

Neural Representation of Multiple Moving Features in Extrastriate Visual Cortical Area MT

By
Jianbo Xiao

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
Physiology

at the
UNIVERSITY OF WISCONSIN-MADISON
2017

Date of final oral examination: 12/8/2016

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

Xin Huang, Associate Professor, Neuroscience

Tom Yin, Professor, Neuroscience

Bradley R. Postle, Professor, Psychology and Psychiatry

Matthew I. Banks, Associate Professor, Anesthesiology

Luis C. Populin, Associate Professor, Neuroscience

Dedication

This dissertation is dedicated to the memory of my father, Yunsheng Xiao.

Acknowledgements

This dissertation would not have been possible without the support and advice of many people.

First of all, I would like to thank my advisor, Xin Huang, for his guidance and constant support. Inside the lab, he taught me how to think rigorously, how to ask hard questions, how to design experiments, and how to efficiently present findings; Outside the lab, he cares about my career interests, supports me to explore a broader research agenda and provides me with tremendous support during the difficult time when my father was terminally ill. Thank you, Xin. I will cherish my time with the lab – discussing project late into the night, isolating a beautiful neuron after multiple tries, and having beer together at terrace.

I would like to thank Donata Oertel for her support and encouragement. She has attended pretty much all the talks that I have given during the past six and half years and always gives me feedback. She provides huge support to help me get through the difficult time when my father was ill. Thank you, Donata.

I would like to thank the members of my thesis committee – Tom Yin, Brad Postle, Matt Banks and Luis Populin – for their guidance on my PhD training and constructive feedback on my dissertation research. Their discussion and ideas have been invaluable.

I would like to thank all my colleagues in the Huang lab – especially Steven Wisner, Jenenifer Gaudio, Jason Chuang, Connie Chuang, and Courtney Schwebach – for assisting with animal training, data collection and providing feedback on my writing and presentation. Our friendship makes it a joy to work in the lab.

I would like to thank Mark Marohl, Rebecca Welch, Mary Walker, Sue Krey, Lisa Goldsby, Physiology Graduate Training Program and Department of Neuroscience for their support on my education and research. I also would like to thank Ravi Kochhar, Daniel Yee and all the animal care staff for their excellent technical assistance.

Finally, I am grateful for the support and love of my mother, father, sister and my “HappyHeart” Qianli. You are the reason that keeps me pushing and facing hardships without fear.

Table of Contents

Acknowledgements.....	ii
Table of contents.....	iii
Abstract	iv
Chapter 1.....	1
Introduction	
Chapter 2	
Normalization of neuronal responses in cortical area MT across signal strengths and motion directions.....	31
Chapter 3	
Distributed and dynamic neural encoding of multiple motion directions of transparently moving stimuli in macaque cortical area MT	79
Chapter 4	
Neural representation of multiple motion directions at different depths and the effects of feature-based attention	130
Appendixes	159

Abstract

Natural scenes often contain multiple entities. The ability to segregate visual scenes into distinct objects and surfaces, referred to as visual segmentation, is fundamental to vision. Transparent motion perception is the perception of multiple motion vectors within the same visual space. Transparent motion presents a difficult case for visual segmentation. Understanding neural representation of transparent motion is important to elucidating how the brain distinguishes multiple stimuli. In this dissertation, I studied the neural representation of transparent motion by recording from neurons in the middle temporal (MT) cortex of macaque monkeys while they performed either a visual fixation task or a motion discrimination task. In the first project, I investigated the neural representation of two transparently moving stimuli that have different signal strengths. I found that neuronal responses to the transparent motion stimuli can be well accounted for by the weighted summation of the component responses plus a non-linear interaction term between the component responses, and MT neurons weight the stimulus component that has the higher signal strength more strongly. In the second project, I investigated how MT neurons represent overlapping random-dot stimuli moving transparently in slightly different directions. I found that, although the population-averaged neuronal activity represents the vector-averaged direction of the two motion components, more than half of MT neurons preferentially represent the component directions. In the third project, I studied the neural representation of overlapping stimuli moving transparently at different depths and the effects of feature-based attention to understand how different motion features, which in this case are motion direction and binocular disparity, interact with each other to achieve visual segmentation. I found a tuned feature-based attention effect and that MT neurons preferentially represent the near stimulus component regardless of the attentional state and the neuron's binocular disparity selectivity. This dissertation research demonstrated the success of divisive normalization model in characterizing the neuronal representation of transparent motion, and discovered sub-populations of MT neurons that were

capable of discriminating similar motion directions, and illustrated the dominant neural representation of the near component stimulus in response to multiple stimuli moving transparently at different depths.

Chapter 1

Introduction

1.1 Visual Segmentation

1.1.1 Importance of Visual Segmentation

In our daily lives, the visual scenes often contain multiple entities. The ability to segregate visual scenes into distinct objects and surfaces, referred to as visual segmentation, is fundamental to visual perception. Visual motion information provides a strong cue for segmenting visual scenes. Imagine how we usually spot the fast-moving shooting star from a night sky with bright moonlight. Imagine how a predator identifies running prey in the jungle. Also, imagine how we find a quick route while we are navigating through a big crowd in the airport to catch our flight. In these cases, it is the motion information that helps us segment the complex visual scene and make a visually guided behavioral decision. However, many aspects of the neural substrate of visual segmentation of multiple moving objects remain unclear. The segmentation of different moving stimuli is called motion segmentation. This thesis seeks to understand how motion segmentation is achieved based on the neural representation of multiple visual motion stimuli in the Middle Temporal cortex (area MT), which is known for processing visual motion information (Stoner & Albright 1993).

We first investigated the neural mechanism that governs the representation of multiple moving visual stimuli across motion directions and signal strengths. Next, we turned to the problem of discriminating two similar motion directions. Given that the neurons in area MT are broadly tuned to motion direction and neural responses to two similar motion directions usually show a uni-modal profile, it is important to understand how to segment similar motion directions based on population neuronal activity. Finally, to elucidate how motion segmentation is achieved in the presence of multiple visual cues and through the interaction between the top-down attentional modulation and the bottom-up stimulus-driven

effect, we studied the joint representation of motion direction and binocular disparity to superimposed stimuli moving at different depths and the feature-based attention effect.

1.1.2 Transparent Motion

Transparent motion occurs when multiple velocity vectors are presented simultaneously within the same visual field, and it poses a difficult problem for visual segmentation. A moving patch of random dots produces the impression of a moving surface. When two overlapping patches of random dots move in different directions within the same spatial region, the perception of two surfaces moving transparently in different directions is generated (Figure 1.1). This ability to perceive more than one motion vector within the same spatial region is called transparent motion perception.

In my study, I employ the transparent motion as the model to investigate the neural mechanism for visual segmentation. As an example of visual segmentation, the transparent motion is challenging because the motion velocity is the only cue for segmenting one object from another. It is therefore also a good model to study the neural mechanism of visual segmentation since only one visual cue is involved.

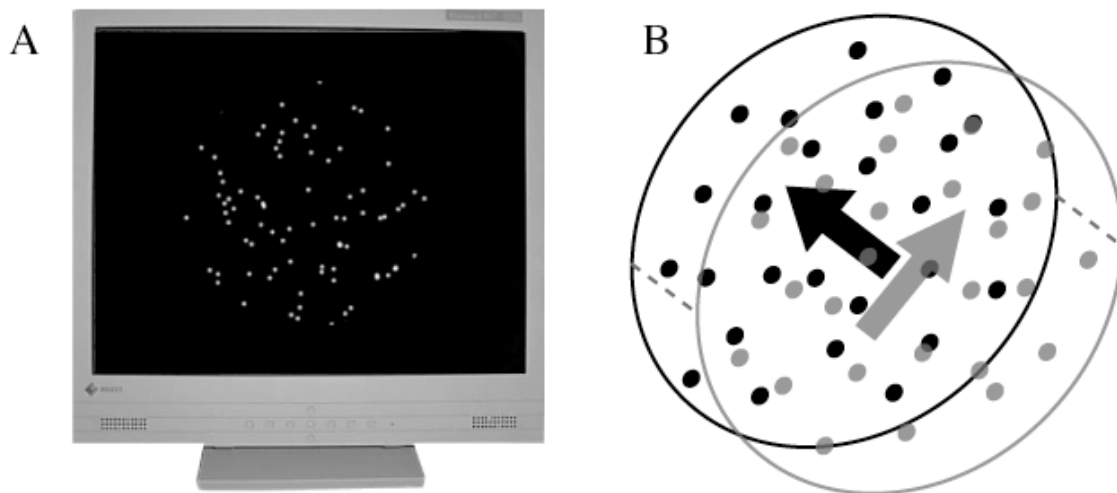


Figure 1.1: **A.** random-dot patch shown on a computer screen; **B.** two random-dot patches moving transparently in two directions. (from Corrado, 1993)

1.1.3 Visual Stimuli for studying Segmentation

Variety types of visual stimuli have been applied to studying the neural mechanisms of visual segmentation and visual integration.

The major visual stimulus used in my study is the random-dot pattern (also called random-dot patch) (Figure 1.1). It can drive most V1 and MT neurons and has been used extensively in studying the neuronal responses in the primary visual cortex (V1) and MT to transparent motion stimuli. Perceptually two superimposed random-dot patterns moving with different velocities generates reliable transparent motion perception with one dot patch sliding on top of another. Therefore, it is an ideal setup for studying the neural mechanisms of motion segmentation.

Another stimulus used in my study is sinusoidal grating, which is also known to drive V1 and MT neurons (Hubel & Wiesel 1962) and has been widely used to studying the neural mechanism of motion integration and segmentation. A plaid is formed when two gratings are superimposed. As a result of motion integration, the perceived motion velocity of the plaid usually is the vector-averaged direction of the directions of two component gratings (Figure 1.2 c). However, varying the luminance of the interaction spot of the two component gratings can change the perception from a coherently moving plaid to two separate component gratings (Stoner & Albright 1993, Stoner et al 1990) (Figure 1.2 d). To understand the neural substrate for motion integration and motion segmentation, previous study has examined the responses of MT neurons to plaid and grating stimuli and found that about 25% MT neurons selectively respond to the velocity of the plaid while another 40% is selective for the velocity of the component gratings and the remaining are unclassified cells (Smith et al 2005). This pattern-component selectivity is recognized as a potential neural substrate for motion integration and motion segmentation. Therefore, I also identified the neuron's pattern-component selectivity with plaid and grating stimuli in my study.

