes (defined by v) that best describe the bi-directional responses of each neuron in our dataset. For this, we used a weighting function (w in Fig 16) defined by a gaussian function centered on the preferred disparity of the neuron ($prefdsp$) with a standard deviation of v: $w = gauss(prefdsp, v)$. S_n and S_f for each neuron are defined using the following equations:

$$S_n = sum(pR_n * w)\ ; S_f = sum(pR_f * w), \quad \text{(15A, 15B)}$$

which are plugged into the DNM in eq. 11 to obtain model fits of the bidirectional tuning curves.

Fig 17 B shows the relationship between v parameter and the area covered by gaussians with different v values as a fraction of area covered by v in "pure global pooling" centered on two example preferred disparity values 0° (solid lines) and -1.1° (dashed lines). Note that this relationship changes depending on where the gaussian is centered i.e., the preferred disparity of the given neuron. Since the difference in the area between two v values could be larger at lower v values than at higher v values (area vs v function steeper at lower v values and plateauing at higher v values in Fig 16B), we divided the search space for parameter v to result in equal steps of the area covered by the gaussians for model fitting, instead of equal steps of v.

We derived the values of the scaling factor β for "Attend Near" and "Attend Far" bidirectional responses separately ($\gamma_{An}$and γ_{Af}, respectively) from the weights of LWS model fitting described in Eq. 8 and 9, shown below again:

$$R_{(An,f)pred} = w_{nAn}R_n + w_{fAn}R_f + c$$
$$R_{(n,Af)pred} = w_{nAf}R_n + w_{fAf}R_f + c$$

If we assume,

$$w_{nAn} = w_n * \gamma_{An}$$
$$w_{fAn} = w_f * 1$$

$$w_{nAf} = w_n * 1$$
$$w_{fAf} = w_f * \gamma_{Af}$$
$$\gamma_{An} = \frac{w_{nAn}}{w_{nAf}}\ ;\ \gamma_{Af} = \frac{w_{fAf}}{w_{fAn}} \quad \text{(16A, 16B)}$$

In Eq. 5A and 5B, $R_{(An,f)pred}$ and $R_{(n,Af)pred}$ are the bidirectional response in the two attention conditions: "Attend Near" and "Attend Far". w_{nAn}, w_{fAn} and w_{nAf}, w_{fAf} are the weights of near and far components in the two attention conditions, respectively. In Eq. 6, we assume that attention to a component of bidirectional stimuli only scales the weight of the attended component (scaled by γ_{An} and γ_{Af} factors), but not the weight of the other/un-attended component (scaled by a factor of 1). Put together, the normalization model takes the following form with β being a coefficient of response to the component being attended (γ_{An} for "Attend Near" and γ_{Af} for "Attend Far" in Eq. 16A and 16B):

$$R_{(An,f)_{pred}} = \frac{\gamma_{An} * S_n^N * R_n + S_f^N * R_f}{\gamma_{An} * S_n^N + \alpha * S_f^N + \sigma} + c \qquad (17A)$$

$$R_{(n,Af)_{pred}} = \frac{S_n^N * R_n + \gamma_{Af} * S_f^N * R_f}{S_n^N + \alpha * \gamma_{Af} * S_f^N + \sigma} + c \qquad (17B)$$

where $S_n = sum(pR_n * gauss(prefdsp, v))$; $S_f = sum(pR_f * gauss(prefdsp, v))$.

Fig 16 shows the implementation of this model. This model fits the bidirectional responses from monkey B and G reasonably well (mean PV=89 % for monkey B, 80% for monkey G). Fig 17 C1 to C3 show the probability distributions of v values obtained from the model fits across near, far and all neurons combining the two attention conditions for each monkey (the error bands show 95% confidence intervals across each indicated set of neurons). To represent the relativity of different v values with respect to v values of pure global pooling, instead of plotting the v values on the x-axis, we plotted the area of gaussian with standard deviation of v as a fraction of area of gaussian in pure global pooling case (see Fig 17B).

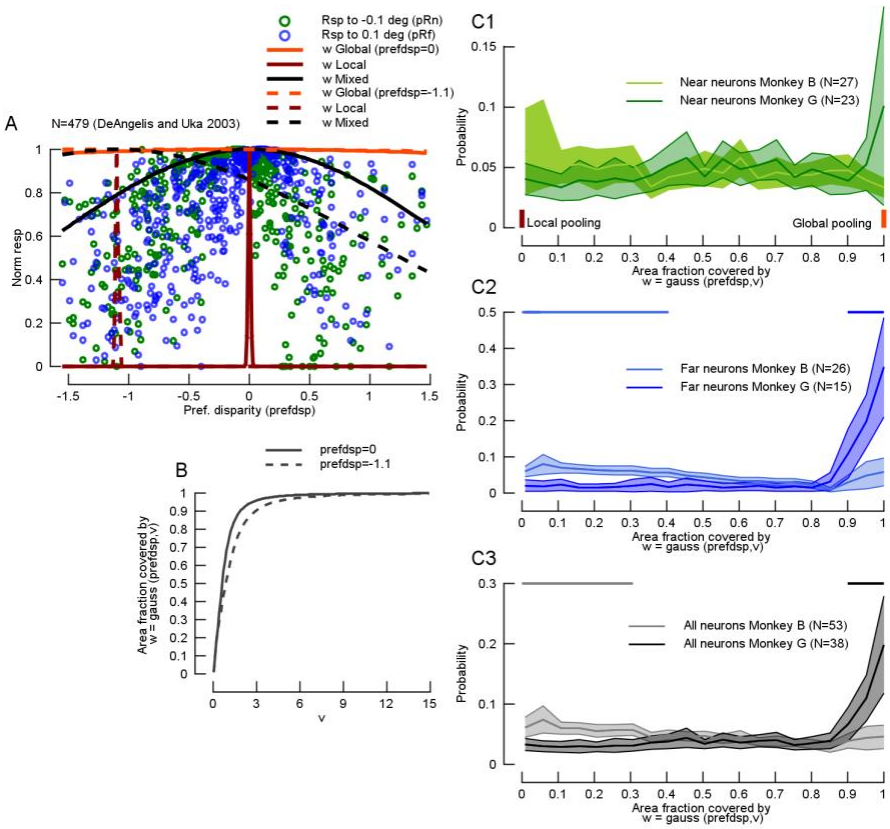

Figure 17. Distribution of parameter v from model 2 fits from Monkey B and G.

(A) Green and blue dots show the responses of N=479 MT neurons (DeAngelis and Uka 2003) to -0.1° and +0.1° disparities moving in their preferred directions, denoted as pR_n and pR_f. Orange, brown, and black lines show examples of Gaussian weight (w) filters of global, local, and mixed pooling strategies. Solid and dashed lines are for example neurons with preferred disparity of 0° and -1.1°, to illustrate the effects of preferred disparity on the amount of activity captured by convolving w with pR_n and pR_f . (B) Relationship between the fraction of area covered by the global pooling Gaussian weight filter and the standard deviation for the two example neurons with preferred disparity of 0° and -1.1° shown in (A). (C1) Probability distributions of the model fits suggesting local and global pooling on the left and right ends of the x-axis for near neurons. Light and dark greens lines indicate the probability values from fit of monkey B and G's near neurons. The error bands are 95% confidence intervals. (C2) Far neurons. Light and dark blue curves are for monkey B and G. (C3) All neurons. Light and dark grey curves are for monkey B and G. The flat lines indicate the portions of significant difference between the monkeys.

While there was no significant difference between the v value distributions from model fits of near neurons between the monkeys (Fig 17C1), monkey B's fits across far neurons resulted in higher occurrence of v values closer to the local pooling (area frac<0.4) than those of monkey G (Fig 17C2). Monkey G's far neurons showed a higher frequency of v values closer to global pooling (area frac>0.9) than monkey B's far neurons indicating a more likely 'global pooling' in

monkey G's case. This means that monkey G's far neurons pooled inputs over neurons with wider range of preferred disparities. This broad normalization pool might explain the near bias of the bidirectional response seen in monkey G's data, which is against their preference to single component i.e., the far component. In contrast, the more likely local pooling of monkey B's far neurons might be in line with the far bias of the bidirectional responses indicating the local preference to far disparity.

Changes in vergence can't explain our results

As MT neurons are selective to disparity, their responses might be affected by changes in the depth of fixation (vergence eye movement). If vergence eye movements were linked to the behavioral choice in our task, this might trivially explain the attentional effects on neuronal responses. For example, one likely scenario is, to report the motion direction of the cued surface, the monkeys could have shifted their gaze to the cued disparity i.e., to the near and far surfaces of the overlapping stimuli in "Attend Near" and "Attend Far" conditions, respectively. To rule out this possibility, for each session with monkey's binocular eye positions recorded (Monkey B: 20 out of 49 sessions), we performed an ANOVA, and found that significant differences between the trial-averaged vergence values (left-right horizontal eye positions) of "Attend Near" and "Attend Far" trials were rare (Monkey B: 15%). For each session with only the left eye position recorded (29 sessions of monkey B and 47 (all) sessions of monkey G), we performed an ANOVA, and found that significant differences between the trial-averaged left eye horizontal positions of "Attend Near" and "Attend Far" trials were rare (monkey B: 6%; monkey G: 21%). A critical test would be to see if the change in vergence between the attention conditions was consistent with the attention cue i.e., a significant positive difference in mean vergence or left eye position values between "Attend Near" and "Attend Far" trials. Only a small proportion of the sessions had the

direction of the vergence shift (or the horizontal position of the left eye) consistent with the attention cue (monkey B: 6%; monkey G: 8% of the all the sessions). Together, these results suggest that vergence changes had a negligible effect on our findings.

Discussion

We investigated the role of binocular disparity and selective attention in the neural encoding of two transparently moving surfaces. We designed a novel motion discrimination task where macaque monkeys were cued using disparity to attend one of the two moving surfaces located at different depths. We found at least two schemes of how area MT encodes transparently moving surfaces separated in depth, both of which could be beneficial to the segmentation of the component surfaces. In one encoding scheme (seen in monkey B and R), in response to transparent motion surfaces separated in depth (one at a near disparity and the other at a far disparity), MT neurons showed a bias to the motion of the component surface closer to their preferred disparity. This response bias might imply that subgroups of neurons defined by their disparity preferences (near and far neurons) selectively pool the responses elicited by the "preferred" component surface over the other surface. This selective pooling by different neuronal subgroups suggests a distributed representation of component surfaces of the overlapping stimuli. This distributed encoding could help in segregating information about the stimulus features of each component possibly aiding in discriminating the overlapping motion directions. In the second encoding scheme (seen in monkey G), the responses of MT neurons to overlapping stimuli separated in depth were always biased to the motion of the near surface. Although, the extent of this bias to the near surface was higher for a subgroup of neurons with preferred disparities closer to the near surface (near neurons) than for the subgroup of neurons with preferred disparities closer to the far surface (far neurons). This subtle but significant difference in the extent of response bias between

near and far neurons was likely enough to enable the segmentation of the surfaces, evident from the decent performance of the animal on the direction discrimination task using depth cues across sessions. The disparity bias (i.e., response bias to the motion at one disparity over the other) we observed in MT responses to overlapping motion surfaces separated in depth, can be deemed as a neural signature of the segmentation of the two surfaces.

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

Previous studies showed that the neuronal response elicited by two stimuli within the RF can be described as a weighted sum of the responses elicited by the individual stimulus components (Qian and Andersen 1994, Recanzone et al.1997, Britten & Heuer 1999, Zoccolan et al. 2005, Xiao et al. 2014, Xiao et al. 2015, Buss et al. 2009, Wiesner et al. 2020). Some of these studies have looked at what attributes of stimulus components might determine their response weights, especially when the components differed in one or more features. For example, Xiao et al. 2014 found that rather than weighting each stimulus component equally, MT neurons weighted the stimulus component that had a higher signal strength more strongly, regardless of whether the signal strength was defined by motion coherence or luminance contrast. Analogously, in our study, the direction tuning curves of the near and far component surfaces had the same selectivity but usually different amplitudes owing to the disparity tuning of MT neurons. In other words, neurons that preferred -0.1° disparity more than 0.1° show a higher response amplitude in the direction tuning to near component motion than the far component motion (near neurons), and vice-versa (far neurons). This natural distinction of two neuronal ensembles that preferred one component surface disparity over the other (across motion directions) when presented alone could be exploited by area MT to distribute the representation of the two surfaces when overlapping in the receptive fields over the two neuronal groups, using higher response weights for the component at preferred

disparity. Supporting this idea, we found a significant positive correlation between the MT neuron response differences to component disparities (CPI) and the difference of response weights of the bi-directional tuning curves (DBI) (Fig 6 and Fig 14). This provides evidence that the segregation of MT activity into near and far-preferred neuron subpopulations is crucial for motion discrimination using depth cues. While these neurons, at a subpopulation level, showed larger response weights to one component than the other, it is important to note that neurons within a subpopulation exhibited a range of response weights. This diversity of response weights within each neuronal group might make the visual system more robust to noise, hence increasing the encoding accuracy of multiple stimuli, compared to a system with homogeneous response weights within neuronal groups (Orhan & Ma, 2015).

Divisive normalization: taking contrast out of the equation and defining the normalization pool

We used a divisive normalization model (DNM) framework to capture the heterogeneity of response weights across neurons as a function of their preferences to the components of overlapping stimuli when presented alone. The preferences were reflected by the population neural responses to individual stimulus components, with additional scaling from selective attention. Several studies have used the normalization model to describe visual neurons' responses to two stimuli in their RFs that had different contrasts (Heuer & Britten 2002; Busse et al.2009, Xiao et al. 2014). The influence of adding a second stimulus to the first, on the neuronal responses, depended on the contrast values of the stimuli. To capture this contrast dependence, these studies used contrast values as gain factors/weights for the excitatory drive from the stimulus components (the numerator) and in the normalization signal (the denominator). Our stimuli had the same contrast but differed in both direction and disparity features, and the neuronal responses to the overlapping stimulus across motion directions depended on neuronal disparity preferences. So, we

took a novel approach to replace the contrast values of the stimulus components (referred to as "signal strength") in a typical DNM equation with population activity elicited by each component summed over a pool of neurons with varying direction and disparity preferences. Additionally, what sets our model apart from the previous adaptations of DNM, is the flexibility to define and vary the extent/size of the pool of neurons (defined in terms of the range of the preferred feature covered i.e., disparity here), that is contributing to the normalization signal. This redefined DNM model, which is currently a fusion of mechanistic and descriptive approaches, can be refined in future into a biologically plausible framework.

Preference to the near surface of overlapping stimuli

In monkey G, MT neurons showed greater response weights to the near motion component than the far motion component of the overlapping stimuli, regardless of the neurons' disparity preference and the attention condition. In addition, we found that the monkey had a better performance in discriminating motion directions at the near surface than at the far surface. Why would there be such a strong bias toward the near component at both neuronal and behavioral level?

Hibbard and Bradshaw (1999) reported that the coherence threshold in detecting motion direction at near plane is lower than that at far plane, which provides a possible explanation for the better detection of the motion direction of the near component for monkey G. But this was not the case for monkey B who performed equally well discriminating the motion directions of near and far surfaces. It is possible that monkey G used a task strategy quite different from that of monkey B. For example, on most "Attend Far/Cue Far" trials, monkey G could have possibly majorly attended the more "salient" near surface while dividing or switching his attention to the far surface within a trial, resulting in a poorer performance on "Attend Far" trials than "Attend Near" trials.

This could also explain why we did not see an increase in weights of the far surface in “Attend Far” condition compared to those in “Attend Near" condition, across monkey G’s MT neurons, making the effects of attention to far surface on the neuronal responses minimal.