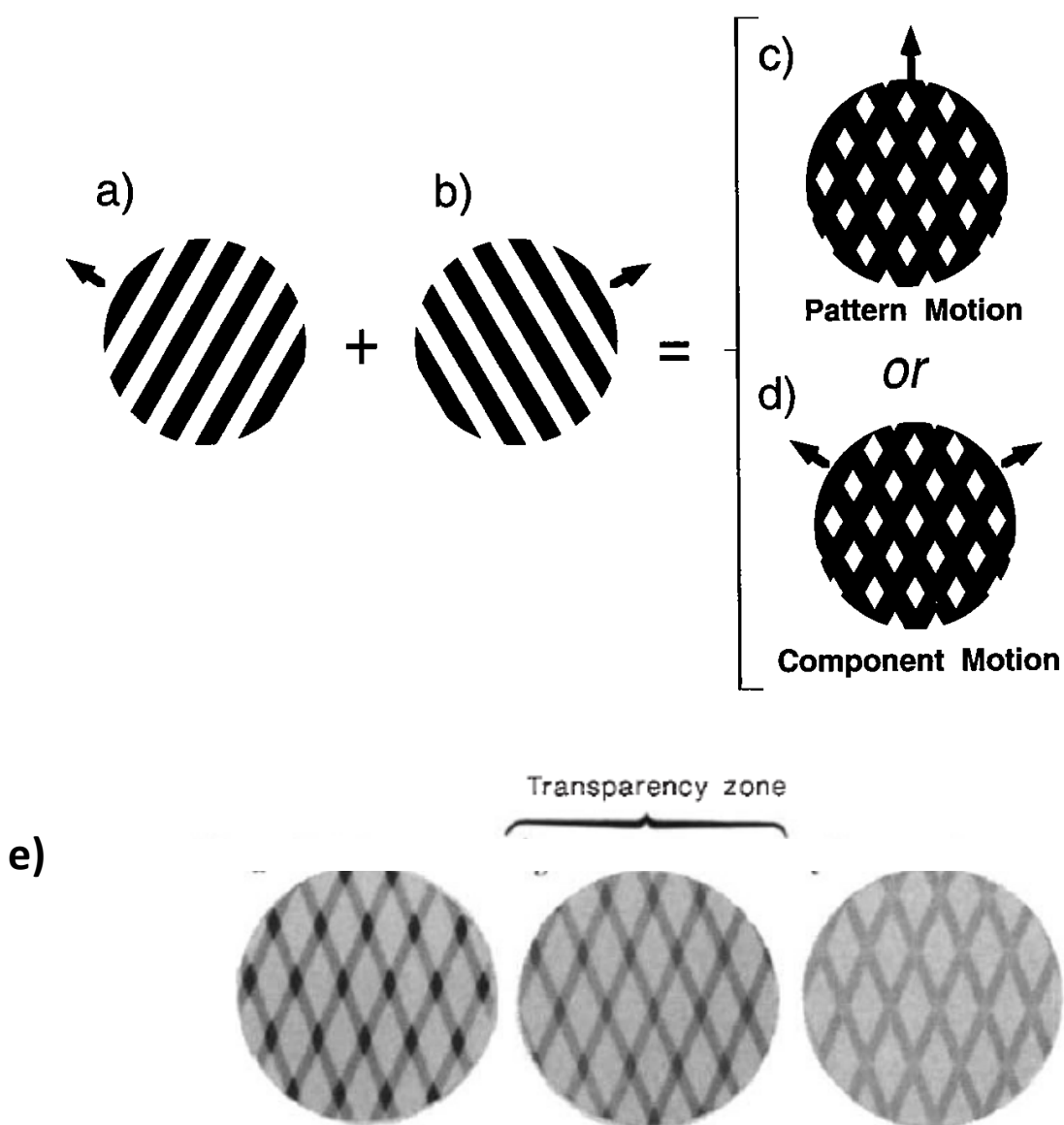


Figure 1.2: *a,b*. Moving plaid patterns are produced by superimposing two drifting gratings. The resultant pattern can be either perceived as a coherent moving plaid (shown in *c*) or two separated component gratings (shown in *d*) depending on a variety of stimulus parameters, such as the luminance of intersections shown in *e*. (adapted from Stoner & Albright 1993, Stoner et al 1990)

1.2 Middle-temporal cortex (area MT)

Over 90% of neurons in the middle temporal cortex (area MT) are motion-selective (Albright 1984, Born & Bradley 2005). Area MT is recognized as the hub for visual motion processing, and it has been widely studied to understand the neural mechanisms of motion perception.

1.2.1 Anatomy

Area MT is located at the posterior bank of the superior temporal sulcus and it receives inputs from both cortical and subcortical regions and sends outputs to regions that are related to the analysis of optic flow (e.g., MST and VIP) and the generation of eye movements (e.g., LIP, FEF, SC, dorsolateral pons) and many other regions like V4, V4t and FST. Its corticocortical connections identify it as one of the main inputs into the dorsal or posterior parietal processing stream (Born & Bradley 2005).

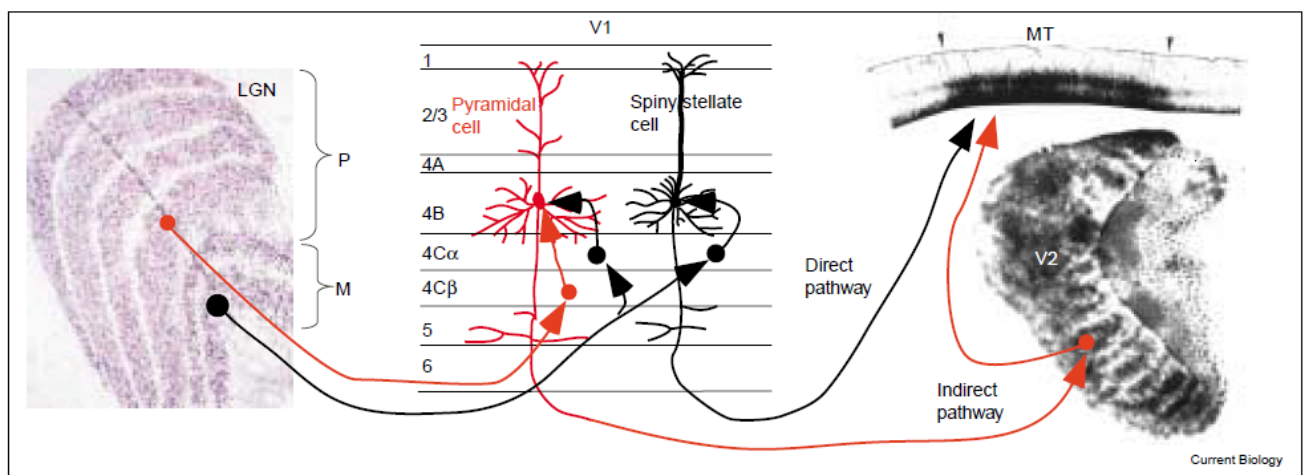


Figure 1.3, Two parallel input pathways to area MT are shown in red (indirect) and black (direct). The magnocellular (M) cells project to layer 4Ca of V1, which projects to both pyramidal (red) and spiny stellate (black) cells of layer 4B. The parvocellular (P) cells project to layer 4CB and then to the pyramidal cells of layer 4B. The spiny stellate cells then send their signals directly to MT. The pyramidal cells relay their presumably mixed M and P signals to MT indirectly via either the thick stripes of V2 or V3 (not shown). (from Born 2001).

Area MT receives inputs from several cortical areas, including V1, V2, V3, V3A, VP and PIP (Figure 1.3 & 1.4) (Born & Bradley 2005); among which the direct input from V1 is the largest one (Maunsell & Van Essen 1983). In particular, the direct inputs from V1 are predominately from spiny stellate cells in layer 4B and have larger axons, larger and more complex terminal boutons than those from V2 and V3, and often form multiple synapses on a single MT neuron (Anderson & Martin 2002, Rockland 1995, Shipp & Zeki 1989). These MT-projecting 4B neurons also appear to receive exclusively magnocellular-inputs (M-inputs) via layer 4Ca (Yabuta et al 2001). In addition, MT receives important indirect cortical

inputs via V3 (Maunsell, 1983) and the thick cytochrome oxidase stripes of V2 (Shipp & Zeki 1989). These indirect inputs also originate from V1 but from a mostly distinct population of neurons within layer 4B (Sincich & Horton 2003) that receives a mixed M and P input (though still dominated by M) (Yabuta et al 2001) and is preferentially distributed beneath the inter-blob regions (Sincich & Horton 2003). These MT-projecting V1, V2, and V3 cells predominantly locate above granular layer IV, while MT sends feedback to neurons located at the supra-granular and infra-granular layers of these areas (Born & Bradley 2005).

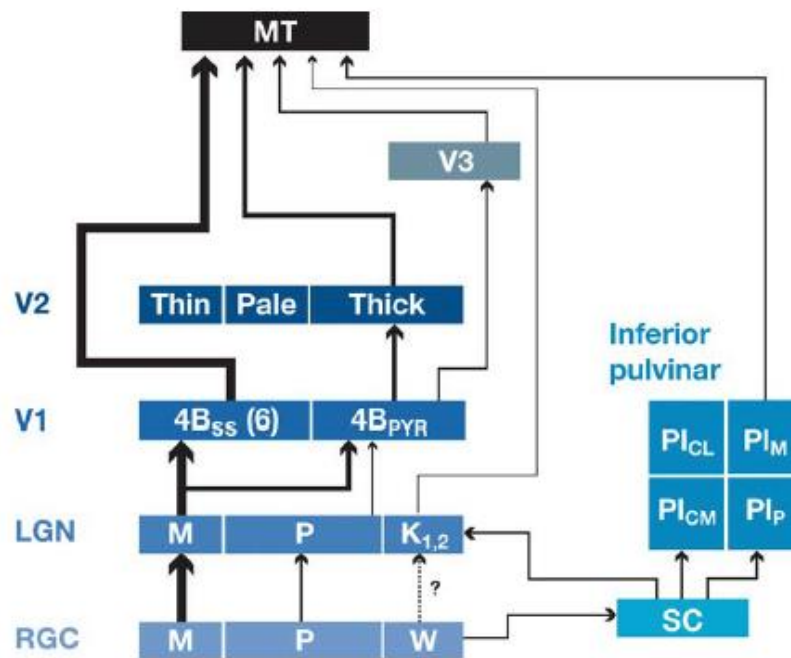


Figure 1.4: Gestalt map of major input pathways to MT. Line thickness is roughly proportional to the magnitude of the inputs. The thickest lines represent the direct cortical pathway (from Born & Bradley 2005).

MT also receives inputs from subcortical areas (Figure 1.3 & 1.4), including superior colliculus (SC) (Standage & Benevento 1983) and the koniocellular neurons in LGN (Fries 1981), which, in theory, explains the remaining motion selectivity in MT after V1 lesions. For example, a previous study has shown

that the majority of the direction selectivity of area MT would be preserved even with the inactivation of area V1, but is reinforced by the input from V1 (Girard et al 1992). In addition, previous studies have indicated that the pathway connecting the subcortical area with MT, such as the pulvinar pathway from SC to MT, is important for visual attention (Berman & Wurtz 2010, Petersen et al 1985b).

MT is also thought to receive top-down feedback signals from frontal cortex, which may serve as the substrates for the observed attention effect in area MT (Buschman & Miller 2007). (Ninomiya et al 2012) applied the retrograde trans-synaptic tracer to study the architecture of multisynaptic top-down pathways to MT and they found that MT receives inputs from area 46v and supplementary eye field (SEF) through frontal eye field (FEF) and lateral intraparietal area (LIP). The inputs from superior colliculus (SC) to MT is also found to be related to the attention effect observed in area MT (Lovejoy & Krauzlis 2010). Future study is required to compare the specific contribution of different pathways to the implementation visual attention in area MT.

1.2.2 Neurophysiology

Area MT is characterized by a high portion of motion selective cells. (Dubner & Zeki 1971) first reported that a preponderance of MT neurons is selective for the motion direction of visual stimuli within their receptive fields, and most of MT neurons are also selective for motion speed (Baker et al 1981, Felleman & Kaas 1984, Maunsell & Van Essen 1983), as shown in **Figure 1.5 A. B.** Later studies provided evidence showing that neural activity in area MT is associated with monkey's decision in a visual motion discrimination task. For example, monkeys were more likely to make a decision in favor of the preferred direction of a neuron when the neuron was firing more vigorously (Shadlen & Newsome 1996) and that magnetic stimulation of area MT (V5) in humans caused deterioration of motion detection (Beckers & Zeki 1995). In addition to the real external stimulus, MT activity also reflects the cognitive experience. For example, another study (Schlack & Albright 2007) has shown that MT neurons which are not responsive

to stationary stimuli would respond to stationary stimulus after the animal was well trained to associate a pattern of moving dots with a stationary arrow. In contrast to the strong selectivity for motion, MT neurons are insensitive to the change of color (Zeki 1983) and form (Albright 1984).

Early study also has shown that MT neurons within the same vertical column perpendicular to the surface tend to have similar motion selectivity and differ from nearby columns in a continuous or systematic manner (Albright 1984). This columnar organization a common feature of the organization of cerebral cortex. For instance, in primary visual cortex (Hubel & Wiesel 1977, Hubener et al 1997), somatosensory (Sretavan & Dykes 1983) and auditory cortex (Abeles & Goldstein 1970), neurons within a column normal to the cortical surface have similar functional properties, and these properties usually vary systematically from column to column, thus forming a topographic map (Mountcastle 1997).

Along with motion direction selectivity, most MT neurons are found to be selective to motion speed. Maunsell and Van Essen (1983) found that MT neurons were generally broadly speed tuned. Their speed tuning can be characterized as low-pass, high-pass and band-pass (Lagae et al 1993) and in general, most MT neurons prefer low speed in the range of 2~32°/s (Priebe & Lisberger 2004). Early studies also have demonstrated that speed perception in macaques can be explained by a firing rate weighted average preferred speed of MT neurons (Priebe & Lisberger 2004). However, in contrast to direction selectivity, neural mechanisms of speed selectivity are less well understood.

In addition to motion, MT neurons are known to be selective to binocular disparity (DeAngelis & Newsome 1999, DeAngelis & Uka 2003) (Figure 1.5). The binocular disparity is the small positional differences between images formed on the two retinae, which generates the perception of depth, also called Stereopsis. Using random-dot stimuli, DeAngelis and Newsome (1999) reported that disparity selectivity often occurred in discrete patches (typically 0.5-1 mm in extent) that were interspersed among similar-sized patches of cortex with weak disparity tuning. Within the disparity tuned patches, preferred disparities changed smoothly across the surface of MT, but there was little change in disparity selectivity along

penetrations normal to the cortical surface, which strongly suggests there are disparity columns in MT, in addition to the well-known columns for the direction of motion. Along with MT neurons, neurons in area V1 (Prince et al 2002b) and the thick stripes of V2 (Hubel & Livingstone 1987, Peterhans & Vonderheydt 1993) are also disparity selective, but weaker than MT neurons. Empirical study has found that there are more cells preferring “Near” disparity (depth) than cells preferring “Far” disparity (depth) in area MT (DeAngelis & Newsome 1999, DeAngelis & Uka 2003). In addition to area MT, there are also more disparity-selective neurons in area V1 and V4 preferring “Near” depth than “Far” depth (Prince et al 2002a, Prince et al 2002b, Tanabe et al 2005) and this bias toward “Near” is even stronger in area V4 than in area MT. However, the neural substrate for this predominance of “Near” disparity-selective cells remains unclear.

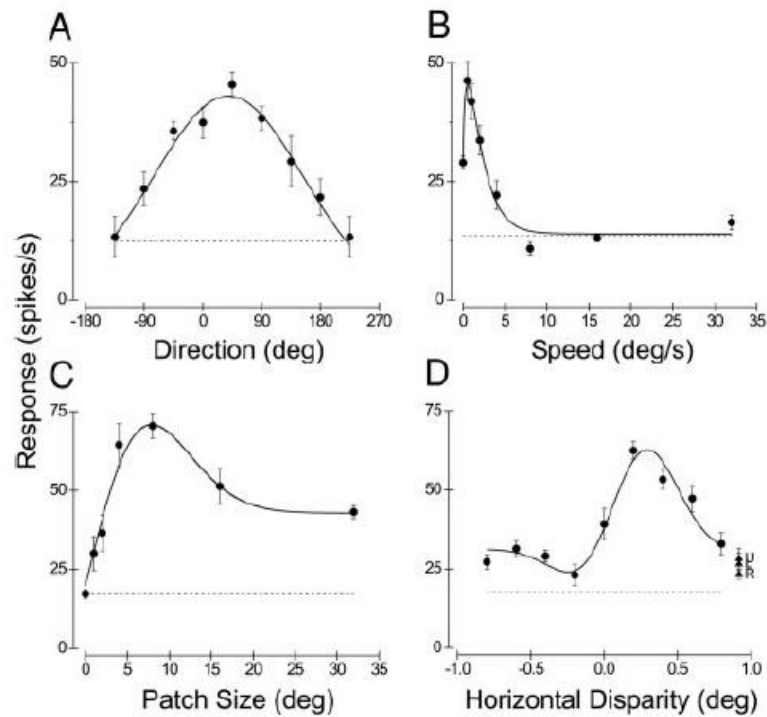


Figure 1.5: Example MT neuron. A. Direction tuning; B. Speed tuning; C. Size tuning; D. Disparity tuning. (from DeAngelis & Uka 2003)

1.2.3 Heterogeneity

The heterogeneity of MT neuron property has been widely documented. Although most MT neurons are motion selective or disparity selective, there are still neurons not selective for motion direction or binocular disparity. For motion direction selective MT neurons, ~83% is also orientation selective to stationary oriented stimuli (Albright 1984). Furthermore, the same study by Albright (1984) shows that 61% of MT neurons have preferred orientation nearly perpendicular to its preferred direction of motion, as is the case for all V1 neurons. However, another 29% of MT neurons have an orientation preference roughly parallel to the preferred direction of motion.

It has also shown that about half MT neurons preferred the same direction of motion throughout a contiguous region of space. This type of receptive field is referred to as “simple” receptive field. While the preferred direction of the other half of MT neurons changed across the RF and there were clear gaps in the RF where neuronal activity is much harder to drive by visual motion. This type of receptive field is referred to as “complex” receptive field (Richert et al 2013).

In response to the superimposed plaid stimuli, (Stoner & Albright 1996) showed that the response tuning of MT neurons can reflect the coherent and non-coherent perception of two superimposed grating stimuli. As mentioned above (Smith et al 2005) in response to the plaid stimuli constituted by two superimposed sinusoidal gratings separated by 135° , about 25% of MT neurons respond maximally when the pattern as a whole moved in the neuron’s preferred direction and therefore has a unimodal tuning curve; another 40% shows two peaks in the response tuning curve corresponding to the two-pattern directions that resulted in one of the gratings moving in the neuron’s preferred direction. The remaining 35% are unclassified (Movshon et al 2006).

1.2.4 Function

Given MT neuron's motion selectivity and disparity selectivity, many studies have explored the link between MT response and perceptual judgment. A study by (Salzman et al 1992) investigated the function of MT neurons by micro-stimulating area MT while the monkey performed a motion discrimination task and found micro-stimulation of a column of MT neurons can bias the perceptual judgment of motion direction towards the preferred direction of the neurons stimulated. Britten and his colleagues (Salzman et al 1992) further demonstrated that response of MT neurons supports behavioral choice in visual discrimination tasks. Later another study (Shadlen et al 1996) showed that trial-to-trial variability in MT response was moderately but significantly correlated with the choices that the monkey made. In other words, the monkey is more likely to make choice in favor of the preferred direction of the cell when that MT cell fires more vigorously.

In the same vein, (DeAngelis & Newsome 1999) showed that electrical stimulation of a cluster of disparity-selective MT neurons can bias perceptual judgments of depth towards the preferred disparity of the stimulated neurons. Further, Kim and his colleagues (Kim et al 2015) investigated the functional linkage between MT responses and depth judgements based on Motion Parallax (MP), and found that single neurons were generally twofold to threefold less sensitive than psychophysical judgment of the animal, with the most sensitive neurons on par with behavior, suggesting that the activity of a small number of MT neurons is sufficient to account for behavioral performance. These results indicate that MT neurons are involved in judging depth from MP, thus extending the functional roles of area MT to the perception of 3D structure during navigation.

1.3 Neuronal representation of transparent motion

1.3.1 V1 and MT response

To understand the neuronal mechanism underlying motion transparency, previous studies have focused on two cortical areas – V1 and MT.

V1 – an earlier visual processing stage than area MT, is known for its orientation-selectivity, but ~30% of V1 neurons also show direction selectivity. A previous study (Erickson 1991) recorded from direction selective V1 neurons and found V1 neurons responded well to a stimulus moving in the PD of the neuron even when another transparently moving stimulus was also present. However, another study (Qian et al., 1994b) demonstrated that V1 neuron could not distinguish transparently moving stimuli from non-transparently moving stimuli, suggesting that the neural activity in V1 is not sufficient to explain transparent motion perception. Further processing is required.

MT -- transparent motion perception is known to be closely related to the neural activity in area MT. (Newsome et al 1985) first showed that lesion of MT damages the animal's transparent motion perception. Albright's study showed that two superimposed gratings moving in different directions are perceptually seen as one coherent moving plaid pattern or as two separated moving gratings (Stoner & Albright 1996) depending on the luminance contrast of the interaction point of two gratings. Interestingly, they also found that the direction tuning of neurons in area MT reflects the coherent and non-coherent perception, suggesting a potential solution to motion transparency at the level of area MT (Stoner & Albright 1992). Later, the study by Qian et al. (1994 a, b) showed that when two random-dot moving patches are made of locally paired dots moving in opposite directions, the perception of transparent motion is lost. Their neurophysiological recordings showed that MT neuron's response in area MT, but not area V1, to unpaired dot patterns was significantly stronger than that to the paired dot patterns. This different response pattern could form the physiological basis for the perceptual difference (Figure 1.6). In addition, they re-confirmed the finding by (Snowden & Rossiter 1997). (Snowden et al 1991) that the neural response to the single dot patch moving in the preferred direction is strongly suppressed by adding the second dot patch moving in the opposite direction and the bi-directional response is well predicted by the averaging of the

two component responses. This suppression of MT neuronal responses is likely to be explained by the rise of the threshold for motion detection in transparent motion.

In addition to the opponent suppression present in the response of MT neuron to transparent motion, MT neurons are reported to show surround suppression in response to multiple visual stimuli. (Petersen et al 1985a) found that the response of MT neuron to the stimulus moving in the preferred direction within its receptive field was suppressed by the stimulus moving in the same direction but outside of the receptive field. Similar center-surround suppression also exists in representing disparity. If the surround stimulus moves at the same depth as the center's stimulus does, the neural response is maximally suppressed (Bradley et al 1995). This feature of center-surround suppression would be useful in detecting motion borders for figure-ground segmentation. (Bradley et al 1995) Bradley and Andersen (1995) also found a more salient transparent motion perception was developed by adding the depth difference, suggesting that the stereopsis can facilitate the buildup of transparent motion perception.

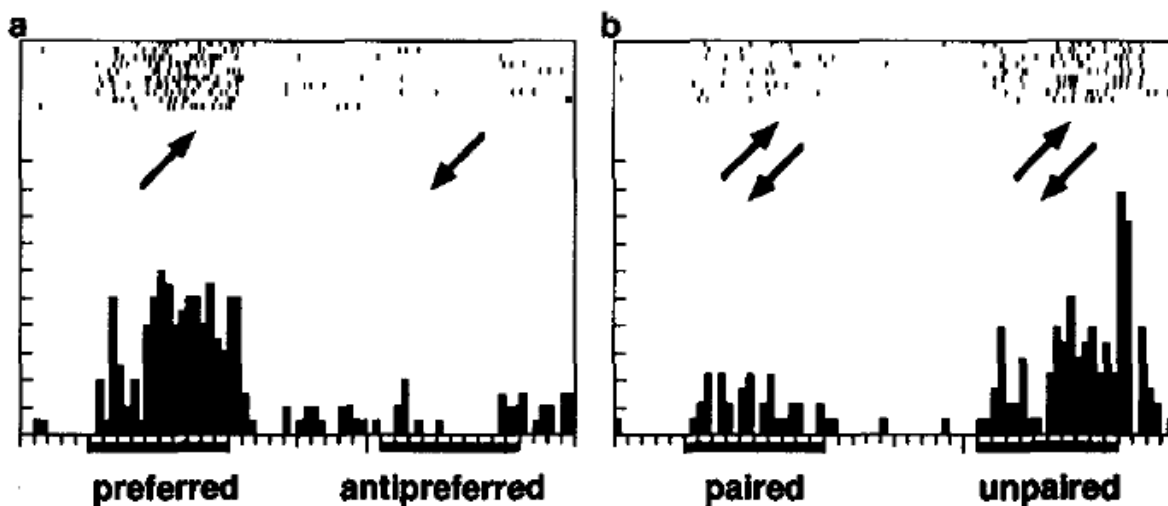


Figure 1.6: (a) An example motion selective MT neuron; (b). This cell responded significantly stronger to the unpaired dot pattern than to the paired dot pattern (from Qian & Andersen 1994).

1.3.2 Divisive Normalization model

Modeling the neural activity is of great values for uncovering the organizational principles that may underlie the responses of cortical neurons. To understand the mechanism underlying the neural representation of transparent motion, several candidate models have been proposed and examined.