Regardless of the task strategy used by monkey G, it is surprising that the far-preferred neurons showed a higher response weight to the near surface than to the far surface of the overlapping stimuli, which is against their preference to far disparity over the near when presented alone. This is in contrast with the higher response weight to the far surface compared to the near surface for monkey B’s far neurons. We speculate a few possible reasons for this. One reason is a possible boosting of the stimulus drive of the near surface of the overlapping stimuli either via the feedforward inputs to, or the lateral inputs from neurons with different disparity preferences within area MT in monkey G (see results from DNM Model 1). Previous studies reported a higher number of binocular neurons that prefer near (crossed) disparities than far (uncrossed) disparities across different brain areas like V1 (Gonzalez et al. 2010, Prince et al. 2002, Samonds et al. 2012), V2 (Prince et al. 2002, Chen et al. 2008), V3 (Adams et al., 2001), V4 (Hinkle et al. 2005, Tanabe et al. 2005), MT (Maunsell & Van Essen 1983, Bradley & Andersen 1998, DeAngelis & Uka 2003), and MST (Gonzalez et al. 2001). This higher proportion of near-preferred neurons at different stages of visual hierarchy might point to an increased stimulus drive elicited by the near component to both near and far-preferred neurons in monkey G. We also showed that this biased stimulus drive, especially to far-preferred neurons, could be explained by pooling inputs over neurons with a broad range of preferred disparity (see results from Model 2).

If a biased stimulus drive is a possible reason for the near bias seen in monkey G, why did we not see this strong bias to near component in monkey B’s far neurons (and monkey R’s)? We suspect this individual difference may be due to different task strategies that animals employed,

which may change how the brain relies on signals from different neuronal populations to solve segmentation. One possible reason for this individual animal difference may come from training history. During the initial training, monkey G was trained exclusively on the "Cue Near" conditions first and then was introduced to the "Cue Far" conditions, while monkey B was trained using a balanced set of "Cue Near" and "Cue Far" trials. This training history might have led monkey G to rely on the near disparity cues more than the far cues, which could have resulted in a higher response weight to the near surface of the overlapping surfaces. Along similar lines, previous studies have reported a training history dependence of either a correlation (Uka & DeAngelis 2004; Lui and Pack 2017; Yan et al., 2014) or a causal link (Chowdhury & DeAngelis 2008; Nienborg & Cumming 2007) between MT activity and disparity discrimination. Although, as the neuronal response biases between monkey G and B are substantially different, whether such a difference can develop during a relatively short time scale of behavioral training (6 months) is an open question. Future work would be required to examine whether training history has a significant influence in shaping the response weights of the components, which in turn might determine which MT signals are preferentially weighted to generate segmentation judgments in a flexible manner.

Effects of attention on neuronal responses

Several studies in the past reported that attention selectively enhances the neural representation of attended stimuli and reduces the influence of unattended stimuli (Treue & Maunsell 1996, Ferrera & Lisberger 1997, Reynolds et al. 1999, Treue & Martínez Trujillo1999, Recanzone & Wurtz 2000, Li & Basso 2005, Lee & Maunsell 2010). In line with these reports, our results suggest that selectively attending to one of the two overlapping surfaces moving in different directions presented at two binocular disparities enhanced the representation of the

attended component in area MT. This enhancement of representation of the attended component could be achieved either by simply enhancing the response weight to a component or by simply suppressing the response weight to the other component or a combination of the two. We found that MT neurons in monkey B used a combination of the two to achieve the enhanced representation of the attended component, while monkey G's MT neurons enhanced their response weight of the near surface when near surface was attended leaving the weight of the far surface unchanged. Hence, our results suggest at least two neural schemes of how selective attention to one component surface can affect the encoding of multiple motion directions with disparity cues, both of which, nevertheless, can aid in segmenting the surfaces overlapping in MT neuron receptive fields.

Disparity-based or Object/surface-based attention?

An early psychophysics study (Snowden and Rossiter 1999) demonstrated that the role of stereo depth cues in transparent motion processing is distinct from other visual cues like color, in the sense that, compared to the color cues, depth cues gave a clearer impression of separate surfaces, which may then be selectively processed through an attentional mechanism. This emphasizes the importance of forming a percept of surfaces for the segmentation process. Is this surface-based perception important in our study? In our task, the monkeys selectively attended to the depth cues to report the motion direction of each surface. This process can be broken down to three steps: 1. conjugate the disparity and direction features forming a percept of a surface, 2. matching the disparity cue with the disparity of one of the overlapping surfaces, 3. retrieving the motion direction feature of that surface. It is important to note that even though disparity cues aided in the selection of one of the overlapping surfaces, only disparity-based attention cannot explain the effects of attention on responses to preferred vs non-preferred direction seen in our

data (from the AMI analysis in Fig 8). Selective attention needs to spread from the disparity to direction domain for the monkey to successfully perform the direction discrimination task using the disparity cues. We believe that this attention spread relies on perceptual grouping of the direction and disparity features via surface-based mechanisms. Hence, the unit of selection for attention in our task is a surface defined by its disparity rather than the disparity feature alone, which posits our findings under the umbrella of "surface/object-based" attention mechanisms reported previously (Wannig et al., 2007). Our study is one of a kind to probe the effects of attention on the neural representation of surfaces defined by two features (motion and disparity) that are conjointly encoded in one brain area (area MT).

Time course of disparity bias and attentional effects: asymmetry between near and far-preferred neurons

It is debatable whether attention is directly involved in segmenting two objects initially or if the attention spreads over already segmented object representations and enhances the representation of the attended object. In this study, we deem the disparity bias in MT responses to two overlapping moving surfaces separated in depth, as a neural signature of segmentation of the two surfaces. Hence, we were interested in examining if the disparity bias (a neural signature of segmentation) of our neuronal responses occurred before the effects of selective attention. In monkey B, amongst far neurons, far disparity bias occurred earlier (~70 ms; see Fig 13 B5 and B6) than the effects of attention (~210 ms; see Fig B2 and B4). For near neurons, it is hard to dissociate the occurrence of the two, as both near disparity bias and the effects of attention on population neuronal responses appeared concurrently soon after the motion onset (disparity bias ~70 ms: Fig 13 A5 and A6; attention effects peaking at ~90 ms but starting at 20 ms: Fig 13 A1 and A3). In monkey G, the effects of attention on neuronal responses were minimal rendering the

data not suitable for studying the temporal dissociation between disparity bias and attention effects. Hence, at least in one subgroup of neurons, we saw evidence toward the possibility that the neural signatures of disparity and motion aided surface segmentation precede attentional modulation of neuronal responses.

Across far neurons, there was a near bias for a brief period after motion onset (20 to 70 ms) and the bias flipped to the far component (also the preferred component) around 70 ms and remains the same throughout the motion period. The effects of selective attention on the population averaged responses occurred the earliest, for near neurons (peaking at 90 ms see Fig 13 A1 and A3) compared to those for far neurons (210 ms see Fig 13 B2 and B4). Does this suggest a possible temporal advantage to encoding of near surface motion by near neurons compared to encoding of far surface motion by far neurons? Some evidence toward this exists in human psychophysics. Patterson and colleagues (1995) reported that observers showed a lower duration limits and depth discrimination thresholds for crossed (near) disparities than for uncrossed (far) disparities. These results suggest that crossed and uncrossed mechanisms differ in terms of their temporal integration properties at processing levels involving the computation and discrimination of depth. Further experiments are needed to pinpoint the underlying neural mechanisms.

Attention away

We found that even with the attention away from the receptive fields of MT neurons, the neural representation of the transparent motion stimuli separated in depth was distributed over two subpopulations defined by their disparity preferences i.e., near-preferred neurons showed a higher response weight to the near component and far-preferred neurons showed a higher response weight to the far component of the overlapping motion stimuli, akin to monkey B's attention task data. Although, we did find a decrease in the differences between the component response weights (i.e.,

the extent of bias to either component) for both near and far neurons compared to that of monkey B's data from the attention task. This implies that MT neurons' responses to transparent motion stimuli with depth cues, across motion directions, are biased to the motion surface closer to the preferred disparity ("preferred component"). The reduced magnitude of response bias, without attention to the RFs found in monkey R, suggests that selective attention may enhance this bias further. Studies are underway in our laboratory to expand our findings towards a set of follow up questions focusing on the role of disparity and attention on motion segmentation. One such effort is towards understanding the modulation of elicited neuronal responses by changing the task strategy. For example, one variant of the selective attention task in the current study is to engage a general form of attention inside the RFs by requiring the animal to distinguish between two surfaces and one surface, without cueing the monkey to either surface. This might result in a neuronal response bias that is intermediary between the selective attention task and the attention away task of the current study.

Chapter 3: Neural representation of multiple spatially separated speeds in cortical area MT

Introduction

In daily vision, several objects might be closely situated in space. Of the various cues that help our visual system to segment objects apart, relative motion is a powerful one. For example, previous studies have shown that speed difference alone, in the absence of direction cues (i.e., moving in the same direction) could be a compelling cue for the segregation of both overlapping and non-overlapping surfaces (Masson et al. 1999, Andersen, 1989, Bravo & Watamaniuk 1995). Thresholds for detecting the presence of two speeds is lower for non-overlapping (5–10% speed difference) than for overlapping sets of dots (20–30% speed difference) as measured by separate studies (Mestre et. al., 2001, Masson et.al. 1999, Sachtler & Zaidi 1995), suggesting that spatial separation benefits the perceptual segregation of two speeds. However, it remains unclear how different speeds of multiple objects influence individual neuronal responses possibly resulting in perceptual segmentation of the objects.

The middle temporal cortex (area MT) of primates plays an important role in speed processing (Liu & Newsome 2005, Orban et al. 1995, Pasternak & Merigan 1994). MT neurons have larger receptive fields (RF) than those in earlier visual areas, making it more likely to encompass multiple objects within the RFs under natural vision. While MT neurons' selectivity to single motion speed has been well documented, it is not fully understood how MT neurons encode two speeds that are non-overlapping within the RFs. In this study, we aim to study how MT neurons represent multiple speeds of spatially separated stimuli within the RFs.

Britten & Heuer (1999) found that the response to two Gabor patches spatially separated within the RFs of area MT, moving in the same direction and speed, could be well explained by a nonlinear summation of the responses to the individual gratings. Previous work from our laboratory has shown that when two stimuli compete in more than one feature domain, the spatial arrangement of the stimuli within the RF is important to understand the neural encoding rule (Wiesner et al. 2020). By changing the spatial arrangement from overlapping to spatially separated/non-overlapping, MT neurons receive their inputs from two distinct pools of V1 neurons, activated by the individual stimuli in different parts of MT RFs. This allows MT neurons to perform spatial and directional pooling of their inputs, as well as perform divisive normalization within MT. Although, the stimuli used in these studies did not share a border, which meant the boundaries between the two stimuli were not defined by differences in motion.

Previous studies reported that motion discontinuity border provides important information for image segmentation (Grossberg 1994, Braddick 1993, Moller and Hulbert 1997, Kohler et al. 2019), and most electrophysiology studies in area MT focused on the border encompassing the RFs in a center-surround arrangement (Allman et al 1985, Born 2000, Huang et al. 2007). An ongoing study (Wiesner & Huang SfN abstract 2019) tested how neurons in area MT encode two adjacent stimuli with a motion direction discontinuity border placed within their classical RFs. They found that responses show a strong spatial location bias towards the component that elicits a stronger response when presented alone, regardless of the orientation of the border. While the spatially separated stimuli tested were different in motion directions, their speeds were the same. Hence, when the two non-overlapping stimuli differ in their speeds, it remains to be probed whether similar rules as mentioned above, such as non-linear summation and/or strong spatial bias, would be at play in area MT.

Another study in our laboratory (Huang et al. unpublished) has investigated how area MT represents two overlapping random dot patterns moving with different speeds. When the stimulus speeds were low (faster speed ≤ 20 deg/s), the responses to two overlapping speeds were biased to the faster speed, even for the slow-speed preferred neurons that responded strongly to the slower component when presented alone. As the component speeds increased, the response to two speeds shifted toward the average of the component responses (response averaging). This suggests that MT neurons show a bias to faster component of overlapping stimuli and the extent of this bias decreases with increasing stimulus speeds. In this study, we aim to examine whether MT neurons display a similar fast bias as with the overlapping stimuli, in a stimulus speed dependent way, when the two speeds are non-overlapping in the RFs. Crucially, as the processing of the two non-overlapping speeds requires a spatial comparison of local speed estimates, we ascertain if and how the speed biases of MT neurons interact with the spatial biases, enabling conjoint encoding of spatial and speed information of the two stimuli in their RFs.

To test how MT neurons encode two stimuli that are adjacent to each other moving at different speeds, we placed two random-dot patches of equal luminance and coherence, but of different speeds side-by-side with the speed discontinuity border centered in the RFs. To capture the speed dependencies of the encoding of multiple speeds, we used five speed combinations of the non-overlapping stimuli, ranging from 1.25 deg/s and 5 deg/s to 20 deg/s and 80 deg/s, while one component speed was always four times the other. To study the interaction between the speed and spatial location, we presented two spatial arrangements: Left-component Fast & Right-component Slow and Left-component Slow & Right-component Fast. For both spatial arrangements, we found that, at low stimulus speeds, population averaged responses in area MT show a bias to the faster component. This is akin to the bias to the faster component of the

overlapping stimuli moving at different speeds. But, at high stimulus speeds, the response bias flips to the slower component. This is different from the case of overlapping stimuli where we see response averaging (i.e., no bias to either component) at high stimulus speeds. Notably, for majority of neurons, at low stimulus speeds, the faster speed alone elicited a stronger response than the slower speed alone (i.e., faster is the stronger component) and vice-versa was true at high stimulus speed (i.e., slower is the stronger component). Hence this points to a population response bias to the stronger component (which are the faster and slower speed components in low and high-speed ranges, respectively).

For individual neurons, we saw that the spatial preferences and speed preferences to single components were generally preserved in their responses to the two speeds with different spatial arrangements. For example, a neuron that preferred left location (across different speeds) and a faster speed (on left and right) presented alone, showed a higher response to "Left-component Fast & Right-component Slow" stimuli than to "Left-component Slow & Right-component Fast". We also identified a small subset of neurons that preferentially represented one component (left/right or fast/slow) over the other, across speed combinations and spatial arrangements of the two non-overlapping stimuli. Put together, our findings reveal a distributed neural code of multiple non-overlapping speed stimuli over neurons with different preferences to locations (within RFs) and speeds, in area MT, which might be beneficial for estimating the local speeds tagged with the spatial information of multiple stimuli.

Methods

Two male adult 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 like those described previously (Xiao et al. 2015). A head post and a recording cylinder were implanted during sterile surgery with the animal under isoflurane anesthesia. For electrophysiological recordings from neurons in area MT, we took a vertical approach and used tungsten electrodes (1-3 MΩ, FHC). For some experiments, we used a 16-channel laminar electrode from Plexon (S-probe) to record multiple neurons at the same time to speed up data collection. 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 & Gross, 1981). 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.

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 × 768 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.

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

Visual stimuli

The main visual stimuli were two spatially separated random-dot patches, each a square of size 5°. The border between the patches was centered on the MT neuron's receptive field. The random dots were achromatic. Each random dot was 6 pixels at a side and had a luminance of 112 cd/m2. The background luminance was 0.2 cd/m2. The dot density of each random-dot patch was 0.89 dots/deg2. In each trial, the animal fixated in a 1°x1° window of a stationary target for 250 ms with a blank screen. In the fixation task, while the animal maintained the fixation, first the main visual stimuli appeared stationary in neuron's receptive field for 300 ms and then they moved for 600 ms. The two random-dot patches translated at two different speeds in the preferred direction (PD) of the MT neuron being recorded (see Fig 1). The speed of the faster component was always four times of the slower component. In five stimulus conditions, the speeds of the slower and faster

stimulus components were 1.25 and 5°/s; 2.5 and 10°/s; 5 and 20°/s; 10 and 40°/s; 20 and 80°/s, respectively. The rationale for choosing a fixed ratio between the two speeds instead of a fixed difference was that the neural representation of speed in the middle temporal cortex (area MT) of macaque monkeys is on a logarithmic scale (Maunsell & van Essen 1983; Lisberger & Movshon 1999; Nover et al. 2005). Therefore, a fixed ratio between the two component-speeds gave rise to a fixed speed difference in the logarithmic scale.

Attention away task

To direct animal's attention away from the speed stimuli centered on the RFs, we trained Monkey R to perform a 2-AFC task of discriminating whether the motion direction of an RDP, presented in the opposite visual hemifield, was clockwise or counter-clockwise to vertically upward direction. In each trial, after the animal fixated within 1°x 1° window of the dot for 200 ms, the "attention stimulus"; a random-dot patch with a diameter of 5°, was placed 10° eccentric in the opposite hemi field from the recording stimulus and was turned on at the same time as the recording stimulus. Both remained stationary for 250ms, before moving either clockwise or counterclockwise direction from the vertical by 10°, 15°, or 20° in randomly interleaved trials. After the 500ms motion period, all the stimuli were turned off, and two reporting targets located on the left or right side of the FP were turned on. The animal made a saccadic eye movement to the left (or right) if the motion was counterclockwise (or clockwise) of vertical and maintained fixation on the chosen target for 200ms to receive a small juice reward.

Neural data analysis

Response firing rate was calculated during the stimulus motion period (600 ms for the attention task and 500 ms for attention away task) averaged across repeated trials. For 4X speed

PD stimuli (see above), we constructed the response tuning curves to bi-speed stimuli (two components in RF) and to the constituent single speed stimuli (one component in RF). We then normalized neuronal responses to two speeds by each neuron's maximum of response to individual components and averaged the normalized tuning curves across neurons. To ensure good similarity between speed tuning curves of left and right patches, we selected neurons with correlation coefficient (using "corr2" Matlab function) of left and right (or center and background) tuning curves greater than or equal to 0.6 for the subsequent analyses.

Correlation between the bi-speed and component tuning curves for various predictions

We used the methods of Smith et al. (2005), and Kumano & Uka (2013) to classify the neuronal bi-speed tuning curves based on their similarity to predicted tuning under different scenarios. The similarity of a bi-speed tuning to prediction 1 and prediction 2 (for example fast component tuning and slow component tuning respectively) was determined the partial correlations Rp1 and Rp2 as follows:

$Rp1 = \frac{r_1 - r_2 * r_{12}}{\sqrt{(1-{r_2}^2)*(1-{r_{12}}^2)}}$; $Rp2 = \frac{r_2 - r_1 * r_{12}}{\sqrt{(1-{r_1}^2)*(1-{r_{12}}^2)}}$, where r_1 and r_2 are the correlations between the neuronal responses to prediction 1 and 2, respectively, and r_{12} is the correlation between the two predictions. We took each value of $Rp1$ or $Rp2$ and converted it to a *Z*-score using the following equation (shown for $Rp1$):

$$Zp1 = \frac{0.5 * ln\left(\frac{1+Rp1}{1-Rp1}\right)}{\sqrt{1/df}}$$

where *df* is the degrees of freedom, equal to the number of values in the tuning curve minus 3 (there were 5 values in our speed tuning curves). The numerator of the equation is the Fisher *r*-to-

Z transformation. We used a criterion of 1.28, equivalent to $P = 0.90$, for testing each value of Z_{p1} or Z_{p2} for significance. For a cell to be grouped under prediction 1 (for example "Fast bias") class, the value of Z_{p1} had to exceed the value of Z_{p2} (or zero, if Z_{p2} is negative) by this amount. Similarly, Z_{p2} had to exceed Z_{p1} by that same amount (1.28) for a cell to belong to prediction 2 (for example "Slow bias") class. If a cell met neither of these conditions, it remained unclassified.

Time course analysis of response tuning

To examine the time-course of the response tuning to the bidirectional stimuli, we calculated the response tuning for each individual neuron using the trial-averaged firing rates within a 50ms time window, sliding at a step of 10-ms, and normalized the responses by the maximum firing rate of single components response across all time windows.

Results

We investigated how MT neurons encode multiple speeds that are spatially separated and whether the encoding rules facilitate the segmentation of multiple objects in natural environments. To address this question, we recorded the electrical activity of neurons in area MT from two male rhesus macaques while they performed either a fixation task (monkey R and monkey Z) or a fine direction discrimination task directing attention away from the RFs (monkey R). Visual stimuli were two spatially separated (i.e., non-overlapping) random-dot patches moving simultaneously at different speeds in (or close to) neurons' preferred direction. We placed the two patches side-by-side with a vertical border centered on the RFs. To capture the speed dependencies of the encoding of multiple speeds, we used five speed combinations of the non-overlapping stimuli, ranging from 1.25 deg/s and 5 deg/s to 20 deg/s and 80 deg/s, while one component speed was always four times the other. To evaluate the effect of spatial arrangement on the encoding of

multiple speeds, we used two arrangements: the faster component was either on the left or the right side of the slower component (Fig 1).

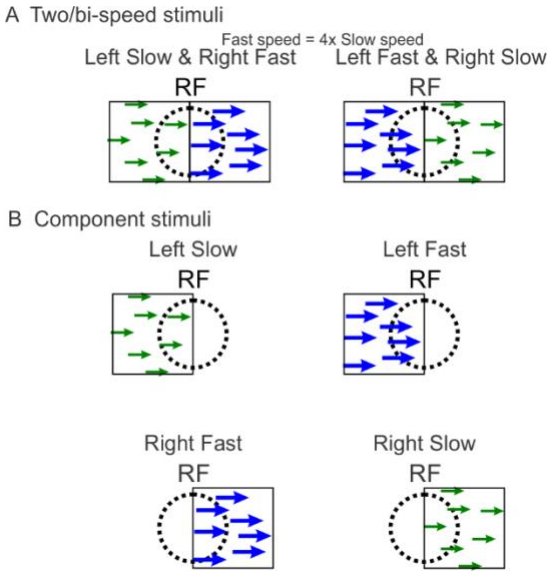

Figure 1. Visual stimuli.

(A) Two spatial configurations of the non-overlapping speeds (bi-speed stimuli): Left Slow & Right Fast and Left Fast & Right Slow (bi-speed). (B) Component stimuli presented alone: Left Slow, Right Fast, Left Fast, and Right Slow. The stimuli are 5°x 5° square random dot patches. All the dots are achromatic. The colors and the black square borders are only for illustration purposes. One component (faster) is always four times faster than the other (slower) for five different combinations of speeds tested. The speed discontinuity border between the two speeds is placed approximately at the center of the receptive fields (dashed lines).

MT neurons response tuning to spatially separated stimuli moving at different speeds: "stronger-component-take-all"

Figure 2 shows the responses to spatially separated stimuli moving at different speeds (shown in the cartoon) from four example neurons sorted by their preferred speeds. Fig 2A1 to D1 show the responses of the neurons to "Left Slow & Right Fast" bi-speed stimuli arrangement and, Fig 2A2 to D2 show the responses of the neurons to "Left Fast & Right Slow" bi-speed stimuli arrangement. Plotted on the X-axis are mean values of the speeds of the two stimulus components (with 4X speed separation, see Methods). The exact speeds of the stimulus components are shown

in Figure 2A. The red curve indicates the response elicited by the bi-speed stimuli with left and right components presented simultaneously in neurons' RFs. The green and blue curves indicate the responses elicited by the constituent slow (presented alone on left for Fig 2A1 to D1 and on right Fig 2A2 to D2) and fast components (presented alone on right for Fig 2A1 to D1 and on left Fig 2A2 to D2), respectively. Fig 2A3 to D3 compare the speed tuning curves of the neurons to left and right components. Note that for most of the neurons in our dataset, the left and right speed tuning curves are well correlated (see Methods).

We examined the relationship between MT responses to two non-overlapping speeds and those to constituent speed stimuli presented alone at the corresponding locations (left or right). For the example neurons shown, for most stimulus speed combinations, the bi-speed response generally followed the higher component response (aka stronger component bias). This stronger component bias was seen whether the strong component was the slower component (Fig 2A1, 2A2) or the faster component (Fig 2D1, 2D2). This phenomenon results in an interesting response profile for neurons with intermediate preferred speeds for which the component tuning curves cross each other (neuron #2 and #3). The bi-speed response for these neurons roughly follows the stronger component switching from the fast speed for the first four mean speed values to the slow speed for the last one. For neurons like #2, we also noticed response facilitation to bi-speed stimuli (above the responses to the single components) when the responses to the component stimuli match.

Fig 3A shows the response tuning curves averaged across all neurons in the population. Across combinations of component stimulus speeds, the population-averaged responses showed a bias toward the stronger component (i.e., the component that elicits a higher response) for both stimulus arrangements (Left Slow: Fig 3A1 to D1 and Right Slow: Fig 3A1 to D2). To examine

whether this overall trend holds for neurons with different preferred speeds, we divided our neuronal population into three groups with “low” (<2.5°/s), “intermediate” (between 2.5 and 25°/s), and “high” (>25°/s) preferred speeds (PS) and examined the speed tuning for each of these groups.

For 3 neurons that preferred low speeds (Fig 3B1, B2), the averaged bi-speed response (red curves) followed the stronger component which was the slow component (green curves). For 42 neurons with high preferred speeds (Fig 3D1, D2), the averaged bi-speed response (red curves) followed the stronger component which was the fast component (blue curves) for most speed combinations and, even overshot the component responses when they match (matching points on blue and green curves shown by arrows). For 36 neurons with intermediate PSs (Fig 3C1, C2), the bi-speed response follows the stronger component, switching from the fast speed for the lower mean speed values to the slow speed for the larger mean speed values. The bi-speed response for these neurons also overshot the component responses when they matched (matching points on blue and green curves shown by arrows), like the high PS neurons (Fig 3D1, D2). This overshooting of bi-speed response relative to matching component responses might suggest response facilitation, when the stimulus components provide a similar drive to an MT neuron. Overall, these results indicate that MT neurons with a range of preferred speeds show a “winner-takes-all” behavior when two speed features are presented next to each other in their receptive fields, regardless of whether the winner (which is the stronger component) is faster or slower on left or right sides of the RFs.

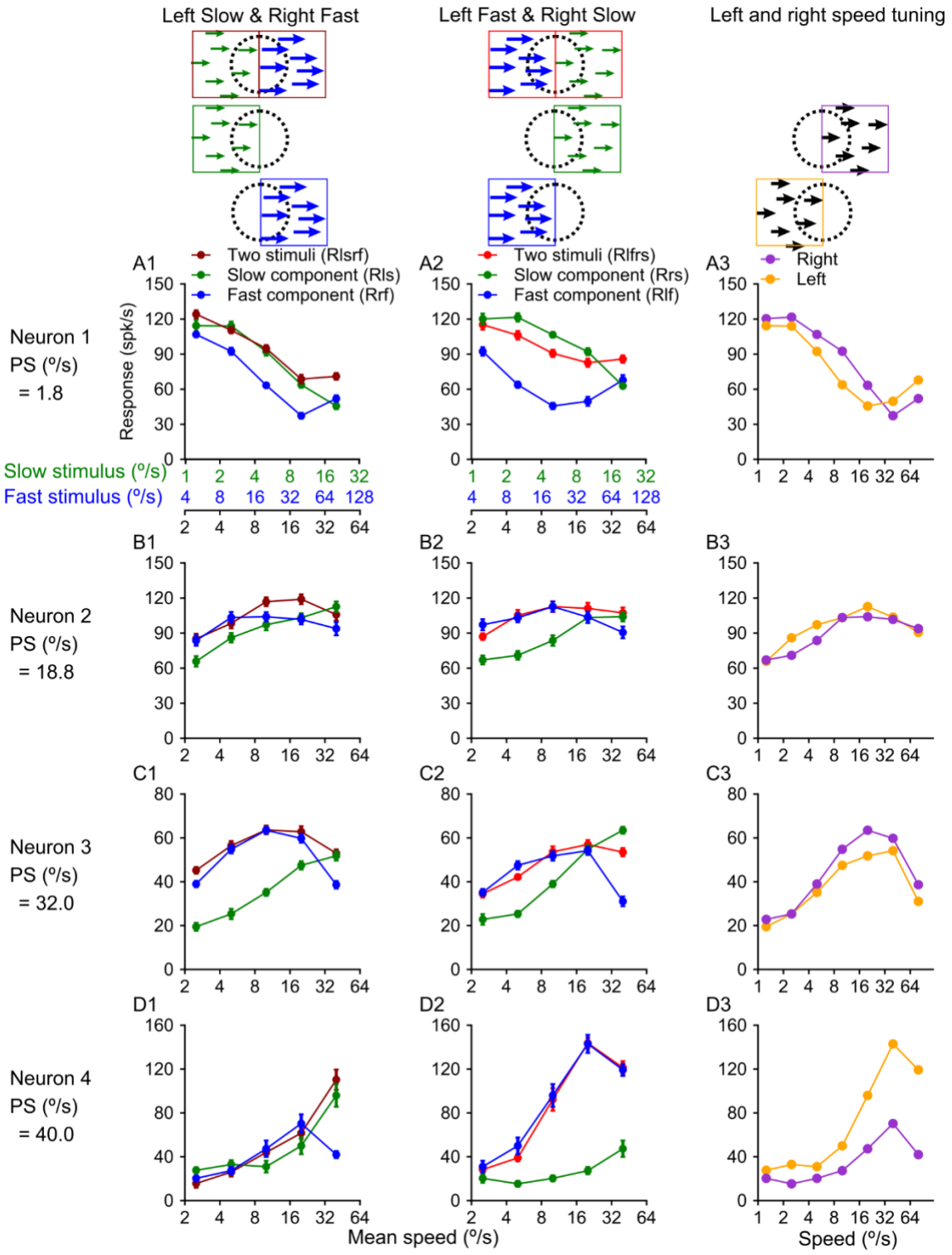

Figure 2. Bi-speed and component speed tuning curves of four example neurons.

(A1-D1) Brown curves show tuning to bi-speed stimuli: Left Slow & Right Fast (Rlsrf). Green and blue curves show tuning to component speeds Left Slow (Rls) and Right Fast (Rrf), respectively. (A2-D2) Red curves show tuning to bi-speed stimuli: Left Fast & Right Slow (Rlfrs). Green and blue curves show tuning

to component speeds Left Fast (Rlf) and Right Slow (Rrs), respectively. x-axis is the mean speed of the two stimuli in log scale (mean speed= exp((ln(slow)+ln(fast))/2). Y-axis is the firing rate. The green and blue axes in A1 show the speeds of the slower and faster stimuli. There were five combinations of speeds tested (5 points in each curve). (A3-D3) Purple and yellow curves show the speed tuning curves of right and left components presented alone. Note that the green and blue curves in A1-D1 and A2-D2 are portions of the right and left speed tuning curves.

Note that the responses shown in Figs 2 and 3 were collected when Monkey R performed the "Attention Away" task with his attention directed away from MT neurons' receptive fields (see Methods). This eliminated any selective deployment of attention to either stimulus component in a simple fixation task (e.g., attention to the more "salient" stimulus component like the one closer to the fixation point or the one with a faster speed). Fig 4 shows the population averaged responses to the main stimuli presented in RFs of 42 MT neurons when the animal simply fixated at the center of the monitor, with no control for attention (note: 25 of the 42 MT neurons were tested with both Attention Away and Fixation, and 17 were tested with only Fixation task). We saw a similar "stronger-component-take-all" trend in bi-speed responses across three groups of neurons: "low", "intermediate" and "high" preferred speeds. We also collected MT responses (N=11) to the non-overlapping stimuli from a second animal, Monkey Z, using the fixation task. As the population trends were similar across the three datasets (Monkey R: Attention Away N=81 and Fixation N=17; Monkey Z: Fixation N=11), we combined them (N=109) for subsequent analyses.