First, linear weighted summation model. Given that area MT serves as an intermediate stage in the hierarchy of motion processing. The receptive fields of MT neurons are larger than those in V1 by as much as a factor of 10 and smaller than those of the neurons in MST (Lagae et al 1993, Maunsell & Van Essen 1983, Shadlen et al 1996). Spatial summation is, therefore, prevalent in the extrastriate cortex. Several studies have reported that linear averaging of the component responses evoked by individual stimulus component can well explain the neural activity evoked by multiple moving stimuli simultaneously (Recanzone et al 1997, Snowden & Rossiter 1997, Snowden et al 1991) and approximate smooth-pursuit eye movement initiated by two moving targets (Lisberger & Ferrera 1997).

Second, power-law summation model. In addition to linear weighted summation, Britten and Heuer (1999) demonstrated that power-law summation, a non-linear model, provide a significantly better prediction relative to linear summation in explaining the relationship between neural responses evoked by a pair of Gabor impulses and that evoked by single Gabor stimulus presented individually.

Third, “winner-take-all” model. This is a process whereby neurons with the strongest response to the visual stimuli determine the overall representation of the motion information. The “winner-take-all” model, essentially another distinct version of non-linear summation model, has been demonstrated to be able to well explain many perceptual consequences of visual attention (Lee et al 1999). The winner-take-all model has also been tested by (Britten & Heuer 1999) to explain the relation between neural responses to a pair of Gabors and that elicited by single Gabor, but its performance is not as good as linear averaging and power-law.

Fourth, divisive normalization model. This model was first developed to account for the non-linear property of neural activity in the primary visual cortex (Albrecht & Geisler 1991, Heeger 1992). The model calculates the ratio between the activity of an individual neuron and the summed activity of a pool of neurons.

Britten and Heuer (1999) also reported that the performance of normalization model in explaining the neural activity in area MT to a pair of Gabor stimuli is good. Their results show that stimuli that are largely ineffective at driving the cell because they lay outside the classical excitatory receptive field center are still effective at normalizing the responses to stimuli within the RF, and summation and normalization are effectively constant across a region substantially larger than the RF of the cell. Later studies further showed that divisive normalization can well account for the attention modulation at a variety of visual cortical areas, including area MT. Divisive normalization model, therefore, has been recognized to be the potential neural mechanism that determines the neural response to multiple moving stimuli. Later, a two-stage model has been proposed to explain MT response, in which a summation field that operates on the population activity of V1 serves as the first stage, then the normalization stage helps to make the selectivity of MT neuron independent of spatial pattern (Heeger et al 1996, Simoncelli & Heeger 1998). However, it remains unclear whether the normalization is computed *de novo* in MT or is simply inherited from V1.

In addition to visual system, normalization model also has been proposed to account for neural activity in the fruit fly olfactory system to odorants (Olsen et al 2010), light adaption in the retina (Boynton et al 1970), size invariance in the fly visual system and associative memory in the hippocampus (McNaughton & Davies 1987). It has also been employed to explain how MST neuron integrates the inputs from both visual modality and vestibular modality (Ohshiro et al 2011), and how responses in visual cortex are modulated by attention (Reynolds & Heeger 2009).

1.3.3 Neural substrate of Divisive Normalization

While the presence of Divisive Normalization is clear, its neuronal mechanism remains unclear. Many hypotheses have been proposed and tested.

It has been demonstrated that GABA-mediated inhibition can explain the normalization in the olfactory system of the Fruit fly (Olsen et al., 2010). However, there is still not a single biophysical and cellular mechanism that can hold across systems and species. For example, (Katzner et al 2011) report that GABA-inhibition in V1 helps enhance stimulus selectivity but is not responsible for competition among superimposed stimuli. Moreover, normalization seems to root on multiple circuits and mechanisms operating in concert and cascading across multiple stages. For instance, in the visual system, the contrast normalization is thought to be progressively strengthened from the retina, via LGN and V1 to MT (Carandini & Heeger 2012).

Regarding the feature of divisive signals, as shown above, normalization generally involves a pool of neurons rather than a single cell. Where do they come from? Feedforward network has been proposed to explain the origin of these divisive signals for the visual system of the housefly and for some aspects of normalization in the mammalian visual cortex, in which neural signals of both numerator and denominator are non-normalized themselves. In addition, feedback mechanism has been proposed to explain the normalization in primary visual cortex, where signals in the denominator have been thought to originate from lateral feedback within V1 or from feedback from higher visual areas. Therefore, further study is needed to explore the origin of the divisive signals of normalization model.

Regarding the source of division (normalization), shunting inhibition has been suggested as the biophysical mechanism of response normalization and related gain modulation because shunting inhibition acts as a divisive factor. On top of inhibition, some studies emphasized the importance of the balance of excitation and inhibition (Chance et al 2002, Shadlen & Newsome 1998), whereas others stressed the

importance of synaptic noise (Mitchell & Silver 2003, Prescott & De Koninck 2003) or short-term synaptic depression (Abbott et al 1997).

1.4 Attention modulation

1.4.1 Definition & Classification

Attention is the behavioral and cognitive process of selectively concentrating on a discrete aspect of information, whether deemed subjective or objective while ignoring other perceivable information. Research has shown that attention can enhance the subject's behavioral performance of detecting and tracking the attended information. It has been well known that visual attention improves behavioral performance in segmentation task and contrast discrimination task (Carrasco 2011, Carrasco et al 2006b, Yeshurun & Carrasco 1998). However, there has been a longstanding debate about whether segmentation guides the deployment of visual attention or visual attention helps segment visual scenes. Traditional view from studies on visual attention has suggested that visual segmentation processes arise pre-attentively and maybe facilitated by attention in the later phase for further processing (Driver et al 2001, Duncan 1984). However, the interaction between visual segmentation and selective attention remains unclear and needs to be addressed.

Attention can be classified into different classes based on how it is implemented, such as, "overt" versus "covert" attention and, "feature-based" versus "spatial" attention. Overt attention is the act of selectively attending to an item or location over others by moving eyes to point in that direction. Overt attention can be directly observed in the form of eye movements. Covert attention is the act of mentally shifting one's focus without moving one's eyes. Feature-based attention refers to the case when attention is directed to a specific feature, while spatial-based attention is directed to a specific spatial location.

1.4.2 Function

In visual cortex, it has been demonstrated that attention can modulate the neural response by increasing signal-noise ratio at both individual cell level and population level. The underlying mechanisms include the gain of contrast sensitivity, rate modulation, fano factor decrease and noise ratio decrease, contrast-dependent modulation (Reynolds & Heeger 2009).

Contrast gain. First, early studies have investigated the relationship between attention modulation and the stimulus contrast. Among them, the study by (Martinez-Trujillo & Treue 2002) found that spatial attention caused a change in contrast gain. In their study, the monkey was trained to detect a differently oriented target at cued location while the attention was directed to either a location within the RF of the recorded cell or a location in the opposite hemifield. They found that attention increased contrast sensitivity of neuronal activity. In another study by (Williford & Maunsell 2006), they showed that spatial attention had a multiplicative scaling effect on the full range of contrast tested, although the modulation gain was higher when contrast was low and lower when contrast was high.

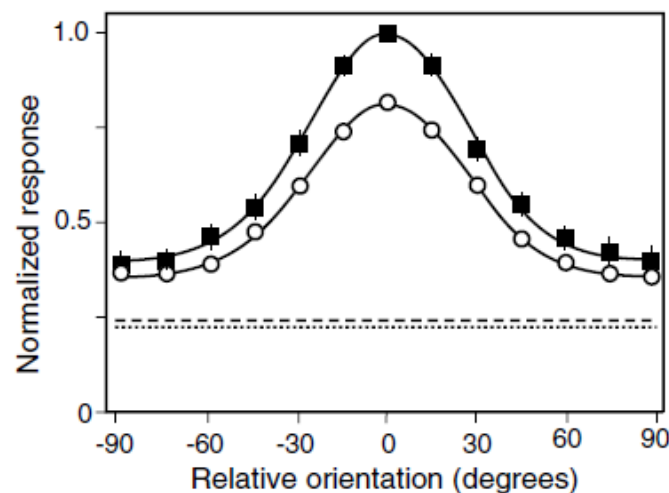


Figure 1.7, Attention increases orientation tuning curves by a gain factor. Averaged normalized orientation tuning curves computed across a population of area V4 neurons while the monkey attended either to a grating located inside the receptive field (filled squares) or to a location outside of the receptive field (empty cycles). (from Mcadams & Maunsell, 1999).

Rate modulation. Many previous studies have indicated that attention increases the neural response to stimuli moving within the receptive field, especially when the stimuli is aligned with the selective features and Neural activity is reduced when attention is directed outside of the RF (Treue & Maunsell 1996). Interestingly, it is also known that this rate modulation is stronger in area MST than that in MT, suggesting that top-down attentional modulation may increase in successive levels of cortical areas. (McAdams & Maunsell 1999) (Figure 1.7) found that attending to the receptive field causes a multiplicative increase in the orientation tuning curve of V4 neurons, without otherwise altering the shape of the tuning curve, and they further demonstrated that this increase in the gain of the orientation tuning curve enabled neuronal signals to better distinguish the orientation of the stimulus.

The decrease of Fano factor. Fano factor is computed as the ratio of spike count variance to mean spike count. Fano factor is a measure of the dispersion of the neuronal activity. Several studies have shown that attention causes a reduction in the Fano Factor of neuron response in area V4 (Cohen & Maunsell 2009, Mitchell et al 2007).

The decrease of Noise correlation. Noise correlation measures the correlation in trial-to-trial fluctuations in responses. It has been found that attention improves behavior performance primarily by reducing the noise correlations among neurons (Cohen & Kohn 2011, Cohen & Maunsell 2009, Mitchell et al 2007) (Figure 1.8b).

Other effects include the expansion of the receptive field and modulation of oscillatory neuronal synchronization. Niebergall and colleagues (2011) reported that MT responses increased when animals tracked stimuli passing across MT neurons' RFs relative to response when the same stimuli were ignored. This effect is stronger at the RF periphery than at its center, which seems like the size of RF is enlarged (Niebergall et al 2011). Fries and colleagues (2001) found that V4 neurons activated by the attended stimulus showed increased gamma-frequency (35~90 Hz) synchronization but reduced low-frequency (<17 Hz) synchronization compared with neurons activated by distractors.

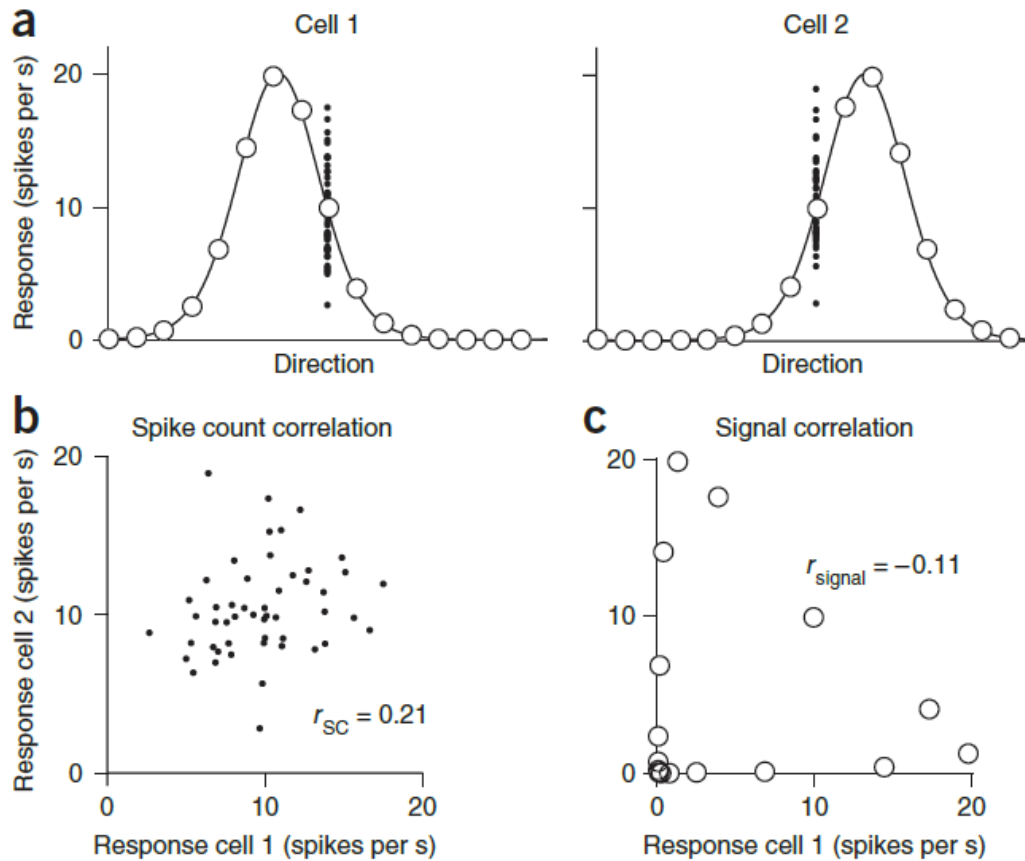


Figure 1.8, Types of pair-wise neuronal correlations. (a) Tuning curves for two hypothetical direction-selective neurons. (b) Spike count correlation, also known as noise correlation, measures the correlation between fluctuations in responses to the same stimulus. Here, each point represents the response of the two neurons on one presentation of an individual stimulus. (c) Signal correlation measures the correlation between the two cell's mean responses to different stimuli. Here each point represents the mean response to a given direction of motion. (from Cohen & Kohn 2011).