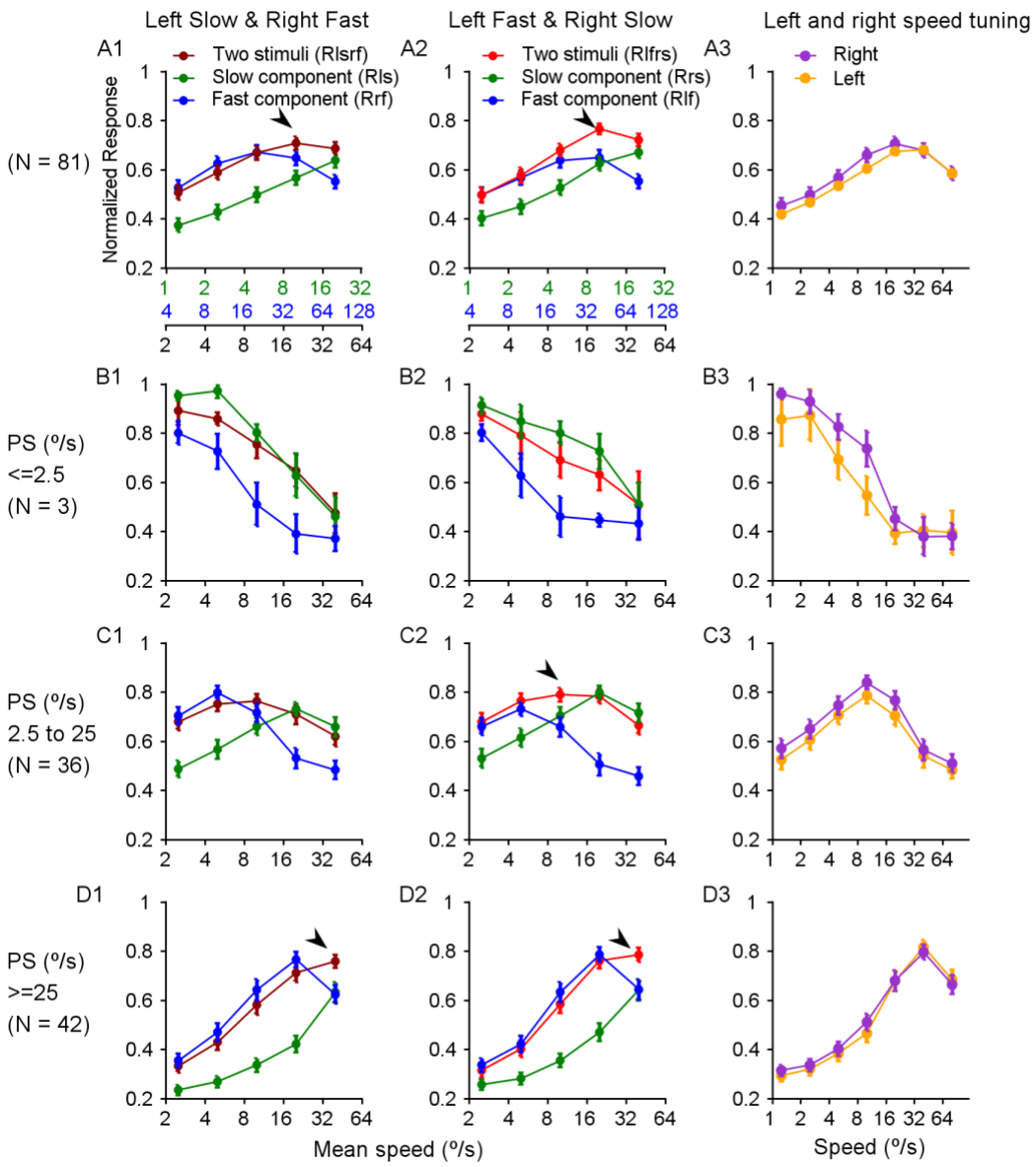

Figure 3. Population averaged bi-speed and component speed tuning curves (Attention Away).

(A1-D1), (A2-D2), (A3-D3) Same as Figure 2 showing tuning curves averaged across the population of neurons instead. (A1-A3) All neurons. (B1-B3) Neurons with preferred speeds (PS)<=2.5°/s. (C1-C3) Neurons with 2.5°/s <PS<25°/s. (D1-D3) Neurons with PS>=25°/s. The arrows show examples of bi-speed response facilitation when the two component responses are similar.

Population response weights of faster and slower stimulus components

To quantify the relationship between the bi-speed response and the corresponding component responses, for each neuron and a given pair of stimulus speeds, we expressed the bi-speed response R as a linear weighted sum of the component responses R_s and R_f elicited by the slower and faster stimulus component, respectively:

$$R = w_s R_s + w_f R_f \quad , \qquad (1)$$

in which, w_s and w_f are the response weights for the slower and faster stimulus component, respectively. Note that, here we did not make any assumption regarding whether w_s and w_f should be constant across neurons, or across different speed combinations. For any given set of three data points of R, R_s and R_f, if $R_f - R_s \neq 0$, R can always be expressed as:

$$R = \frac{R_f - R}{R_f - R_s} R_s + \frac{R - R_s}{R_f - R_s} R_f \quad , \qquad (2)$$

Therefore, the response weights can be determined by how close the bi-speed response was to one of the component responses, normalized by the distance between the two component responses: $w_s = \frac{R_f - R}{R_f - R_s}$, $w_f = \frac{R - R_s}{R_f - R_s}$ (note that $w_f + w_s = 1$).

For each neuron, one stimulus component would have a higher response weight than the other, if the bi-speed response was closer to the response elicited by the component compared to that elicited by the other component. To estimate the response weights of the faster component across a population of neurons (N=109), in Fig 4, we plotted $R - R_s$ on Y-axis and $R_f - R_s$ on X-axis, for five combinations of the stimulus speeds in separate panels (Fig 5A1 to A5). The brown and red symbols represent the data from "Left Slow & Right Fast" and "Left Fast & Right Slow" stimulus arrangements, respectively. Across the neuron population, the relationship between R -

R_s and R_f - R_s is remarkably linear (see Table 1 for type II linear regression parameters), and can be expressed as:

$$R - R_s = k_f(R_f - R_s) + b_f \qquad (3)$$

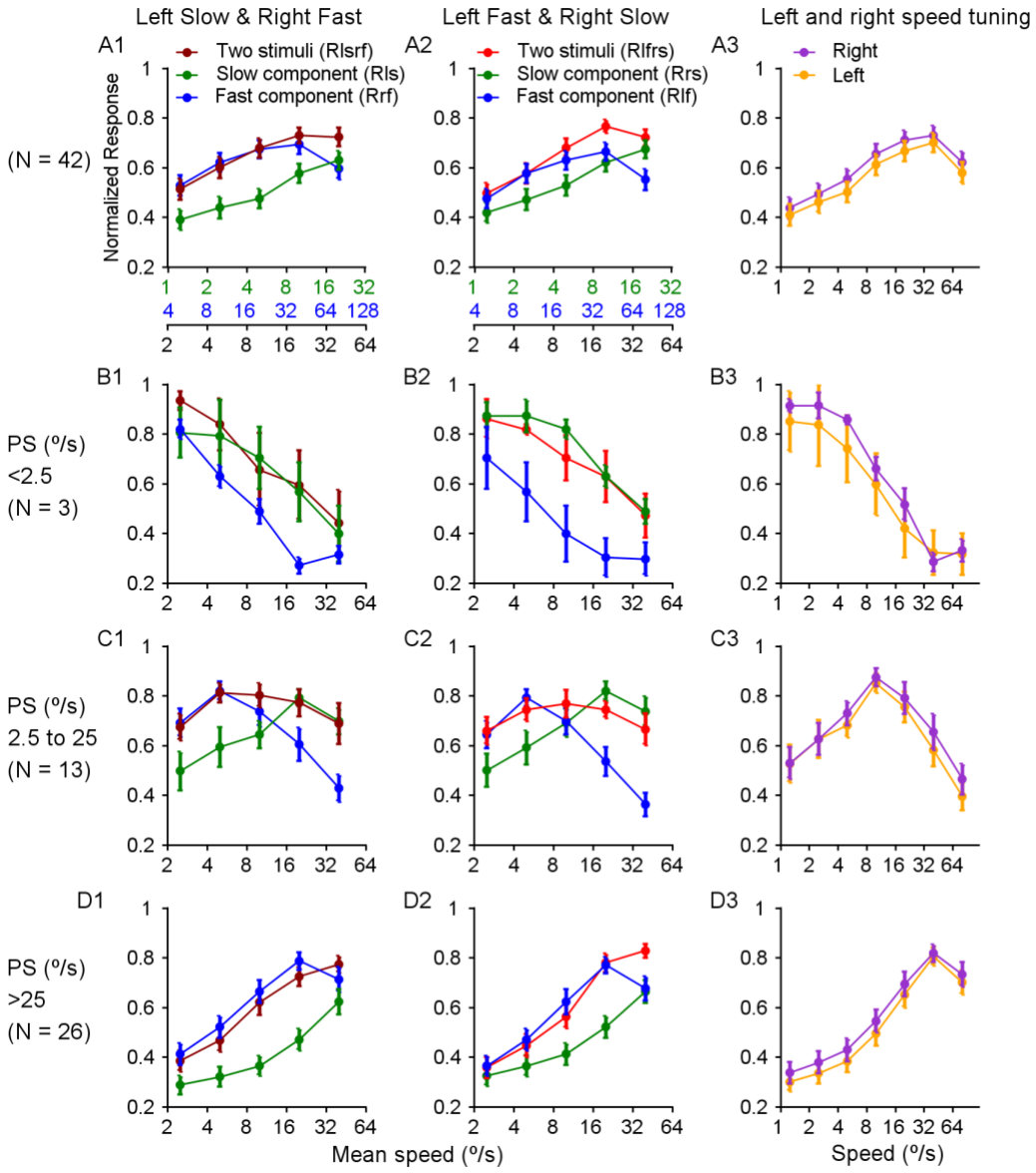

Figure 4. Population averaged bi-speed and component speed tuning curves (Fixation task).

(A1-D1), (A2-D2), (A3-D3) Same as Figure 2 showing tuning curves averaged across the population of neurons (Monkey R N=31; Monkey Z N=11). (A1-A3) All neurons. (B1-B3) Neurons with preferred speeds (PS)<=2.5°/s. (C1-C3) Neurons with 2.5°/s <PS<25°/s. (D1-D3) Neurons with PS>=25°/s.

Since the intercept b is small (Table 1), the slope k_f of the linear regression approximates $\frac{R-R_s}{R_f-R_s}$, which is the response weight for the faster component ($w_f\ in\ Eq.2$). Therefore, we can estimate the population response weight of the faster component (w_f) using the slope of the linear regression. The weight for the slower component can be estimated as k_s given by the following equation of a similar form as (3).

$$R-R_f = k_s(R_s - R_f) + b_s \qquad (4)$$

Note that $k_f + k_s \approx 1$ if the intercepts b_f and b_s are close to zero (since $w_f + w_s = 1$ according to Equation 2).

Fig 5B summarizes the response weights of the faster and slower components (solid and dashed lines respectively), estimated by the slope values for different speed combinations. The faster component response weight was high (close to 1) for the low stimulus speeds (1.25 & 5 °/s and 2.5 & 10 °/s) and decreased with increasing stimulus speeds. On the other hand, the slower component weight increased with increasing stimulus speeds. For the high stimulus speeds, slower component weight matched the faster component weight (10 & 40 °/s) and even raised above the faster component weight (20 & 80 °/s) (note: type II linear regression lines estimating slower component weight are not shown in the figure).

This suggests that MT population responses to non-overlapping fast and slow speeds are biased to faster components at low stimulus speeds and to slower components at very high stimulus speeds. This trend holds true for two non-overlapping spatial arrangements tested (red and brown lines). Notably, for majority of neurons, at low stimulus speeds, the faster speed alone elicited a stronger response than the slower speed alone (i.e., faster is the stronger component) and vice-versa was

true at high stimulus speed (i.e., slower is the stronger component). Hence this points to a bias toward the stronger population response elicited by the faster and slower speed components in low and high-speed ranges, respectively. When fast and slow speeds were overlapping in MT neuron RFs, Huang et al (unpublished) observed a similar faster component bias at low stimulus speeds (black lines in Fig 5B). At very high stimulus speeds (20 & 80 ° /s), the authors reported equal weights for the faster and slower components in contrast to a stronger bias to the slower component we observed when the stimuli are spatially separated.

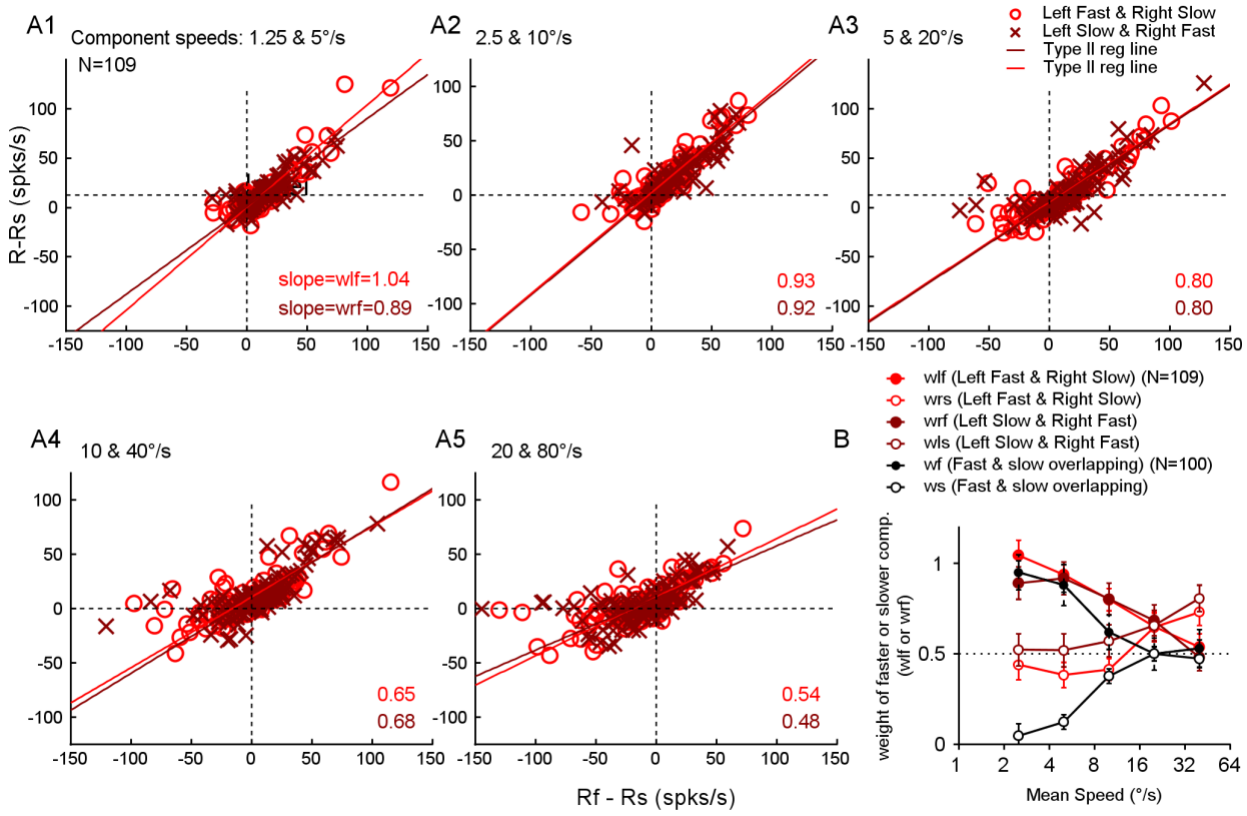

Figure 5. Weights of faster components for different stimulus speed combinations.

(A1) 1.25°/s & 5°/s. X values are the differences between the responses to faster and slower components ($R_f - R_s$) for each neuron (N=109). Y values are the differences between the responses to bi-speed stimuli and slower component ($R - R_s$). Red circles and brown crosses show the values from Left Fast & Right Slow and Left Slow & Right Fast (bispeed) conditions. The solid lines show type II linear regression fits. The slopes of the fits (i.e., the weight of the faster component) are labeled on the lower right. (A2-A4)

2.5°/s & 10°/s, 5°/s & 20°/s, and 20°/s & 80°/s (B) Weight of faster (solid circles) or slower (open circles) component vs mean speeds of the components in A1-A4. Red, brown, and black are from Left Fast & Right Slow, Left Slow & Right Fast, and Fast & Slow overlapping (Huang et al., unpublished), respectively.