1.4.3 Attention model and Neural Correlates

To describe this wide variety of attention effects observed across many different visual areas, several models, such as biased competition model (Kastner & Ungerleider 2001) and feature similarity model (Boynton 2005, Martinez-Trujillo & Treue 2004, Treue 2001) have been proposed and demonstrated to fit the attentional modulation fairly well in many specific settings depending on task strategy and sensory

conditions. Among these models, the normalization model (Lee & Maunsell 2009, Reynolds & Heeger 2009) well accounts for all these attentional effects, in particular, it provides a mechanism that explains how attention might alter the relative influence of two or more stimuli. Later, (Ni et al 2012) introduced the tuned Normalization to explain the variability of attentional modulations across neurons with different motion selectivity.

Although the normalization model has successfully characterized attentional modulation for multiple visual areas. Many questions remain to be addressed: For example, whether the feature attention can be explained by the attentional normalization or not? How does the attention normalization model relate to the normalization of neuronal response?

With respect to these attentional modulation effects, many studies have been conducted to explore the source of visual attention. Areas like pulvinar (Sherman 2001), the frontal eye field (FEF) (Schall et al 1995) superior colliculus (Basso & Wurtz 1998) and parts of the parietal cortex (Bisley & Goldberg 2010, Colby & Goldberg 1999, Mountcastle et al 1987) have been identified as sources of the feedback signals that potentially causes the attentional modulation. For instance, to examine the role of FEF in visual attention modulation, (Moore & Fallah 2004) measured changes in contrast detection threshold following the micro-electrical stimulation of FEF neurons, and they found micro-stimulations at FEF improved the contrast sensitivity behaviorally and increased the neuronal response in area V4. Since the identification of the source for feedback signals related to attention identification is not the focus of my study, I will not expand this topic here.

1.5 Neural Coding Mechanisms

Spiking activity has been well recognized as the major way how neurons communicate information. Two major coding schemes are spiking intensity modulation (also called rate coding) and spiking timing

profile (also called temporal coding). The information coding process can be separated into two phases – encoding and decoding.

Rate coding refers to coding information based on the neural response to the external stimuli. Multiple types of neuronal response, for instance, spiking intensity (spikes per second), local field potential, have been demonstrated to carry information. Temporal coding refers to coding information through the timing profile of cortical spike trains (Richmond & Optican 1987). In my dissertation research, I focused on rate coding based on spiking intensity.

Encoding refers to the neural representation of external stimuli. In response to visual motion stimuli moving in different directions, response tuning of neurons in area MT can be estimated by a Gaussian function. Given a specific visual stimulus, the evoked neural population activity roughly follows a bell curve as well (Figure 1.9 a). The closer the stimulus motion direction is to a neuron's PD, the stronger the neuron responds. Therefore, the neuron that has a PD equals to the motion direction of the stimulus responds maximally, while neurons that have PDs far away from the stimulus direction respond weakly. The population response profile also seems like a bell curve (Figure 1.9 b).

Decoding refers to the process to recover the external sensory information from the observed neuronal responses. Several statistical methods have been developed to decode stimulus information from the neuronal response. Among them, Bayesian estimation and Maximum Likelihood estimation (MLE) have been extensively applied in decoding neuronal activity. Bayesian estimation is to estimate the posterior distribution of stimulus s given the neuronal response r , or called $P(s|r)$. (Equation 1.1), in which $P(r|s)$ is the likelihood function and $P(s)$ and $P(r)$ are called the priors over s and r . Based on the posterior probability distribution $P(s|r)$, we can predict how likely each stimulus is given the observed responses. Usually, the prediction (or called decoding result) is the estimate of the maximum posterior likelihood probability. However, Maximum Likelihood method selects the stimulus that can maximize the likelihood function $P(r|s)$ as the prediction result. Essentially, it learns (estimates) parameters to specify the joint distribution

of $P(r,s)$, such that an idealized response function (or called discriminant function) can account for the observed neuronal response. Prediction result is the one corresponding to the maximum neuronal response.

$$P(s|r) = \frac{P(r|s) * P(s)}{P(r)} \quad .(1.1)$$

$$\text{With, } P(r) = \sum_s p(r|s) * p(s)$$

In practice, besides the well-established Bayesian decoder, many other decoding algorithms have been developed, such as Linear discriminant analysis (LDA) and Support vector machine (SVM). LDA assumes a linear relationship between predictor and response after introducing a dimensionality reduction and performs one-vs-one classification to minimize the ratio of within-class variance against the across-class variance. SVM aims to maximize the hyperplane width that can optimally separate the data. SVM is robust to high dimensional data space and can apply to both linear and non-linear predictor-response data set.

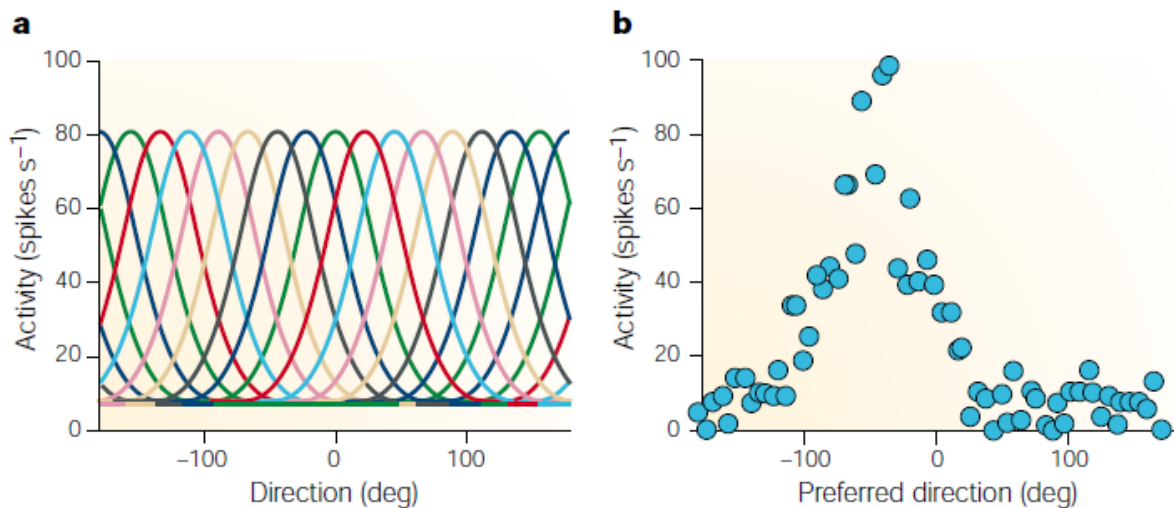


Figure 1.9, Population coding of Motion direction in are MT. (a) Direction tuning of a population of MT neurons with preferred directions span across all directions. (b) Profile of population activity elicited by a single moving stimulus. (from Pouget et al 2000).

1.6 Motivation and objectives for the thesis study

1.6.1 What is the neural representation of transparent motion across signal strengths?

To understand how our brain efficiently segments the visual objects from the background, it is critical to study the neural representation of multiple stimuli and how this internal representation gives rise to the perception of multiple objects and surfaces. Previous studies have shown that linear averaging of the neuronal responses to component stimuli can well account for the neuronal responses evoked by multiple stimuli simultaneously in area MT. However, to fully understand the interaction among component stimuli, it is necessary to study the interaction among stimuli with varying signal strength. Study at area V1 has demonstrated the divisive normalization model can well explain the neural activity to two superimposed gratings with varying luminance contrast, however, it remains unclear whether the divisive normalization model can also explain the neural representation of multiple moving stimuli with varying signal strength at area MT or not.

1.6.2 How similar motion directions are represented in area MT?

Human psychophysics study has shown that human can segment two similar motion directions (Braddick 1993). It has also been shown that neuronal response evoked by two stimuli roughly follows the average of the responses evoked by each component stimulus alone and that MT neurons usually have a broad direction tuning. In response to two stimuli moving in slightly different directions, MT neurons tend to respond maximally when the vector-averaged direction of the two component directions instead of one of the component directions moves in a cell's preferred direction. It remains unknown how to segment two similar motion directions based on neural activity in area MT. Therefore, to better understand the underlying neural mechanism for motion segmentation, it is important to investigate the representation of two slightly different motion stimuli.

1.6.3 What is the neural representation of transparent stimuli moving at different depths?

In motion transparency, one surface is often perceived as moving on top of the other in spite of no depth cue in the display. MT neurons are also known for having selectivity for both motion direction and binocular disparity. Then what is the neuronal representation of transparent motion at different depths? The depth perception could be part of the reason for the buildup and maintenance of motion segmentation, and it could also be the consequence of motion segmentation. Early studies have shown that binocular disparity facilitates the detection of transparent motion in a variety of stimuli settings, such as random-dot stimuli, grating-plaid and motion after-effects. However, the mechanisms underlying this facilitation remain unclear. Directing attention to a stimulus moving at a certain depth is known for assisting motion segmentation (Greenwood & Edwards 2006). Is the facilitatory effect for detecting transparent motion by adding disparity cue the result of an attention effect or low-level neural processing independent of attention may also contribute to it? In this project, I would investigate the neural representation of multiple stimuli moving at different depth and the effects of feature-based attention.

References

- Abbott LF, Varela JA, Sen K, Nelson SB.** 1997. Synaptic depression and cortical gain control. *Science* 275: 220-4
- Abeles M, Goldstein MH, Jr.** 1970. Functional architecture in cat primary auditory cortex: columnar organization and organization according to depth. *J Neurophysiol* 33: 172-87
- Albrecht DG, Geisler WS.** 1991. Motion Selectivity and the Contrast-Response Function of Simple Cells in the Visual-Cortex. *Visual Neurosci* 7: 531-46
- Albright TD.** 1984. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol* 52: 1106-30

- Anderson JC, Martin KA.** 2002. Connection from cortical area V2 to MT in macaque monkey. *J Comp Neurol* 443: 56-70
- Baker JF, Petersen SE, Newsome WT, Allman JM.** 1981. Visual response properties of neurons in four extrastriate visual areas of the owl monkey (*Aotus trivirgatus*): a quantitative comparison of medial, dorsomedial, dorsolateral, and middle temporal areas. *J Neurophysiol* 45: 397-416
- Basso MA, Wurtz RH.** 1998. Modulation of neuronal activity in superior colliculus by changes in target probability. *J Neurosci* 18: 7519-34
- Beckers G, Zeki S.** 1995. The Consequences of Inactivating Areas V1 and V5 on Visual-Motion Perception. *Brain* 118: 49-60
- Berman RA, Wurtz RH.** 2010. Functional identification of a pulvinar path from superior colliculus to cortical area MT. *J Neurosci* 30: 6342-54
- Bisley JW, Goldberg ME.** 2010. Attention, Intention, and Priority in the Parietal Lobe. *Annual Review of Neuroscience*, Vol 33 33: 1-21
- Born RT.** 2001. Visual processing: parallel-er and parallel-er. *Curr Biol* 11: R566-8
- Born RT, Bradley DC.** 2005. Structure and function of visual area MT. *Annu Rev Neurosci* 28: 157-89
- Boynton GM.** 2005. Attention and visual perception. *Curr Opin Neurobiol* 15: 465-9
- Boynton JE, Gillham NW, Burkholder B.** 1970. Mutations Altering Chloroplast Ribosome Phenotype in *Chlamydomonas*, II. A New Mendelian Mutation. *Proc Natl Acad Sci U S A* 67: 1505-12
- Braddick O.** 1993. Segmentation versus integration in visual motion processing. *Trends Neurosci* 16: 263-8
- Bradley DC, Qian N, Andersen RA.** 1995. Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature* 373: 609-11
- Britten KH, Heuer HW.** 1999. Spatial summation in the receptive fields of MT neurons. *J Neurosci* 19: 5074-84
- Buschman TJ, Miller EK.** 2007. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315: 1860-2
- Carandini M, Heeger DJ.** 2012. Normalization as a canonical neural computation. *Nature Reviews Neuroscience* 13: 51-62
- Carrasco M.** 2011. Visual attention: the past 25 years. *Vision Res* 51: 1484-525
- Carrasco M, Loula F, Ho YX.** 2006. How attention enhances spatial resolution: Evidence from selective adaptation to spatial frequency. *Percept Psychophys* 68: 1004-12
- Chance FS, Abbott LF, Reyes AD.** 2002. Gain modulation from background synaptic input. *Neuron* 35: 773-82
- Cohen MR, Kohn A.** 2011. Measuring and interpreting neuronal correlations. *Nature Neuroscience* 14: 811-19
- Cohen MR, Maunsell JH.** 2009. Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12: 1594-600
- Colby CL, Goldberg ME.** 1999. Space and attention in parietal cortex. *Annu Rev Neurosci* 22: 319-49
- DeAngelis GC, Newsome WT.** 1999. Organization of disparity-selective neurons in macaque area MT. *Journal of Neuroscience* 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
- Driver J, Davis G, Russell C, Turatto M, Freeman E.** 2001. Segmentation, attention and phenomenal visual objects. *Cognition* 80: 61-95
- Dubner R, Zeki SM.** 1971. Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Res* 35: 528-32
- Duncan J.** 1984. Selective Attention and the Organization of Visual Information. *Journal of Experimental Psychology-General* 113: 501-17

- Erickson GM.** 1991. The Cambridge Ancient-History - the Rise of Rome to 220 Bc, Vol 7, Pt 2, 2nd Edition, the Cambridge Ancient-History - Rome and the Mediterranean to 133 Bc, Vol 8, 2nd Edition - Walbank,Fw, Astin,Ae, Frederiksen,Mw, Ogilvie,Rm. *Mod Lang J* 75: 515-16
- Felleman DJ, Kaas JH.** 1984. Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. *J Neurophysiol* 52: 488-513
- Fries W.** 1981. The projection from the lateral geniculate nucleus to the prestriate cortex of the macaque monkey. *Proc R Soc Lond B Biol Sci* 213: 73-86
- Girard P, Salin PA, Bullier J.** 1992. Response selectivity of neurons in area MT of the macaque monkey during reversible inactivation of area V1. *J Neurophysiol* 67: 1437-46
- Greenwood JA, Edwards M.** 2006. Pushing the limits of transparent-motion detection with binocular disparity. *Vision Res* 46: 2615-24
- Heeger DJ.** 1992. Normalization of Cell Responses in Cat Striate Cortex. *Visual Neurosci* 9: 181-97
- Heeger DJ, Simoncelli EP, Movshon JA.** 1996. Computational models of cortical visual processing. *P Natl Acad Sci USA* 93: 623-27
- Hubel DH, Livingstone MS.** 1987. Segregation of form, color, and stereopsis in primate area 18. *J Neurosci* 7: 3378-415
- Hubel DH, Wiesel TN.** 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160: 106-54
- Hubel DH, Wiesel TN.** 1977. Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc R Soc Lond B Biol Sci* 198: 1-59
- Hubener M, Shoham D, Grinvald A, Bonhoeffer T.** 1997. Spatial relationships among three columnar systems in cat area 17. *J Neurosci* 17: 9270-84
- Kastner S, Ungerleider LG.** 2001. The neural basis of biased competition in human visual cortex. *Neuropsychologia* 39: 1263-76
- Katzner S, Busse L, Carandini M.** 2011. GABAA inhibition controls response gain in visual cortex. *J Neurosci* 31: 5931-41
- 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
- Lagae L, Raiguel S, Orban GA.** 1993. Speed and direction selectivity of macaque middle temporal neurons. *J Neurophysiol* 69: 19-39
- Lee DK, Itti L, Koch C, Braun J.** 1999. Attention activates winner-take-all competition among visual filters. *Nature Neuroscience* 2: 375-81
- Lee J, Maunsell JH.** 2009. A normalization model of attentional modulation of single unit responses. *PLoS One* 4: e4651
- Lisberger SG, Ferrera VP.** 1997. Vector averaging for smooth pursuit eye movements initiated by two moving targets in monkeys. *J Neurosci* 17: 7490-502
- Lovejoy LP, Krauzlis RJ.** 2010. Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. *Nat Neurosci* 13: 261-6
- Martinez-Trujillo JC, Treue S.** 2002. Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron* 35: 365-70
- Martinez-Trujillo JC, Treue S.** 2004. Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* 14: 744-51
- Maunsell JH, Van Essen DC.** 1983. 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
- McAdams CJ, Maunsell JH.** 1999. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 19: 431-41
- McNaughton L, Davies P.** 1987. The effects of a 16 week aerobic conditioning program on serum lipids, lipoproteins and coronary risk factors. *J Sports Med Phys Fitness* 27: 296-302

- Mitchell JF, Sundberg KA, Reynolds JH.** 2007. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55: 131-41
- Mitchell SJ, Silver RA.** 2003. Shunting inhibition modulates neuronal gain during synaptic excitation. *Neuron* 38: 433-45
- Moore T, Fallah M.** 2004. Microstimulation of the frontal eye field and its effects on covert spatial attention. *J Neurophysiol* 91: 152-62
- Mountcastle VB.** 1997. The columnar organization of the neocortex. *Brain* 120 (Pt 4): 701-22
- Mountcastle VB, Motter BC, Steinmetz MA, Sestokas AK.** 1987. Common and differential effects of attentive fixation on the excitability of parietal and prestriate (V4) cortical visual neurons in the macaque monkey. *J Neurosci* 7: 2239-55
- Movshon JA, Rust NC, Mante V, Simoncelli EP.** 2006. How MT cells analyse the motion of visual patterns. *Perception* 35: 12-13
- Newsome WT, Wurtz RH, Dursteler MR, Mikami A.** 1985. Deficits in Visual-Motion Processing Following Ibotenic Acid Lesions of the Middle Temporal Visual Area of the Macaque Monkey. *Journal of Neuroscience* 5: 825-40
- Ni AM, Ray S, Maunsell JH.** 2012. Tuned normalization explains the size of attention modulations. *Neuron* 73: 803-13
- Niebergall R, Khayat PS, Treue S, Martinez-Trujillo JC.** 2011. Expansion of MT neurons excitatory receptive fields during covert attentive tracking. *J Neurosci* 31: 15499-510
- Ninomiya T, Sawamura H, Inoue K, Takada M.** 2012. Segregated pathways carrying frontally derived top-down signals to visual areas MT and V4 in macaques. *J Neurosci* 32: 6851-8
- Ohshiro T, Angelaki DE, DeAngelis GC.** 2011. A normalization model of multisensory integration. *Nat Neurosci* 14: 775-82
- Olsen SR, Bhandawat V, Wilson RI.** 2010. Divisive normalization in olfactory population codes. *Neuron* 66: 287-99
- Peterhans E, Vonderheydt R.** 1993. Functional-Organization of Area V2 in the Alert Macaque. *Eur J Neurosci* 5: 509-24
- Petersen SE, Baker JF, Allman JM.** 1985a. Direction-Specific Adaptation in Area Mt of the Owl Monkey. *Brain Res* 346: 146-50
- Petersen SE, Robinson DL, Keys W.** 1985b. Pulvinar Nuclei of the Behaving Rhesus-Monkey - Visual Responses and Their Modulation. *Journal of Neurophysiology* 54: 867-86
- Pouget A, Dayan P, Zemel R.** 2000. Information processing with population codes. *Nat Rev Neurosci* 1: 125-32
- Prescott SA, De Koninck Y.** 2003. Gain control of firing rate by shunting inhibition: Roles of synaptic noise and dendritic saturation. *P Natl Acad Sci USA* 100: 2076-81
- Priebe NJ, Lisberger SG.** 2004. Estimating target speed from the population response in visual area MT. *J Neurosci* 24: 1907-16
- Prince SJ, Cumming BG, Parker AJ.** 2002a. Range and mechanism of encoding of horizontal disparity in macaque V1. *J Neurophysiol* 87: 209-21
- Prince SJ, Pointon AD, Cumming BG, Parker AJ.** 2002b. Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. *J Neurophysiol* 87: 191-208
- Qian N, Andersen RA.** 1994. Transparent Motion Perception as Detection of Unbalanced Motion Signals .2. Physiology. *Journal of Neuroscience* 14: 7367-80
- Recanzone GH, Wurtz RH, Schwarz U.** 1997. Responses of MT and MST neurons to one and two moving objects in the receptive field. *J Neurophysiol* 78: 2904-15
- Reynolds JH, Heeger DJ.** 2009. The Normalization Model of Attention. *Neuron* 61: 168-85
- Richert M, Albright TD, Krekelberg B.** 2013. The complex structure of receptive fields in the middle temporal area. *Front Syst Neurosci* 7: 2

- Richmond BJ, Optican LM.** 1987. Temporal Encoding of Two-Dimensional Patterns by Single Units in Primate Inferior Temporal Cortex .2. Quantification of Response Wave-Form. *Journal of Neurophysiology* 57: 147-61
- Rockland KS.** 1995. Morphology of Individual Axons Projecting from Area V2 to Mt in the Macaque. *Journal of Comparative Neurology* 355: 15-26
- Salzman CD, Murasugi CM, Britten KH, Newsome WT.** 1992. Microstimulation in Visual Area Mt - Effects on Direction Discrimination Performance. *Journal of Neuroscience* 12: 2331-55
- Schall JD, Hanes DP, Thompson KG, King DJ.** 1995. Saccade Target Selection in Frontal Eye Field of Macaque .1. Visual and Premovement Activation. *Journal of Neuroscience* 15: 6905-18
- Schlack A, Albright TD.** 2007. Remembering visual motion: Neural correlates of associative plasticity and motion recall in cortical area MT. *Neuron* 53: 881-90
- Shadlen MN, Britten KH, Newsome WT, Movshon JA.** 1996. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci* 16: 1486-510
- Shadlen MN, Newsome WT.** 1996. Motion perception: seeing and deciding. *Proc Natl Acad Sci U S A* 93: 628-33
- Shadlen MN, Newsome WT.** 1998. The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience* 18: 3870-96
- Sherman SM.** 2001. Thalamic relay functions. *Vision: From Neurons to Cognition* 134: 51-69
- Shipp S, Zeki S.** 1989. The Organization of Connections between Areas V5 and V2 in Macaque Monkey Visual-Cortex. *Eur J Neurosci* 1: 333-54
- Simoncelli EP, Heeger DJ.** 1998. A model of neuronal responses in visual area MT. *Vision Res* 38: 743-61
- Sincich LC, Horton JC.** 2003. Independent projection streams from macaque striate cortex to the second visual area and middle temporal area. *Journal of Neuroscience* 23: 5684-92
- Smith MA, Majaj NJ, Movshon JA.** 2005. Dynamics of motion signaling by neurons in macaque area MT. *Nature Neuroscience* 8: 220-28
- Snowden RJ, Rossiter M.** 1997. Stereoscopic depth cues can segment motion information. *Perception* 26: 1337-37
- Snowden RJ, Treue S, Erickson RG, Andersen RA.** 1991. The Response of Area Mt and V1 Neurons to Transparent Motion. *Journal of Neuroscience* 11: 2768-85
- Sretavan D, Dykes RW.** 1983. The organization of two cutaneous submodalities in the forearm region of area 3b of cat somatosensory cortex. *J Comp Neurol* 213: 381-98
- Standage GP, Benevento LA.** 1983. The Organization of Connections between the Pulvinar and Visual Area Mt in the Macaque Monkey. *Brain Res* 262: 288-94
- Stoner GR, Albright TD.** 1992. Neural Correlates of Perceptual Motion Coherence. *Nature* 358: 412-14
- Stoner GR, Albright TD.** 1993. Image segmentation cues in motion processing: implications for modularity in vision. *J Cogn Neurosci* 5: 129-49
- Stoner GR, Albright TD.** 1996. The interpretation of visual motion: evidence for surface segmentation mechanisms. *Vision Res* 36: 1291-310
- Stoner GR, Albright TD, Ramachandran VS.** 1990. Transparency and Coherence in Human Motion Perception. *Nature* 344: 153-55
- 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.** 2001. Neural correlates of attention in primate visual cortex. *Trends Neurosci* 24: 295-300
- Treue S, Maunsell JHR.** 1996. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382: 539-41
- Williford T, Maunsell JH.** 2006. Effects of spatial attention on contrast response functions in macaque area V4. *J Neurophysiol* 96: 40-54