Table 1. Weights of faster and slower components, Type II Linear Regression (N = 109)

Components Speeds (°/s)	1.25 & 5		2.5 &10		5 & 20		10&40		20&80	
Spatial arrangement	LS & RF	LF & RS	LS & RF	LF & RS	LS & RF	LF & RS	LS & RF	LF & RS	LS & RF	LF & RS
Slope (k_f) *and 95% CI*	**0.89** ± 0.09	**1.04** ± 0.09	**0.92** ± 0.07	**0.94** ± 0.09	**0.80** ± 0.09	**0.80** ± 0.06	**0.68** ± 0.09	**0.65** ± 0.08	**0.48** ± 0.07	**0.54** ± 0.07
Intercept (b_f) *and 95% CI*	**1.20** ± 1.15	**0.15** ± 0.85	**-0.71** ± 1.56	**2.09** ± 0.85	**3.62** ± 1.48	**4.91** ± 0.80	**8.58** ± 0.57	**10.7** ± 0.12	**9.87** ± 0.77	**10.7** ± 0.99
Slope (k_s) *and 95% CI*	**0.52** ± 0.09	**0.44** ± 0.09	**0.52** ± 0.09	**0.38** ± 0.07	**0.57** ± 0.09	**0.41** ± 0.06	**0.66** ± 0.08	**0.65** ± 0.08	**0.81** ± 0.07	**0.73** ± 0.07
Intercept (b_s) *and 95% CI*	**6.56** ± 1.15	**5.09** ± 0.85	**7.35** ± 1.56	**6.06** ± 0.85	**9.95** ± 1.48	**7.73** ± 0.80	**10.8** ± 0.57	**11.2** ± 0.12	**6.87** ± 0.77	**7.04** ± 0.99

Time course of the responses

To shed light on the neural mechanisms underlying the response biases, we examined the time course of the neuronal responses in the spatially separated (our dataset) and overlapping (Huang et al unpublished) conditions. Figure 6 shows the time course of the response averaged across neurons.

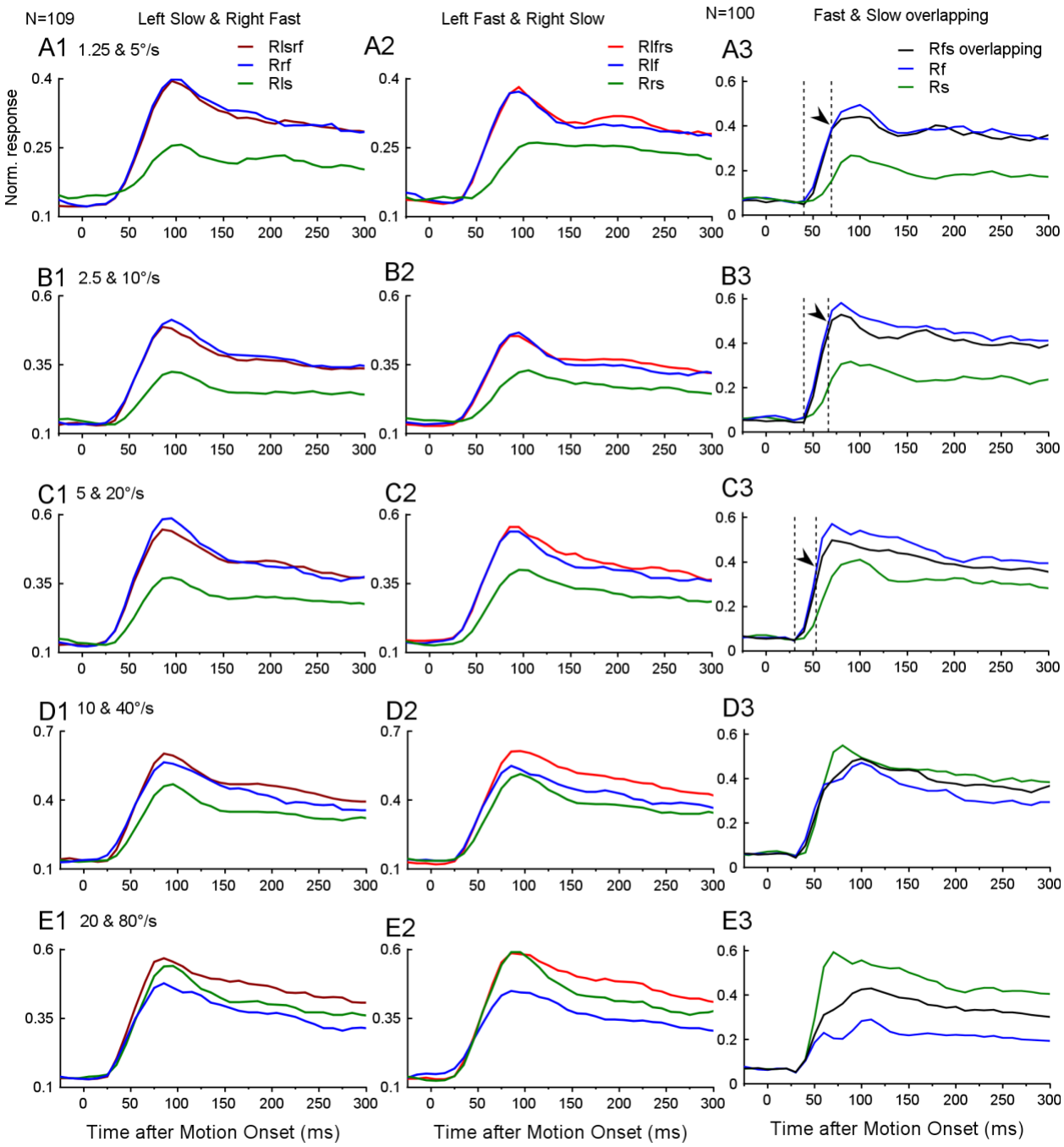

Figure 6. Time course of the population averaged responses to bi-speed and single speed stimuli.

X and Y axes are the time after motion onset and the normalized responses. (A1-E1) Brown, blue, and green curves are the responses to Left Slow & Right Fast stimuli, Right Fast component, and Left Slow component stimuli. (A2-E2) Red, blue, and green curves are the responses to Left Fast & Right Slow stimuli, Left Fast component, and Right Slow component stimuli. (A3-E3) Black, blue, and green curves are the responses to Fast & Slow overlapping stimuli, Fast component, and Slow component stimuli (Huang et al., unpublished). The vertical dashed lines mark the start and the end of the time during which the bi-

speed responses closely follow the response to faster components. (A1-A3) 1.25°/s & 5°/s. (B1-B3) 2.5°/s & 10°/s. (C1-C3) 5°/s & 20°/s. (D1-D3) 20°/s & 80°/s.

When two speeds were overlapping, the bi-speed responses (red curves in Fig. 6A3, B3, C3) followed the responses elicited by the faster speed right after the motion onset for ~25 ms and was then "pulled down" by the slower component (indicated by the arrows). When stimuli were spatially separated, the response elicited by both the faster and slower components (brown and red curves in Fig. 5 column 1 and 2, respectively) closely followed the higher response elicited by a single component alone (faster component at low stimulus speeds: Fig 6A1 to D1, A2 to D2, and slower component at higher stimulus speed: Fig 6E1 and E2). This means that the weaker stimulus (be it slower or faster) had little effect on the population-averaged responses to the two non-overlapping speeds throughout the motion period. Interestingly, when the responses to the two components, presented one at a time, are similar, the bi-speed responses overshoot the component responses prominently after the responses peak (see Fig 6D2, E1, and E2). This implies a response facilitation (more than what's predicted by a winner-takes-all scheme: sub-linear summation) when the component stimuli drive the neurons to a similar extent. These results point to a weaker or almost no response normalization occurring when two speeds are spatially separated compared to when the two speeds are overlapping in MT neurons' RFs.

Effects of spatial arrangement of fast and slow speeds on MT neuronal responses

For a good number of MT neurons, we noticed that one side (left or right) of the receptive field elicited a higher speed tuning response than the other. For example, for neurons in Fig 1A3 and 1C3, the right component speed tuning (purple) has a higher response than that of the left component (orange). To quantify the extent of differential activation to left and right components for each neuron, we defined a location preference index (*lpi*) as below:

$$lpi = \frac{\sum R_{left}(s) - \sum R_{right}(s)}{\sum R_{left}(s) + \sum R_{right}(s)} \quad (5)$$

where $R_{left}(s)$ and $R_{right}(s)$ are the responses to the left and right components presented alone, moving at speed s (values of s range from 1.25 to 80 °/s), respectively. Positive *lpi* values indicate a higher response to the left component compared to the right component, across speeds, and negative values indicate a higher response to the right component compared to the left one. Hence, the distribution of *lpi* captures the spectrum of spatial preferences of neurons from strong right component preference to strong left component preference.

We also found that for some neurons, "Left Slow & Right Fast" stimuli elicited a higher response than "Left Fast & Right Slow" stimuli (e.g., brown curve > red curve in Fig 1C1 and C2) and vice-versa for other neurons (e.g., brown curve < red curve in Fig 1D1 and D2), at different stimulus speeds. Across neurons with different preferred speeds (PS), we asked how the differences in responses to the two spatial arrangements of the faster and slower speed components relate to the spatial preferences to single components (measured by *lpi*). Fig 7 shows the normalized responses to "Left Fast & Right Slow" (size of the red circles: Rlfrs) and "Left Slow & Right Fast" (size of the brown circles: Rlsrf) for different *lpi* values plotted on Y-axis and log (pref. speed) values plotted on X-axis. Panels 7A1 to A5 show this spread for different stimulus speed combinations (green and blue dotted lines indicate the speeds of the slower and faster components).

We identified four different cases that showed differences in responses to the two spatial arrangements of the faster and slower speed components:

(1: Left & Faster) neurons with preferred speeds (PS) closer to the faster component speed and had a high spatial preference to the left component (*lpi*>0) showed a higher response to "Left

Fast & Right Slow" than "Left Slow & Right Fast" (i.e., Rlfrs > Rlsrf) (examples marked by grey dashed lines).

(2: Left & Slower) neurons with PS closer to the slower component speed and had a high spatial preference to the left component (*lpi*>0) showed the opposite trend i.e., Rlsrf > Rlfrs (examples marked by orange dashed lines).

(3: Right & Slower) neurons with PS closer to the slower component speed and had a high spatial preference to the right component (*lpi*<0) showed a trend like (1) i.e., Rlfrs > Rlsrf (examples marked by dark green dashed lines).

(4: Right & Faster) neurons with PS closer to the faster component speed and had a high spatial preference to the right component (*lpi*<0) showed a trend like (2) i.e., Rlsrf > Rlfrs (examples marked by purple dashed lines).

To measure the extent of differential activation to the faster and slower components for each neuron at each speed combination, we defined a speed preference index (*spi*) as below:

$$spi(s_{slow}, s_{fast}) = \frac{R_{left}(s_{fast}) - R_{left}(s_{slow}) + R_{right}(s_{fast}) - R_{right}(s_{slow})}{R_{left}(s_{fast}) + R_{left}(s_{slow}) + R_{right}(s_{fast}) + R_{right}(s_{slow})} \qquad (6)$$

where $R_{left}(s)$ and $R_{right}(s)$ are the responses to the left and right components presented alone, moving at a speed s, respectively. s_{slow} and s_{fast} are the speeds of slower and faster components of each of the five speed combinations of the non-overlapping stimuli (Fig 7A1 to A5). For a given speed combination, positive *spi* values indicate a higher response to the faster component compared to the slower component, across left and right locations, and negative values indicate vice-versa.

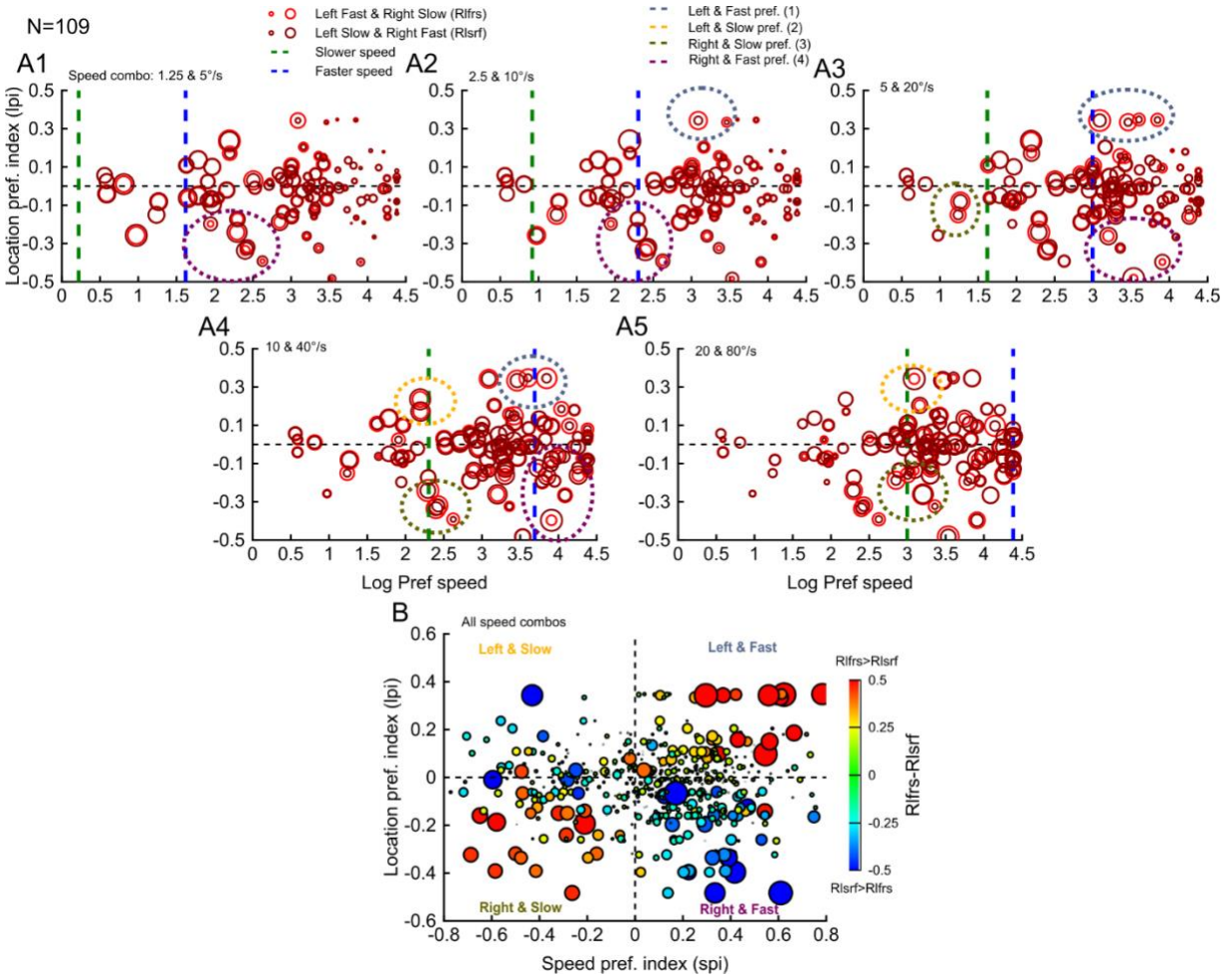

Figure 7. Population response codes of the two spatial arrangements.

(A1) 1.25°/s & 5°/s. X values are the log (preferred speeds) of all neurons (N=109). Y values are the location preference indices (lpi). The sizes of the red and brown bubbles are proportional to the normalized responses to Left Fast & Right Slow (Rlfrs) and Left Slow & Right Fast (Rlsrf) bi-speed stimuli, respectively. Green and blue vertical lines mark the speeds of slower and faster components (indicated on the top). (A2-A4) 2.5°/s & 10°/s, 5°/s & 20°/s, and 20°/s & 80°/s. Baby blue, yellow, dark green, and purple dashed lines mark example neurons with Left & Fast, Left & Slow, Right & Slow, and Right & Fast preferences. (B) Location preference index (lpi) vs Speed preference index (spi) vs Rlfrs-Rlsrf. The size and color of the bubbles indicate the values of Rlfrs-Rlsrf (hotter and cooler colors indicate high positive and low negative values, respectively). Each quadrant is occupied by neurons with a combination of speed and location preferences indicated by colored text.

Fig 7B summarizes the relationship between spatial/speed preferences and their differential response to two spatial arrangements across stimulus speed combinations seen in Fig 7A1 to A5. We plotted speed preference indices (spi), location preference indices (lpi), and Rlfrs-Rlsrf values

on the X, Y and Z-axes (magnitude indicated by the size and the color of each bubble), respectively. Some neurons preferred one component speed over the other (faster or slower) and one location over the other (left or right), populating the four quadrants of Fig 7B (Left & Fast, Left & Slow, Right & Slow, and Right & Fast preferences). The response to the spatial arrangement of two speeds with the preferred speed component placed at the preferred location is usually higher than the one with the preferred speed placed at a less preferred location (and less preferred speed placed at the preferred location). In other words, neurons that preferred Left & Fast (Fig 7B quad 1) or Right & Slow (Fig 7B quad 3) usually showed a higher response to "Left Fast & Right Slow" than "Left Slow & Right Fast" (Rlfrs> Rlsrf). Neurons that preferred Left & Slow (Fig 7B quad 2) or Right & Fast (Fig 7B quad 4) usually showed a higher response to "Left Slow & Right Fast" than "Left Fast & Right Slow" (Rlfrs< Rlsrf). This suggests that the spatial preferences and speed preferences of neurons, measured using the responses to single components, were generally preserved in their responses to the two stimuli with different spatial arrangements.

We asked how spatial and speed preferences of MT neurons, together, shaped the response weights of the components of two non-overlapping stimuli. For each neuron, based on its location and speed preferences, we assigned the two stimulus components as 1) the "preferred speed at preferred location (SL)" and the "non-preferred speed at non-preferred location (nSnL)" or 2) the "preferred speed at non-preferred location (SnL) and "non-preferred speed at preferred location (nSL)". For neurons that preferred Left & Fast or Right & Slow (Fig 7 quad 1 and 3), "Left Fast & Right Slow" is assigned as "SL & nSnL" and "Left Slow & Right Fast" as "SnL & nSL". The assignment is vice-versa for neurons with Left & Slow or Right & Fast preferences (Fig 7 quad 2 and 4).

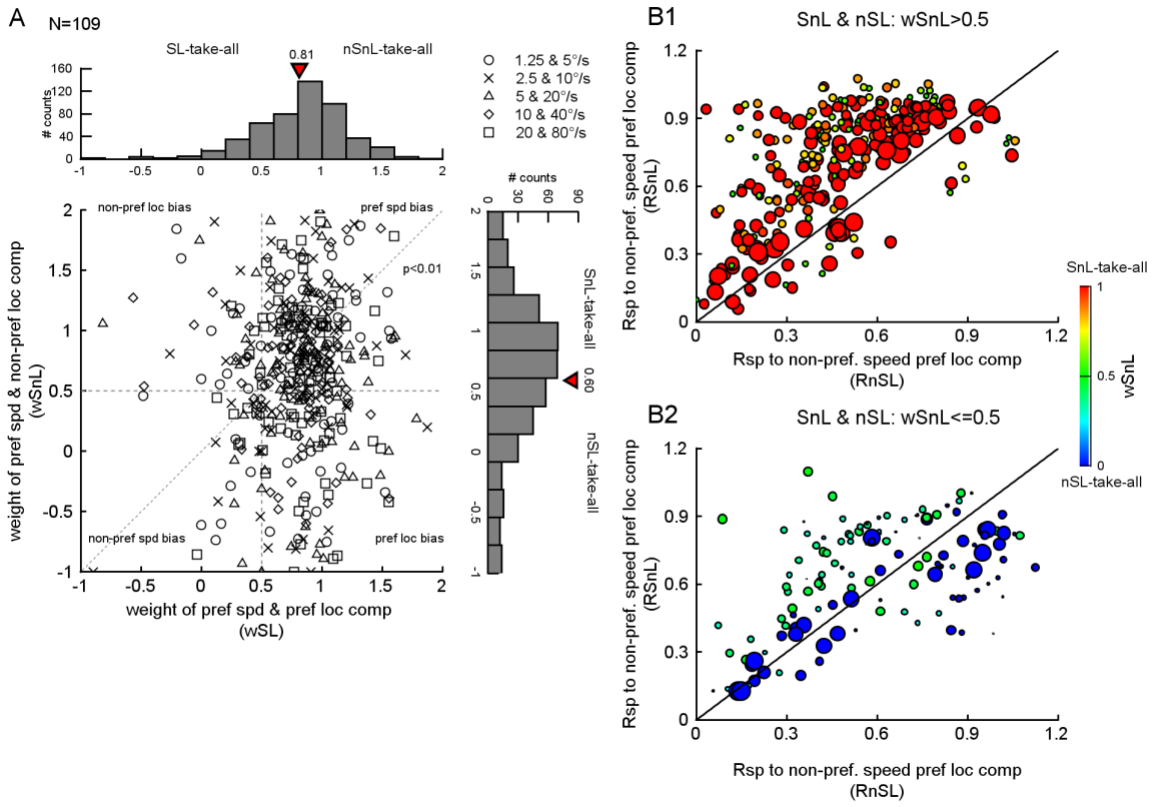

Figure 8. Component weights when speed and locations preferences of MT neurons are pitted against each other.

(A) X values are the weights of component speed closer to neurons' preferred speed placed at preferred location (wSL). Note the other component of this condition is non-preferred speed placed at non-preferred location (wnSnL=1-wSL). Y values are the weights of component speed closer to neurons' preferred speed placed at non-preferred location (wSnL). The other component of this condition is non-preferred speed placed at preferred location (wnSL=1-wSnL). Note that in this condition the speed and location preferences of MT neurons are pitted against each other. The red downward triangles indicate the mean values of the marginal distributions. Weights of a component close to 1 indicate a "winner-takes-all" bias to that component; values close to 0 indicate a bias to the other component. p-value is for one-sided signed rank tests comparing x and y sample sets. (B1, B2) Normalized responses to preferred speed at non-preferred location (y) vs those to non-preferred speed at preferred location (x) vs weights of component speed closer to neurons' preferred speed placed at non-preferred location i.e., wSnL (z). Hotter colors/larger bubbles indicate wSnL close to 1 and cooler colors/smaller bubbles indicate wSnL close to 0.

Like in Eq 1, the response weights of component 1 of the two stimuli (or bi-speed) can be estimated by the difference between bi-speed response and the response to component 2, divided by the

difference between the responses to component 1 and component 2. Weight values close to 1 suggest a strong bias to component 1 (component 1-take-all). Values equal to 0.5 indicate that bi-speed response is the average of the responses to the components, presented one at a time (response averaging). Values close to zero indicate a strong bias to component 2 (component 2-take-all). Using this definition, the weight of preferred Speed at preferred Location (SL) component is given by

$$wSL = (R(SL \,\&\, nSnL) - R(nSnL)) / (R(SL) - R(nSnL)), \qquad (7)$$

where R (SL & nSnL) is the response to "SnL & nSL" stimuli, R (SL) and R (nSnL) are their responses to the component stimuli alone i.e., preferred speed at preferred location and non-preferred speed at non-preferred location stimuli, respectively. Note that weight of non-preferred Speed at non-preferred Location (SL) component is simply wnSnL=1- wSL.

Similarly, the weight of preferred Speed at non-preferred Location (SnL) is given by

$$wSnL = (R(SnL \,\&\, nSL) - R(nSL)) / (R(SnL) - R(nSL)), \qquad (8)$$

where R (SnL & nSL) is the response to "SnL & nSL" stimuli, R (SnL) and R (nSL) are their responses to the component stimuli alone i.e., preferred speed at non-preferred location and non-preferred speed at preferred location stimuli, respectively. Weight of non-preferred Speed at preferred Location (SL) component is simply wnSL=1- wSnL.

Fig 8A shows the scatter plot of wSL vs wSnL for all neurons for all five speed combinations. For most neurons, wSL values (X-axis) are higher than 0.5 suggesting a bias to the component moving at the preferred Speed at preferred Location (mean wSL=0.81, one-sided signed rank test $p<0.001$). This suggests that when two speeds are juxtaposed in MT RFs, a component with both spatial and speed preferences aligned has a higher weight in contributing to

MT responses than the other. This could be because, when presented alone, the component with both spatial and speed preferences aligned (SL) also most likely elicits a higher response than other i.e., moving at non-preferred speed at non-preferred location (by definition). This confirms the strong component bias described earlier using population responses.

The case of preferred Speed at non-preferred Location & non-preferred Speed at preferred Location (SnL & nSL) is interesting because the preferred speed and preferred location are pitted against each other. We found that for more neurons, wSnL values (X-axis) are higher than 0.5 suggesting a bias to the component moving at preferred Speed at non-preferred Location (mean wSnL=0.60, one-sided signed rank test $p<0.001$) (Fig. 8A, marginal distribution on the y-axis), likely owing to slightly higher response elicited by SnL component than nSL component, presented alone (Fig 8B1, for wSnL greater than 0.5, most dots above unity line). These neurons represent their preferred speed component more than their preferred location (indicated by "pref speed bias" in Fig 8A quad 1). But there are few neurons that had wSnL values less than 0.5 showing a bias to non-preferred speed at preferred location (note that for some of these nSL components elicited higher response than SnL; see Fig 8B2 below unity line). These neurons might be selectively coding for the component at their preferred location more than their preferred speed (indicated by "pref. location bias" in Fig 8A quad 4).

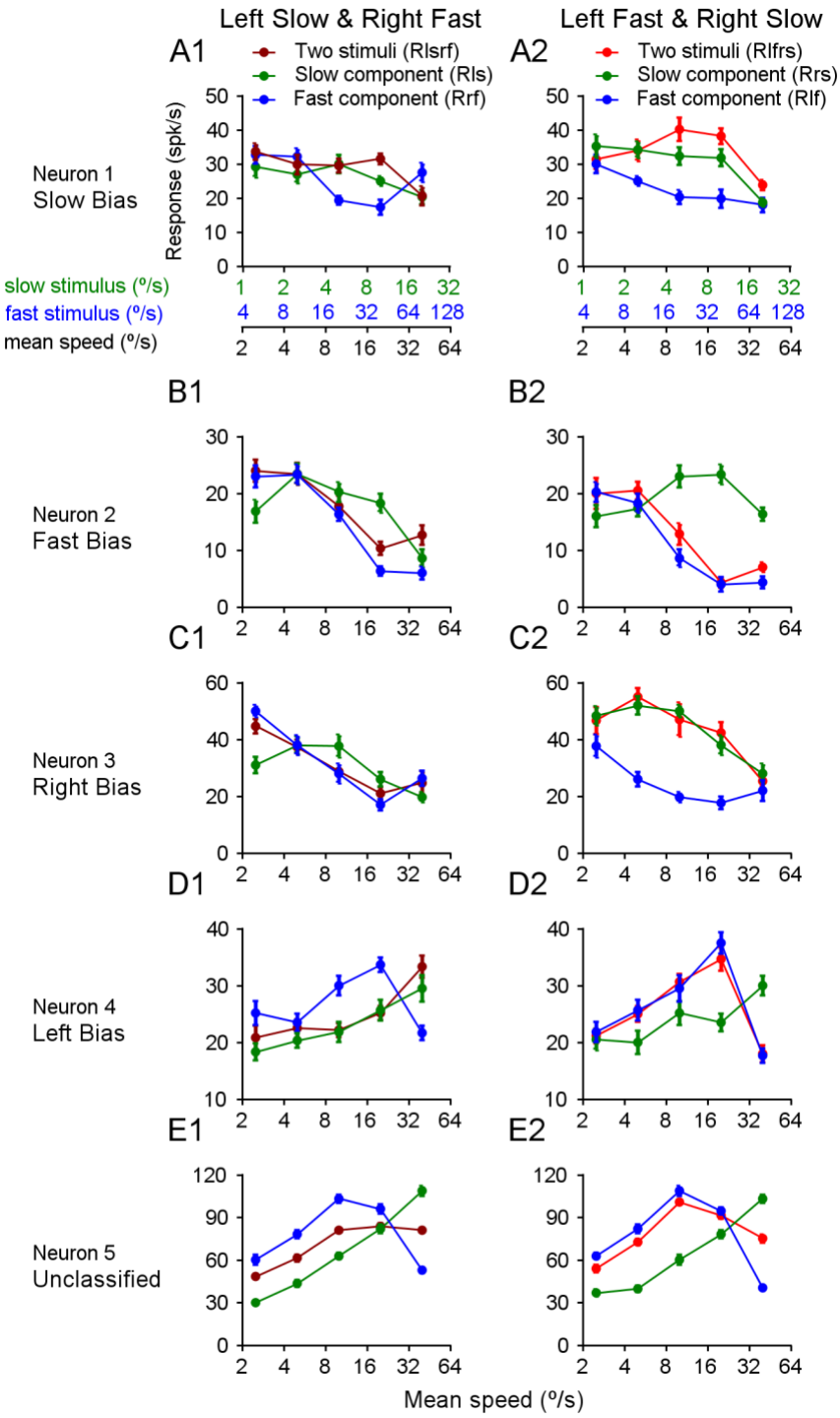

Figure 9. Bi-speed and component speed tuning curves of five example neurons with different classes of tuning biases.

Description same as in Fig 2. (A1-E1) Responses to Left Slow & Right Fast stimuli. (A2-E2) Responses to Left Fast & Right Slow stimuli. (A1, A2) Slow bias neuron. (B1, B2) Fast bias neuron. (C1, C2) Right bias neuron. (D1, D2) Left bias neuron. (E1, E2) Unclassified neuron.

Correlation between bi-speed tuning and component speed tuning curves: classifying neuronal tuning biases to spatially separated bi-speed stimuli

For a subset of neurons, we saw that the shapes of the bi-speed tuning curves were more like one component tuning curve than the other. Example neuronal responses shown in Fig 9 illustrate this- for neuron 1,2,3 and 4, in both spatial arrangements (Fig 9 columns 1 and 2), the bi-speed tuning curves follow the slow, fast, right, and left component speed tuning curves, respectively (Fig 9 A1 & A2, B1 & B2, C1 & C2, and D1 & D2).

To quantify how similar the bi-speed tuning curves to the component single speed tuning curves is, we calculated the partial correlation coefficients of R_{lfrs} with R_{lf} and R_{rs} indicated as zLF and zRS, respectively, and of R_{lsrf} with R_{rf} and R_{ls} indicated as zRF and zLS, respectively (see Methods). To measure the bi-speed tuning correlation bias to the component tuning curves, we defined the following indices:

Two component response bias	Left Slow & Right Fast	Left Fast & Right Slow
Fast bias index	zRF - zLS	zLF - zRS
Right bias index	zRF - zLS	zRS - zLF

Fig 10A1 summarizes these indices compared between the two stimulus configurations. Neurons that followed the fast component tuning more than the slow component tuning, regardless of spatial arrangement (with statistical significance), had high values (above a criterion of 1.28; see Methods) of fast bias index for both stimulus configurations. These are classified as "Fast bias" units. Similarly, those that had lower negative values (<-1.28) are labeled as "Slow bias" units. Neurons with high values (>1.28) of right bias index for both stimulus configurations are classified as "Right bias" units and those with lower negative values (<-1.28) as "Left bias" units. Neurons that passed these criteria for one stimulus configuration but not the other are identified as "Intermediate bias" units with four categories- "Right Fast", "Left Fast", "Right Slow" and "Left Slow" biases. The rest are unclassed.