- Yabuta NH, Sawatari A, Callaway EM.** 2001. Two functional channels from primary visual cortex to dorsal visual cortical areas. *Science* 292: 297-300
- Yeshurun Y, Carrasco M.** 1998. Attention improves or impairs visual performance by enhancing spatial resolution. *Nature* 396: 72-5
- Zeki S.** 1983. Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience* 9: 741-65

Chapter 2

Normalization of neuronal responses in cortical area MT across signal strengths and motion directions

This chapter has been published as:

Xiao J, Niu YQ, Wiesner S, and Huang X. Normalization of neuronal responses in cortical area MT across signal strengths and motion directions. *J Neurophysiol* 112: 1291-1306, 2014.

Abstract

Multiple visual stimuli are common in natural scenes, yet it remains unclear how multiple stimuli interact to influence neuronal responses. We investigated this question by manipulating relative signal strengths of two stimuli moving simultaneously within the receptive fields (RFs) of neurons in the extrastriate middle temporal (MT) cortex. Visual stimuli were overlapping random-dot patterns moving in two directions separated by 90°. We first varied the motion coherence of each random-dot pattern and characterized, across the direction tuning curve, the relationship between neuronal responses elicited by bi-directional stimuli and by the constituent motion components. The tuning curve for bi-directional stimuli showed response normalization and can be accounted for by a weighted sum of the responses to motion components. Allowing nonlinear, multiplicative interaction between two component responses significantly improved the data fit for some neurons and the interaction mainly had a suppressive effect on the neuronal response. The weighting of the component response was not fixed, but dependent on relative

signal strengths. When two stimulus components moved at different coherence levels, the response weight for the higher-coherence component was significantly greater than that for the lower-coherence component. We also varied relative luminance levels of two coherently moving stimuli and found that MT response weight for the higher-luminance component was also greater. These results suggest that competition between multiple stimuli within a neuron's RF depends on relative signal strengths of the stimuli, and multiplicative nonlinearity may play an important role in shaping the response tuning for multiple stimuli.

2.1 Introduction

Natural visual scenes are normally constituted by multiple visual stimuli. The buildup of an accurate perception of the environment around us relies on the ability to segment multiple visual stimuli and extract information from each correctly. However, it remains unclear how multiple stimuli interact to influence neuronal responses. For neurons in extrastriate cortex that have larger RFs than those in the primary visual cortex, this question is even more prominent because larger RFs are more likely to encompass multiple stimuli. In this study we investigate how multiple, overlapping moving stimuli interact within the RFs of neurons in area MT.

It has been well known that cortical area MT plays an important role in visual motion perception (Born and Bradley 2005) and in providing motion signals for guiding the initiation of smooth pursuit of eye movements (Lisberger 2010). Most neurons in area MT are motion direction selective (Albright 1984; Maunsell and Van Essen 1983). Previous studies have indicated that response of MT neurons to two separate stimuli moving in different directions within the receptive fields (RFs) roughly follows the average of the response evoked by individual stimulus presented alone (Qian and Andersen 1994; Recanzone et al. 1997; Snowden et al. 1991; but see Krekelberg and van Wezel 2013). These results can be explained by the divisive normalization model in which each neuron computes a linear weighted sum of its inputs, divided

by the summed activity of a pool of neurons excited by the visual stimuli (Heeger et al. 1996; Simoncelli and Heeger 1998). Using two stimuli moving in the same direction at different locations, Britten and Heuer (1999) have further shown that spatial summation within RFs of MT neurons follows a scaled version of weighted sum and divisive normalization. In these studies, simultaneously presented stimuli had the same signal strength.

To understand the interaction of multiple stimuli, it is important to study the interaction between stimuli that have different visual signal strength. MT neurons may follow a linear averaging scheme and equally weight the responses elicited by component stimuli that have different signal strengths (Qian and Andersen 1994; Recanzone et al. 1997; Snowden et al. 1991). Alternatively, MT neurons may weigh the stimulus components that have a higher signal strength more strongly. Heuer and Britten (2002) have shown that response normalization in area MT depends on the luminance contrast of visual stimuli: normalization is stronger when two Gabor stimuli presented in the RF both have a high contrast, and normalization is weaker when at least one of the two Gabor stimuli has a low contrast. These findings suggest that response weighting in MT can vary depending on the luminance contrast. However, how the response weighting in MT is governed by relative signal strengths remains unclear.

Understanding the interactions between stimuli with different signal strengths is pivotal to uncovering the mechanisms on how multiple stimuli interact with each other. MT neurons may equally weight the responses elicited by component stimuli with different signal strengths or weight the component stimulus with stronger signal strength more strongly. Heuer and Britten (2002) have shown that response normalization in area MT depends on the luminance contrast of visual stimuli: normalization is stronger when two Gabor stimuli presented in the RF both have a high contrast, and normalization is weaker when at least one of the two Gabor has a low contrast. These findings suggest that response weighting in MT can vary depending on the luminance contrast. However, how the response weighting in MT is governed by signal strengths is still unclear.

In primary visual cortex, neuronal response elicited by overlapping gratings that have different orientations can be accounted for by a weighted sum of the response to individual gratings; moreover, the response weight is greater for the stimulus component that has a higher contrast (Busse et al. 2009; MacEvoy et al. 2009). These results can be explained by a contrast normalization model, in which the weight for the response to one grating is determined by the grating's contrast, normalized by the overall contrast of all gratings (Busse et al. 2009; Carandini et al. 1997). It remains unknown whether the changes of response weights based on relative luminance contrast also occur in the extrastriate cortex, and it remains unclear whether the scheme of weighting component responses based on relative signal strengths applies to other visual attributes.

In visual motion, motion coherence is another feature to manipulate the signal strength. Usually, the coherence of the random-dot pattern stimulus is controlled by the proportion of dots that move in the same direction at the same speed. As the coherence increases, the motion signal becomes stronger and the direction of the motion is more salient. Neurons in area MT are reported to be sensitive to motion coherence (Britten et al. 1992, 1993), so in order to investigate the rule governing the interaction/competition between multiple stimuli that have the same or different signal strengths, we manipulated the relative coherence levels of two overlapping, random-dot stimuli moving in different directions. We set out to test the hypothesis that stimulus competition depends on relative motion coherence and favors the stimulus component that has the higher motion coherence. To test the generality of the hypothesis, we also varied signal strength by manipulating relative luminance levels of moving stimuli.

We found that 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. In addition to the weighted linear summation, we also found evidence suggesting nonlinear interaction tends to have a suppressive effect on the neuronal response. Our results provide important constraint on neural

models of encoding multiple visual stimuli and suggest a general rule of competition between multiple stimuli within neurons' RFs.

2.2 Materials and Methods

Three adult male rhesus monkeys (*Macaca mulatta*) were used in the neurophysiological experiments. Experimental protocols were approved by local *IACUCs* and followed the NIH *Guide for the Care and Use of Laboratory Animals*. Procedures for surgical preparation and electrophysiological recording were routine and similar to those described previously (Ramachandran and Lisberger, 2005; Huang and Lisberger, 2013). During sterile surgery with the animal under isoflurane anesthesia, a head post and a recording cylinder were implanted to allow recording from neurons in cortical area MT. Eye position was monitored at 1000 Hz using a video-based eye tracker (EyeLink, SR Research) for monkeys GE and BJ, and the search coil method (Judge et al. 1980) for monkey RG.

For electrophysiological recordings from neurons in area MT, we lowered tungsten electrodes (1~3M Ω , FHC) into the posterior bank of the superior temporal sulcus. 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), and relatively low preferred speed (typically less than 40 deg/s) compared to that of MST neurons (Nover et al., 2005; Churchland et al. 2007). Electrical signals were amplified and single units were identified with a real-time template matching system and an offline spike sorter (Plexon).

Visual stimuli and behavioral paradigm

Stimulus presentation, the behavioral paradigm, and data acquisition were controlled by a real-time data acquisition program (<https://sites.google.com/a/srscicomp.com/maestro/>). Visual stimuli were

presented on a 25" CRT monitor at a viewing distance of 63 cm. Monitor resolution was 1024×768 pixels and the refresh rate was 100 Hz. Visual stimuli were generated by a Linux workstation using an OpenGL application that communicated with an experimental-control computer. The output of the video monitor was measured with a photometer (LS-110, Minolta) and was gamma corrected.

Visual stimuli were random-dot patterns, presented within a stationary, circular aperture that was 7.5° across. Each dot was a square of 2 pixels (0.08°) on a side. The dot density of a single random-dot pattern was 3.4 dots/deg². In the first two experiments investigating the effects of motion coherence, the luminance levels of the dots and the background were 15.3 and 1.9 cd/m², respectively. In the third experiment investigating the effects of the stimulus luminance, the luminance of the dots in each random-dot pattern was one of three values of 2.5, 10, and 40 cd/m² and the background was 0.2 cd/m². These three luminance values gave rise to the root mean square (RMS) contrasts (Peli, 1990; Kukkonen et al., 1992) of 0.34, 1.43, and 5.8 cd/m², respectively.

To generate a random-dot pattern moving at N% of motion coherence (after Newsome and Pare, 1988; Britten et al., 1992), N% of the dots were selected to move coherently, while the rest of the dots were repositioned randomly within the outer boundary of the stimulus. Random selections of signal and noise dots occurred at each monitor frame. Therefore, a given dot would switch back and forth between a signal dot and a noise dot as it travelled across the circular aperture. The life time of each dot was as long as the motion duration. All visual stimuli were presented in individual *trials* while the monkeys fixated within a 1.5°×1.5° window of a spot of light to receive juice rewards. Visual stimuli were illuminated after the animal fixated for 200 ms. To separate the neuronal response to stimulus motion from that of the stimulus onset, visual stimuli remained stationary on the display for 200 ms before starting to move. In the first experiment, visual stimuli moved for 1000 ms. Because we found essentially the same results when we analyzed the neuronal responses during the whole motion period and during the initial 500-ms or 300-ms motion period, in the second and third experiments, we used shorter motion durations of 500 ms and 300

ms, respectively, to speed up data collection. The monkeys continued to fixate for 200 ms after the visual stimuli were turned off.

Experimental design

We first characterized the direction selectivity of an isolated neuron by randomly interleaving trials of $30 \times 27^\circ$ random-dot patches moving at $10^\circ/\text{s}$ in eight different directions at 45° steps. Directional tuning was evaluated on-line using matlab (MathWorks). We next characterized the speed tuning by interleaving $30 \times 27^\circ$ random-dot patches moving in the preferred direction (PD) of the neuron at one of eight different speeds, ranging from 1 to $128^\circ/\text{s}$, evenly spanning a log scale of speed. Next, we mapped the RF of the neuron by recording responses to a series of $5 \times 5^\circ$ patches of random dots that moved in the PD and at the preferred speed (PS) of the neuron. The location of the patch was varied randomly to tile the screen in 5° steps without overlap and to cover an area of $35 \times 25^\circ$. The raw map of the RF was interpolated at an interval of 0.5° and the location giving rise to the highest firing rate was taken as the center of the RF. In the following experiments, visual stimuli were centered on the RF.