We saw a heterogenous mix of these classes in our dataset with many of them being unclassified. We observed a few right, left, and fast bias units and only one slow bias unit (owing to the low number units with low preferred speeds < 5 deg/sec). Fig 10A2 shows the spread of these classes with respect to neurons' location preference (lpi) and preferred speeds. There is no systematic relationship between the classes and their speed/location preferences, except left bias and right bias classes generally preferred left and right locations (positive and negative lpi values), respectively. These neuronal classes, designated based on the correlation between tuning to two spatially separated speeds and tuning to one at either location, indicate a bias to a component's speed or location relative to the other, across a range of absolute speed values. Hence, these subgroups of neurons (even though only a few in our dataset) might represent the speed and or the location of the components of two non-overlapping stimuli in a relative sense, generalized over a range of absolute speeds. For example, "Fast bias" neurons might be informative of the faster speed of the two stimuli invariant to the absolute speed of the components (e.g., in 1.25 & 5 deg/s

and 5 & 20 deg/s conditions, the speed 5 deg/s is faster and slower, respectively, a fast bias neuron might simply code for the faster speed i.e., 5 deg/s and 20 deg/s, respectively). Similarly, "Left bias" neurons might be informative of the relative location of the component regardless of whether it is fast, or slow. "Left Fast bias" neurons might code for the conjugation of left location and faster speed over a speed range.

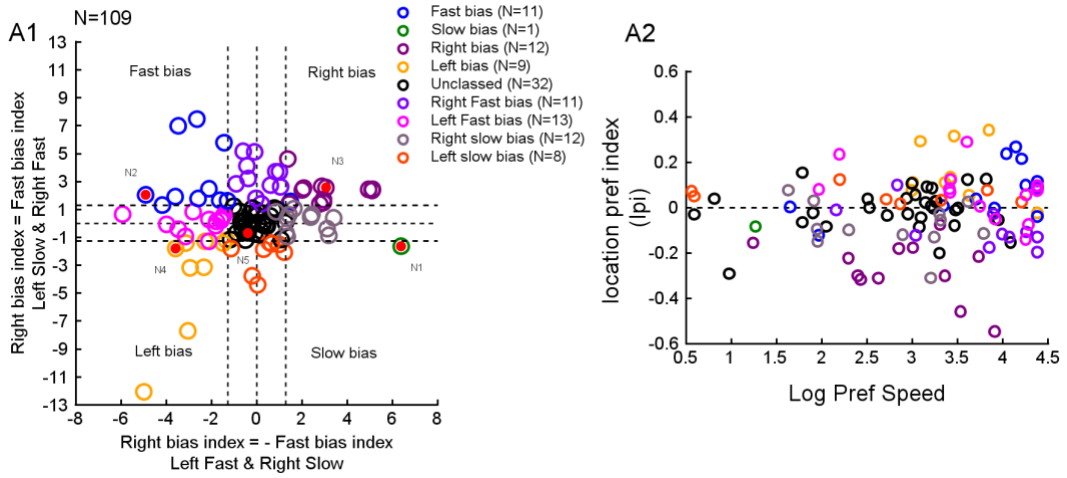

Figure 10. Classification of neurons based on the correlations between their bi-speed tuning curves and component tuning curves.

(A1) Right bias index (zRF – zLS) of Left Slow & Right Fast vs that of Left Fast vs Right Slow. Different colors indicate different classes as indicated in the legend. The red dots mark the values from the example neurons in Fig 6. (A2) Spread of the preferred speeds (x-axis) and location preference indices (y-axis) of different classes of the neurons.

Discussion

We investigated how MT neurons encodes two non-overlapping speeds enclosed in their RFs. A standard way to ascertain neural encoding of multiple stimuli is to understand how neuronal responses to multiple stimuli relate to their responses to the constituent stimuli presented alone (Xiao et al. 2014, Xiao et al. 2015, Wiesner et al. 2020). The responses to multiple stimuli might

equal the average of responses to the components (response averaging) or a weighted sum of responses to the components or might just simply follow the response to one of the components (winner-take-all). We found that at a neuronal population level the encoding rule (defined by this relationship) depended on the speeds of the two stimuli. At low stimulus speeds, the population-averaged responses to two non-overlapping speeds (bi-speed) closely followed the responses to the faster speed when presented alone. At high stimulus speeds, the bi-speed responses followed the responses to the slower speed. Notably, for most neurons, at low stimulus speeds, the faster speed alone elicited a stronger response than the slower speed alone (i.e., faster is the stronger component), and vice-versa was true at high stimulus speeds (i.e., slower is the stronger component). Hence our findings suggest a "stronger-takes-all" behavior (i.e., max-like operation) of MT population averaged responses to two non-overlapping faster and slower components. When the neuronal responses to the individual stimulus components were similar, the response to the bi-speed stimulus exceeded the stronger component response, following a rule of sublinear summation of the component responses.

The bi-speed response following the response of the faster speed at low stimulus speeds is consistent with the bias to faster speed found by Huang et al. (unpublished) when the fast and slow stimuli are overlapping. But, at high stimulus speeds, we found that the response bias flips to the slower component of the non-overlapping stimuli. This is different from the case of overlapping stimuli where the authors see response averaging (i.e., no bias to either component) at high stimulus speeds. How might the encoding of overlapping and non-overlapping speeds relate to the perceptual segmentation or discrimination over a range of speeds? In a follow-up psychophysics experiment, Huang and colleagues found that, as the mean speed of two overlapping speeds increased, human observers' ability to segment two speeds decreased to an extent that at very high

speeds (for combinations of 20 & 40 °/s and 20 & 80 °/s), they could not segment the speeds. Whereas, when the two speeds are spatially separated, one might still perceive two surfaces moving at different speeds, one faster and one slower over a range of speeds (the slower component is likely to be more salient than the faster component at very high speeds). In the same vein, Masson and colleagues (1999) found that thresholds for speed segmentation of two overlapping stimuli increased with increasing speeds while the thresholds for speed discrimination of two non-overlapping stimuli (presented sequentially) remained low over a broader range of stimulus speeds. It remains to be evaluated how the neural encoding of the spatially separated speeds contributes to perceptual discrimination for a range of stimulus speeds.

One striking property of MT population averaged responses to two spatially separated speeds is how little they are influenced by the weaker component responses i.e., the "winner-takes-all" (WTA) behavior seen throughout the motion period. Studies of spatially separated motion stimuli in area MT have found a variety of responses behaviors from averaging of component responses (Recanzone et al.1997 using two discrete spots of light moving in two directions) to power-law summation of component responses with divisive inhibition (Britten & Heuer 1999 using two Gabor patches moving in two directions). Using non-overlapping stimuli akin to ours, but with different motion directions instead of speeds, Wiesner and colleagues (unpublished) found that MT neurons exhibited a robust response bias toward the stimulus component located at their preferred location within RFs, akin to a soft-max operation. What could be the implications of a max-like or WTA encoding rule to the "read-out" of the information about the stimulus components? Microstimulation studies in the past have reported that a WTA mechanism might allow the segregation of distinct motion signals and may therefore aid in their perceptual

discrimination (Groh et al. 1997, Cohen & Newsome 2004). Hence, a WTA encoding in area MT might help segregate the two non-overlapping speeds.

Since the processing of the two non-overlapping speeds might require a spatial comparison of local speed estimates, we characterized the effects of spatial arrangement on the neuronal responses to single speeds and two non-overlapping speeds. We found that subsets of neurons responded more to a single speed presented on the left side of their RFs than the right side, and vice-versa, showing a preference for one side of the RF over the other. We saw that the spatial preferences and speed preferences of neurons to single components were preserved in their responses to the two stimuli with different spatial arrangements. For example, a neuron that preferred a left location (across different speeds) and a faster speed (on left and right) presented alone, showed a higher response to faster on left and slower on right stimuli than to faster on right and slower on left. Such neurons that respond differentially to the two spatial arrangements might code for a particular combination of location and speed. Thus, neuronal groups with different spatial and speed preferences might enable conjoint encoding of spatial and speed information of the two stimuli in their receptive fields in a distributed manner.

We found a small subset of neurons for which the bi-speed tuning curves followed the speed tuning of one component significantly more than the other. This is different from the WTA/ ”stronger-takes-all” behavior seen for most neurons, in the sense that the bi-speed tuning curves followed both the strong and weak portions of the component speed tuning curves (see Fig 7). These neurons might invariantly code for spatial location (left or right) or speed (faster or slower of the two speeds presented) for a range of speeds. It remains to be assessed how these different encoding schemes might be read out by a decoding model that can estimate both the two non-overlapping speeds and their relative placement. Together, these findings reveal a distributed

neural code of multiple non-overlapping stimuli over neurons with different preferences to locations (within RFs) and speeds, in area MT, which might be beneficial for estimating the local speeds tagged with the spatial information of multiple stimuli.

Chapter 4: Discussion and summary

The goal of the studies in my thesis was to understand how neurons in the visual area MT encode multiple moving stimuli and how these encoding rules might benefit the extraction of component stimuli from neuronal populations to facilitate visual segmentation. Our approach of characterizing the full response tuning curves of MT neurons allowed us to characterize the population neuronal responses in area MT to multiple stimuli. Using this approach, we made novel findings that revealed how MT neurons pool the inputs from the components of multiple stimuli and how the encoding rules of MT neurons create a population neural code that facilitates visual segmentation.

In chapter 2, we investigated the role of binocular disparity and selective attention in the neural encoding of two transparently moving surfaces. We designed a novel motion discrimination task where macaque monkeys were cued using disparity to attend one of the two overlapping moving surfaces located at different depths. We determined how much of an influence each component motion direction at near or far disparities has on the responses to overlapping directions. For this, we fitted MT neuronal responses to overlapping motion as a weighted sum of the responses to component directions alone, using the whole direction tuning curves. Higher response weight to one component than the other is termed as bias to that component. We found that the neuronal responses to two overlapping motion directions with disparity cues (bidirectional responses) were more biased to the motion component presented at one disparity than the other. In one animal, the bidirectional responses of MT neurons were biased to the motion at the near disparity. In two other animals, the bidirectional responses were biased to the motion component at a disparity closer to MT neurons' preferred disparities. Nevertheless, across animals, we found that the extent of the bidirectional response bias to one disparity component over the other was

well correlated with the extent of preference to that component presented alone. In other words, near-preferred neurons (i.e., prefer single motion at near disparity than far) showed a higher bias to motion at near disparity of the two overlapping motions than far-preferred neurons. Importantly, the bias toward the more preferred disparity is more than what is expected by simply averaging the responses to the individual components. This implies that MT neurons' intrinsic preferences to single surfaces are preserved in their representation of two simultaneously presented competing surfaces. These preferences to single surfaces are decided by neurons' selectivity to the features of the surfaces i.e., disparity and motion direction in this case.

This bias in neuronal responses to two motion components toward "preferred" component has been a recurring theme of encoding of multiple motion directions, corroborated by other studies in the lab. For example, Xiao and others (2014) showed that when one of the two overlapping motion directions had a higher contrast than the other, MT neurons showed a greater response weight for the component with higher contrast which also elicited the stronger direction tuning response when presented alone. Wiesner and colleagues (SfN abstract 2019) noticed a similar trend when two directions were spatially separated, where there was a robust response bias toward the stimulus component located at the preferred location of MT neurons within their RFs, akin to a soft-max operation. In the same vein, in chapter 3, we found a winner-takes-all scheme in MT neuronal responses to two spatially separated motion speeds, the "winner" stimulus component being the preferred speed placed at the preferred location within the RFs. Together, these findings suggest that the neural encoding of multiple motion stimuli is distributed over neuronal subpopulations, with different subpopulations preferentially representing one motion component over the other. Across a range of motion directions, these subpopulations might be defined by neuronal preferences to other cues (like disparity, 2D spatial separation, or contrast

differences) that help in segmenting and discriminating multiple components. Future experiments are required to determine the causal relationship between this encoding scheme and the perceptual segmentation/discrimination of the stimuli.

While these neurons, at a subpopulation level (e.g., near and far-preferred neurons), showed larger response weights to one component than the other (i.e., bias), it is important to note that neurons within a subpopulation exhibited a range of response biases. This diversity of response weights within each subpopulation might make the visual system more robust to noise, hence increasing the encoding accuracy of multiple stimuli, compared to a system with homogeneous response weights within neuronal groups (theoretically tested by Orhan & Ma, 2015). Furthermore, decoding models incorporating different encoding rules can be tested and compared for their performance in extracting the component stimulus features to evaluate their role in perception.

In chapter 2, we used a divisive normalization model (DNM) framework to capture the heterogeneity of response weights across neurons as a function of their preferences to the components of overlapping stimuli when presented alone. The preferences were reflected by the population neural responses to individual stimulus components, with additional scaling from selective attention. Several studies have used the normalization model to describe visual neurons' responses to two stimuli in their RFs that had different contrasts (Heuer & Britten 2002, Busse et al. 2009, Xiao et al. 2014). The influence of adding a second stimulus to the first, on the neuronal responses, depended on the contrast values of the stimuli. To capture this contrast dependence, these studies used contrast values as gain factors/weights for the excitatory drive from the stimulus components (the numerator of the DNM) and in the normalization signal (the denominator). Our stimuli had the same contrast but differed in both direction and disparity features, and the neuronal responses to the overlapping stimulus across motion directions depended on neuronal disparity

preferences. So, we took a novel approach to replace the contrast values of the stimulus components (referred to as "signal strength") in a typical DNM equation with the population activity elicited by each component summed over a pool of neurons with varying direction and disparity preferences. Additionally, what sets our model apart from the previous adaptations of DNM, is the flexibility to define and vary the size of the pool of neurons that contribute to the gain factor in the numerator, as well as the normalization signal in the denominator of DNM equation. This redefined DNM model, which is currently a fusion of mechanistic and descriptive approaches, can be refined in the future into a biologically plausible framework.

What do we know about the circuit/network mechanisms that enable the encoding rules described above? Medathati and colleagues (2017) proposed a single, adaptive recurrent network where the combined effect of the excitation and inhibition could be tuned to input statistics to implement different computational rules such as vector averaging, winner-take-all or superposition of the components of the multiple overlapping motion stimuli. Their model of recurrent connections in MT was defined only in motion direction space. The model can be extended to combining motion with other features, e.g., binocular disparity to assess the effects of cue combination on the tuning properties resulting from simple recurrent connections within MT.

It is also interesting to test how such a network behaves when multiple stimuli are overlapping vs when multiple stimuli are separated in 2D space within the RFs of MT neurons. For example, in chapter 3, we found a winner-takes-all (WTA) scheme of encoding two spatially separated speeds for different mean speeds tested. When the two speeds were overlapping, the encoding rule switched from a WTA scheme to averaging the component responses as the mean speed increased (see Chapter 3 Fig 5). This suggests a weaker normalization of the inputs when the two stimuli are non-overlapping than when they are overlapping. One distinction between these

two scenarios is whether the feedforward inputs elicited by the component motions in lower visual areas like V1 are mixed or separated. When the components are separated within the MT RFs, two distinct pools of V1 neurons with RFs spread over the two non-overlapping components provide inputs to a given MT neuron. When the components are completely overlapping, the feed forward inputs are mixed. Importantly, in both cases, the interaction between the inputs via normalization mechanisms can happen at the level of V1 and MT. The strength of this normalization mechanism in MT might depend on the mixing or separation of its feed forward inputs, resulting in different response profiles when stimuli are overlapping vs when they are separated. This could be tested by a hierarchical feedforward model simulating the interactions between V1 and MT (Simoncelli & Heeger 1998, Wiesner et al. 2020).

Our study in chapter 2 is one of a kind to probe the effects of attention on the neural representation of surfaces defined by two features (motion and disparity) that are conjointly encoded in one brain area (area MT). It is important to note that even though disparity cues aided in the selection of one of the overlapping surfaces, disparity-based attention itself cannot explain the effects of attention on responses to preferred vs non-preferred direction seen in our data (from the attention modulation index AMI analysis). Selective attention needs to spread from the disparity to direction domain for the monkeys to successfully perform the direction discrimination task using the disparity cues. Several studies in the past reported that attention selectively enhances the neural representation of attended stimuli and reduces the influence of unattended stimuli (Treue & Maunsell 1996, Ferrera & Lisberger 1997, Reynolds et al. 1999, Treue & Martínez Trujillo1999, Recanzone & Wurtz 2000, Li & Basso 2005, Lee & Maunsell 2010). In line with these reports, our results show that selectively attending to one of the two overlapping surfaces moving in different directions presented at two binocular disparities enhanced the representation of the

attended component in area MT. Our results suggest that this enhancement of representation of the attended component could be achieved either by simply enhancing the response weight to a component (seen in Monkey G) or by simply suppressing the response weight to the other component or a combination of the two (seen in Monkey B).

In chapter 2, in one monkey we saw that the bidirectional responses were predominantly biased to the near surface (near bias) of the overlapping motion stimuli. In addition, we found that the monkey had a better performance in discriminating motion directions at the near surface than at the far surface. In another monkey, we saw that the near disparity bias in near-preferred neurons occurred earlier than the far bias in far-preferred neurons and attending near surface also had an earlier effect than attending far surface. Human studies have reported that processing of near disparities might have an advantage over far disparities, using sensitivity and temporal measures (Manning et al. 1987, Becker et al., 1999; Patterson et al. 1995, Landers & Cormack 1997; Birch et al 1982). Previous studies have reported a higher proportion of neurons selective for crossed (i.e., near) disparities than uncrossed (i.e., far) disparities in several brain areas like V1 (Gonzalez et al. 2010, Prince et al. 2002, Samonds et al. 2012), V2 (Prince et al. 2002, Chen et al. 2008), V3 (Adams et al., 2001), V4 (Hinkle et al. 2005, Tanabe et al. 2005), MT (Maunsell & Van Essen 1983, Bradley & Andersen 1998, DeAngelis & Uka 2003), and MST (Gonzalez et al. 2001). Some of these studies have speculated the near bias may underlie the salience of stimuli presented in front of the fixation plane. In the same vein, a recent study on natural scene statistics of depth in figure-ground segregation reported that surfaces labeled as “figures” by human observers tend to be closer to the observers, and therefore occur more commonly at near disparities (Huang et al., SfN abstract 2019). They also found that because figure regions tend to be nearer, world motion in these regions is associated with higher speed. Our results taken together with the studies above,

suggest a representational bias to near surfaces that might efficiently code the natural scene irregularities in the disparity domain. Future experiments are needed to shed light on how disparity and motion statistics of natural scenes may be exploited by cortical neurons to help segregate figures from background and objects from each other.

References

Adams DL, Zeki S. 2001. Functional organization of macaque V3 for stereoscopic depth. *J Neurophysiol* 86: 2195-203

Albright TD. 1984. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol* 52: 1106-30

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: 105-26

Andersen RA, Bradley DC. 1998. Perception of three-dimensional structure from motion. *Trends Cogn Sci* 2: 222-8

Becker S, Bowd C, Shorter S, King K, Patterson R. 1999. Occlusion contributes to temporal processing differences between crossed and uncrossed stereopsis in random-dot displays. *Vision Res* 39: 331-9

Birch EE, Gwiazda J, Held R. 1982. Stereoacuity development for crossed and uncrossed disparities in human infants. *Vision Res* 22: 507-13

Born RT. 2000. Center-surround interactions in the middle temporal visual area of the owl monkey. *J Neurophysiol* 84: 2658-69

Boynton GM. 2009. A framework for describing the effects of attention on visual responses. *Vision Res* 49: 1129-43

Braddick O. 1993. Segmentation versus integration in visual motion processing. *Trends Neurosci* 16: 263-8

Bradley DC, Andersen RA. 1998. Center-surround antagonism based on disparity in primate area MT. *J Neurosci* 18: 7552-65

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

Bradshaw MF, Cumming BG. 1997. The direction of retinal motion facilitates binocular stereopsis. *Proc Biol Sci* 264: 1421-7

Bravo MJ, Watamaniuk SN. 1995. Evidence for two speed signals: a coarse local signal for segregation and a precise global signal for discrimination. *Vision Res* 35: 1691-7

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

Burr D, Thompson P. 2011. Motion psychophysics: 1985-2010. *Vision Res* 51: 1431-56

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

Carandini M, Heeger DJ. 1994. Summation and division by neurons in primate visual cortex. *Science* 264: 1333-6

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

Chen G, Lu HD, Roe AW. 2008. A map for horizontal disparity in monkey V2. *Neuron* 58: 442-50

Chowdhury SA, DeAngelis GC. 2008. Fine discrimination training alters the causal contribution of macaque area MT to depth perception. *Neuron* 60: 367-77

Cohen MR, Newsome WT. 2004. What electrical microstimulation has revealed about the neural basis of cognition. *Curr Opin Neurobiol* 14: 169-77

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

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

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

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

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

Gonzalez F, Bermudez MA, Vicente AF, Romero MC. 2010. Orientation preference and horizontal disparity sensitivity in the monkey visual cortex. *Ophthalmic Physiol Opt* 30: 824-33

Gonzalez F, Perez R, Justo MS, Bermudez MA. 2001. Receptive field organization of disparity-sensitive cells in Macaque medial superior temporal cortex. *Eur J Neurosci* 14: 167-73

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

Greenwood JA, Edwards M. 2006a. An extension of the transparent-motion detection limit using speed-tuned global-motion systems. *Vision Res* 46: 1440-9

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

Groh JM, Born RT, Newsome WT. 1997. How is a sensory map read Out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *J Neurosci* 17: 4312-30

Grossberg S. 1994. 3-D vision and figure-ground separation by visual cortex. *Percept Psychophys* 55: 48-121

Harrison WJ, Ayeni AJ, Bex PJ. 2019. Attentional selection and illusory surface appearance. *Sci Rep* 9: 2227

Heeger DJ. 1992. Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9: 181-97

Hejja-Brichard Y, Rima S, Rapha E, Durand JB, Cottereau BR. 2020. Stereomotion Processing in the Nonhuman Primate Brain. *Cereb Cortex* 30: 4528-43

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

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

Hinkle DA, Connor CE. 2005. Quantitative characterization of disparity tuning in ventral pathway area V4. *J Neurophysiol* 94: 2726-37

Howard IP, Rogers BJ, Oxford University Press. 1995. Binocular vision and stereopsis. In *Oxford psychology series no 29*, pp. 1 online resource (736 p. 4 p. New York ; Oxford: Oxford University Press,

Huang X, Albright TD, Stoner GR. 2007. Adaptive surround modulation in cortical area MT. *Neuron* 53: 761-70

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

Kim HR, Angelaki DE, DeAngelis GC. 2015. A functional link between MT neurons and depth perception based on motion parallax. *J Neurosci* 35: 2766-77

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

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

Kumano H, Uka T. 2013. Responses to random dot motion reveal prevalence of pattern-motion selectivity in area MT. *J Neurosci* 33: 15161-70

Lagae L, Maes H, Raiguel S, Xiao DK, Orban GA. 1994. Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. *J Neurophysiol* 71: 1597-626

Lamme VA, van Dijk BW, Spekreijse H. 1993. Contour from motion processing occurs in primary visual cortex. *Nature* 363: 541-3

Landers DD, Cormack LK. 1997. Asymmetries and errors in perception of depth from disparity suggest a multicomponent model of disparity processing. *Percept Psychophys* 59: 219-31

Lee J, Maunsell JH. 2010. The effect of attention on neuronal responses to high and low contrast stimuli. *J Neurophysiol* 104: 960-71

Li X, Basso MA. 2005. Competitive stimulus interactions within single response fields of superior colliculus neurons. *J Neurosci* 25: 11357-73

Lisberger SG, Movshon JA. 1999. Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. *J Neurosci* 19: 2224-46

Liu J, Newsome WT. 2005. Correlation between speed perception and neural activity in the middle temporal visual area. *J Neurosci* 25: 711-22

Liu LD, Pack CC. 2017. The Contribution of Area MT to Visual Motion Perception Depends on Training. *Neuron* 95: 436-46 e3

Manning ML, Finlay DC, Neill RA, Frost BG. 1987. Detection threshold differences to crossed and uncrossed disparities. *Vision Res* 27: 1683-6

Masson GS, Mestre DR, Stone LS. 1999. Speed tuning of motion segmentation and discrimination. *Vision Res* 39: 4297-308

Maunsell JH, Van Essen DC. 1983a. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol* 49: 1127-47

Maunsell JH, Van Essen DC. 1983b. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol* 49: 1148-67

McDonald JS, Clifford CW, Solomon SS, Chen SC, Solomon SG. 2014. Integration and segregation of multiple motion signals by neurons in area MT of primate. *J Neurophysiol* 111: 369-78

Medathati NVK, Rankin J, Meso AI, Kornprobst P, Masson GS. 2017. Recurrent network dynamics reconciles visual motion segmentation and integration. *Sci Rep* 7: 11270

Mestre DR, Masson GS, Stone LS. 2001. Spatial scale of motion segmentation from speed cues. *Vision Res* 41: 2697-713

Mikami A, Newsome WT, Wurtz RH. 1986. Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *J Neurophysiol* 55: 1308-27

Moller P, Hurlbert AC. 1997. Motion edges and regions guide image segmentation by colour. *Proc Biol Sci* 264: 1571-7
Mysore SG, Vogels R, Raiguel SE, Orban GA. 2006. Processing of kinetic boundaries in macaque V4. *J Neurophysiol* 95: 1864-80
Nawrot M, Blake R. 1989. Neural integration of information specifying structure from stereopsis and motion. *Science* 244: 716-8
Ni AM, Ray S, Maunsell JH. 2012. Tuned normalization explains the size of attention modulations. *Neuron* 73: 803-13
Nienborg H, Cumming BG. 2007. Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. *Nat Neurosci* 10: 1608-14
Nover H, Anderson CH, DeAngelis GC. 2005. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J Neurosci* 25: 10049-60
Ohshiro T, Angelaki DE, DeAngelis GC. 2011. A normalization model of multisensory integration. *Nat Neurosci* 14: 775-82
Orban GA, Saunders RC, Vandenbussche E. 1995. Lesions of the superior temporal cortical motion areas impair speed discrimination in the macaque monkey. *Eur J Neurosci* 7: 2261-76
Orhan AE, Ma WJ. 2015. Neural population coding of multiple stimuli. *J Neurosci* 35: 3825-41
Pasternak T, Merigan WH. 1994. Motion perception following lesions of the superior temporal sulcus in the monkey. *Cereb Cortex* 4: 247-59
Patterson R, Cayko R, Short GL, Flanagan R, Moe L, et al. 1995. Temporal integration differences between crossed and uncrossed stereoscopic mechanisms. *Percept Psychophys* 57: 891-7
Prince SJ, Cumming BG, Parker AJ. 2002. Range and mechanism of encoding of horizontal disparity in macaque V1. *J Neurophysiol* 87: 209-21
Qian N, Andersen RA. 1994. Transparent motion perception as detection of unbalanced motion signals. II. Physiology. *J Neurosci* 14: 7367-80
Read J. 2005. Early computational processing in binocular vision and depth perception. *Prog Biophys Mol Biol* 87: 77-108
Recanzone GH, Wurtz RH. 2000. Effects of attention on MT and MST neuronal activity during pursuit initiation. *J Neurophysiol* 83: 777-90
Recanzone GH, Wurtz RH, Schwarz U. 1997. Responses of MT and MST neurons to one and two moving objects in the receptive field. *J Neurophysiol* 78: 2904-15
Reynolds JH, Chelazzi L, Desimone R. 1999. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19: 1736-53
Reynolds JH, Heeger DJ. 2009. The normalization model of attention. *Neuron* 61: 168-85
Rosenberg A, Wallisch P, Bradley DC. 2008. Responses to direction and transparent motion stimuli in area FST of the macaque. *Vis Neurosci* 25: 187-95
Sachtler WL, Zaidi Q. 1995. Visual processing of motion boundaries. *Vision Res* 35: 807-26
Salzman CD, Murasugi CM, Britten KH, Newsome WT. 1992. Microstimulation in visual area MT: effects on direction discrimination performance. *J Neurosci* 12: 2331-55
Samonds JM, Potetz BR, Lee TS. 2012. Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. *Proc Natl Acad Sci U S A* 109: 6313-8

Simoncelli EP, Heeger DJ. 1998. A model of neuronal responses in visual area MT. *Vision Res* 38: 743-61

Smith MA, Majaj NJ, Movshon JA. 2005. Dynamics of motion signaling by neurons in macaque area MT. *Nat Neurosci* 8: 220-8

Smolyanskaya A, Ruff DA, Born RT. 2013. Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable. *J Neurophysiol* 110: 2806-16

Snowden RJ, Rossiter MC. 1999. Stereoscopic depth cues can segment motion information. *Perception* 28: 193-201

Snowden RJ, Treue S, Erickson RG, Andersen RA. 1991. The response of area MT and V1 neurons to transparent motion. *J Neurosci* 11: 2768-85

Solomon SS, Tailby C, Gharaei S, Camp AJ, Bourne JA, Solomon SG. 2011. Visual motion integration by neurons in the middle temporal area of a New World monkey, the marmoset. *J Physiol* 589: 5741-58

Stoner GR, Albright TD. 1998. Luminance contrast affects motion coherency in plaid patterns by acting as a depth-from-occlusion cue. *Vision Res* 38: 387-401

Tanabe S, Doi T, Umeda K, Fujita I. 2005. Disparity-tuning characteristics of neuronal responses to dynamic random-dot stereograms in macaque visual area V4. *J Neurophysiol* 94: 2683-99

Treue S, Hol K, Rauber HJ. 2000. Seeing multiple directions of motion-physiology and psychophysics. *Nat Neurosci* 3: 270-6

Treue S, Martinez Trujillo JC. 1999. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399: 575-9

Treue S, Maunsell JH. 1996. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382: 539-41

Tse PU. 2005. Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Res* 45: 1095-8

Uka T, DeAngelis GC. 2003. Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *J Neurosci* 23: 3515-30

Uka T, DeAngelis GC. 2004. Contribution of area MT to stereoscopic depth perception: choice-related response modulations reflect task strategy. *Neuron* 42: 297-310

Valdes-Sosa M, Bobes MA, Rodriguez V, Pinilla T. 1998a. Switching attention without shifting the spotlight object-based attentional modulation of brain potentials. *J Cogn Neurosci* 10: 137-51

Valdes-Sosa M, Cobo A, Pinilla T. 1998b. Transparent motion and object-based attention. *Cognition* 66: B13-B23

Valdes-Sosa M, Cobo A, Pinilla T. 2000. Attention to object files defined by transparent motion. *J Exp Psychol Hum Percept Perform* 26: 488-505

Wallace JM, Mamassian P. 2003. The efficiency of speed discrimination for coherent and transparent motion. *Vision Res* 43: 2795-810

Wallace JM, Mamassian P. 2004. The efficiency of depth discrimination for non-transparent and transparent stereoscopic surfaces. *Vision Res* 44: 2253-67

Wannig A, Rodriguez V, Freiwald WA. 2007. Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron* 54: 639-51

Wiesner S, Baumgart IW, Huang X. 2020. Spatial Arrangement Drastically Changes the Neural Representation of Multiple Visual Stimuli That Compete in More Than One Feature Domain. *J Neurosci* 40: 1834-48

Wiesner S, Huang X. 2019. Image segmentation and neural representation of multiple spatially-separated stimuli in cortical area MT. Society for Neuroscience annual meeting.

Xiao DK, Marcar VL, Raiguel SE, Orban GA. 1997. Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. *Eur J Neurosci* 9: 956-64

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

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

Yan Y, Rasch MJ, Chen M, Xiang X, Huang M, et al. 2014. Perceptual training continuously refines neuronal population codes in primary visual cortex. *Nat Neurosci* 17: 1380-7

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