In the first experiment, we set the levels of motion coherence of two superimposed random-dot patterns both at 100%, or one at 100% and the other at 60%. The direction separation between the two random-dot patterns, referred to as two stimulus components, was fixed at 90° . Note that, when a random-dot pattern moved at 60% coherence in a given direction, the visual stimuli were drastically different from a situation if 60% of the dots moved at 100% coherence and 40% of dots moved randomly. Because the random selections of signal and noise dots occurred at each monitor frame, a noise dot at one frame may turn into a signal dot in the next frame and moved in the direction of the random-dot pattern. As a result, the noise dots did not appear to be separable from the random-dot pattern. When a random-dot pattern of 60% coherence was superimposed with a random-dot pattern of 100% coherence, the frame-to-frame noise dots of the 60% coherence pattern did not appear to interfere with the random-dot pattern of 100%

coherence. Trials with different vector-averaged (VA) directions of the two stimulus components were randomly interleaved to characterize the response *tuning* to the bi-directional stimuli. The VA direction was typically varied at a step of 15° . Trials that contained unidirectional component stimuli presented alone were interleaved with the bi-directional stimuli.

In the second experiment, we used visual stimuli that contained two superimposed random-dot patterns, one moving in the PD of the neuron and the other moving in a direction that was 90° away from the PD. The motion coherence of the random-dot pattern moving in the PD (referred to as the PD component) was set at one of 5 levels, from 60 to 100% at 10% steps. The motion coherence of the orthogonal motion component was set at 100%. Trials of different levels of motion coherence were randomly interleaved. Also randomly interleaved were trials that presented the PD or the orthogonal motion component alone.

In the third experiment, we set the luminance of the stimulus component moving at the clockwise side of two component directions at 40 cd/m^2 and the luminance of the other stimulus component at either 10 or 2.5 cd/m^2 . We varied the VA direction of the bi-directional stimuli to characterize direction tuning curves. In all three experiments, the speeds of the two stimulus components were equal and were set within the range from 1.5 to $20^\circ/\text{s}$ and closest to the preferred speeds of the recorded neurons. The median stimulus speed across experiments was $10^\circ/\text{s}$.

Data analysis

Response firing rate was calculated during the time interval of stimulus motion and averaged across repeated trials. We next fitted raw direction tuning curves using splines at a resolution of 1° , and rotated the spline-fitted tuning curve elicited by the bi-directional stimuli to align the VA direction of 0° with the PD of each neuron. We normalized each neuron's responses by the neuron's maximum bi-directional response and averaged the aligned, normalized tuning curves across neurons.

To determine the impact of relative coherence levels of motion components on the response tuning curve, we computed the *center of gravity* (CG) of the aligned tuning curve:

$$CG(R) = \frac{\sum [R(\theta) \cdot \vec{\theta}]}{\sum R(\theta)}, \quad (1)$$

where θ is the motion direction and $R(\theta)$ is the neuronal response to that direction. We calculated a *center of gravity shift index* (CGSI) to quantify the CG of bi-directional responses in relation to the CGs of corresponding component responses:

$$CGSI(x) = \frac{CG(R_x) - CG(R_{LC})}{CG(R_{HC}) - CG(R_{LC})}, \quad (2)$$

where R_x refers to either the response tuning to the bi-directional stimuli, or the averaged response tuning of the component responses; R_{HC} and R_{LC} refer to the response tuning curves elicited by the higher- and lower-coherence component, respectively. We took into account the circular nature of motion direction when calculating the difference between CGs in Eq. 2.

Model fits of bi-directional responses and evaluation of goodness-of-fit

We fitted neuronal responses elicited by the bi-directional stimuli across the VA directions (in Experiments 1 and 3) or multiple levels of motion coherence (in Experiment 2) using four different models (see Results), by minimizing the sum of squared error. To evaluate the goodness-of-fit of a model, we computed the percentage of variance (PV) accounted for by each model fit:

$$PV = 100 \times \left(1 - \frac{SSE}{SST}\right), \quad (3)$$

where SSE is the sum of squared errors between a model fit and the data, and SST is the sum of squared differences between the data and the mean of data (cf. Morgan et al., 2008). When occasionally SSE exceeded SST and gave rise to a negative PV, we forced the PV to be zero.

To evaluate whether adding a nonlinear interaction term to a linear weighted sum model significantly improved the goodness-of-fit, we used sequential F tests (Draper and Smith, 1998). The F ratio was calculated by:

$$F(P_{NL} - P_L, N - P_{NL}) = \frac{(SSR_L - SSR_{NL}) / (P_{NL} - P_L)}{SSR_{NL} / (N - P_{NL})}, \quad (4)$$

where SSR_{NL} is the residual sum of squares from the fit using the nonlinear model that has P_{NL} parameters, SSR_L is the residual from the fit using the linear model that has P_L parameters, and N is the total number of data points. This test takes into consideration the difference in the number of parameters of the to-be-compared models.

2.3 Results

To determine how multiple stimuli interact within the RFs of MT neurons, we investigated the relationship between the responses evoked by two overlapping random dot stimuli moving in different directions and those evoked by the constituent stimulus components. In the first two experiments, we varied relative levels of motion coherence of the stimulus components. In the third experiments, we varied relative luminance contrasts of the stimulus components. We collected 224 single unit neurons in area MT (128 from monkey GE, 88 from monkey BJ, and 8 from monkey RG). Among them, 141, 66 and 38 neurons were tested in the first, second and third experiment, respectively.

Direction tuning of neuronal responses to bidirectional stimuli moving at the same or different levels of motion coherence.

We first determined how direction tuning curves of MT neurons in response to the bidirectional stimuli varied with relative levels of motion coherence. The direction separation between the two motion components was 90° . To characterize the response tuning curve, we varied the Victor-averaged (VA) direction of the two motion components either both at 100% (referred to as the “equal-coherence condition”) or one at 100% and the other at 60% (referred to as the “different-coherence condition”) (illustrated in Fig.1, *A* and *B*).

Figure 1 shows the results from two example neurons. To visualize the relationship between the responses to the bidirectional stimuli and those evoked by the unidirectional stimulus components, we superimposed the response tuning curves evoked by the bidirectional stimuli and by the stimulus components. The blue and green curves show the direction tuning curves evoked by the stimulus components presented alone (Fig.1). These two tuning curves were shifted by 90° relative to each other (note the color-coded component directions in the axis of abscissas). A vertical line at a given VA direction of the bidirectional stimuli intersects with the blue and green elicited by the two component directions of the bidirectional stimuli (see the vertical dashed line in Fig. 1C).

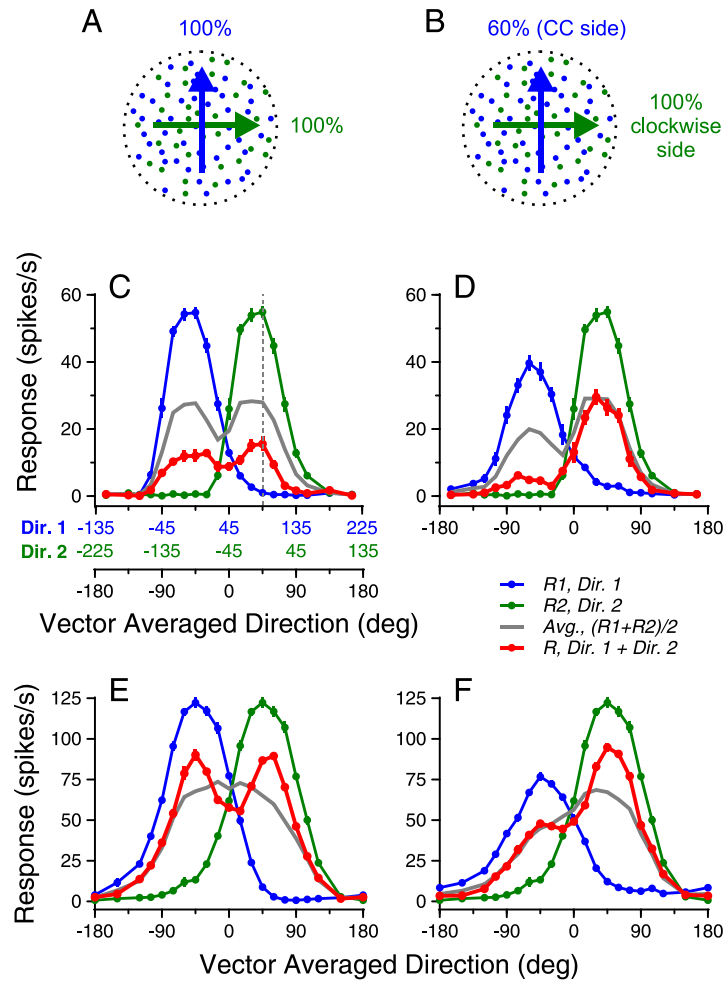


Figure 1, Visual stimuli and direction tuning curves of 2 example neurons elicited by bidirectional stimuli and the constituent unidirectional components. **A** and **B**: cartoon illustration of the visual stimuli. Blue and green dots indicate 2 achromatic random-dot patterns. The stimulus component shown in blue (direction 1, *Dir 1*) moved at the counterclockwise (CC) side of the 2 component directions. The stimulus component shown in green (direction 2, *Dir 2*) moved at the clockwise side of the 2 component direction. **A**: 2 motion component both moved at 100% coherence. **B**: the clockwise-side component moved at 100%, whereas the CC-side component moved at 60% coherence. **C** and **D**: response from 1 example neuron. **E** and **F**: responses from another example neuron. The abscissa in black refers to the vector-averaged (VA) direction of the bidirectional stimuli. The abscissas in blue and green (**C**) refer to *Dir.1* and 2, respectively, of the unidirectional stimulus components. The blue and green axes are shifted 90° relative to each other. VA direction of 0° is aligned with the preferred direction (PD) of the neuron. The dotted vertical line intersects with the bidirectional response at VA direction of 45° and the 2 component response at direction of 90° (*Dir.1*) and 0° (*Dir. 2*), respectively. Error bars indicate SE. See text for definitions.

When both stimulus components moved at 100% coherence, the tuning curves of the two example neurons to bidirectional stimuli were roughly symmetric (Fig.1, *C* and *E*). When one stimulus component moved at 100% and the other moved at 60% coherence, the tuning curves showed a larger peak at the side where the component response elicited by the higher coherence component was stronger than the other component response (Fig.1, *D* and *F*). Although the average of the component responses also showed such bias toward the stronger component response elicited by the higher-coherence component (gray curves in Fig.1, *D* and *F*), the observed bias of the bidirectional response was stronger than that predicted by the average of the component responses.

The two example neurons also showed some differences in their bi-directional responses. Under the equal-coherence condition, the bi-directional response of the first neuron was significantly less than the averaged component response across the whole direction tuning curve (Fig. 1C). In contrast, the second neuron roughly followed the average of the component responses when neither stimulus component moved near the PD of the neuron; When one stimulus component moved in a direction close to the PD (i.e. VA direction of $\pm 45^\circ$), the bi-directional response was stronger than the average of the component responses; When the VA direction was near the PD (i.e. 0°), the bi-directional response was weaker than the average of the component responses; As a result, the tuning curve showed two clearly separate peaks, although the averaged component response only had a single response peak (Fig. 1E).

Figure 2 shows the population-averaged tuning curves under the equal- and different-coherence conditions. Before averaging, the PD of the tuning curve for each neuron was aligned (see Methods). When both stimulus components moved at 100% coherence, the population-averaged tuning curve had two symmetric peaks, located around $\pm 45^\circ$ at which one of the component directions was in the PD (Fig. 2A). The peak responses were significantly stronger than the average of the component responses (one-tailed paired t-test, $p < 10^{-9}$). Moreover, the bi-directional response at the VA direction of the PD (i.e. 0°) was

slightly weaker than the average of the component responses (one-tailed paired t-test, $p < 0.02$), reinforcing the appearance of two response peaks in the tuning curve.

When one stimulus component moved at 100% and the other moved at 60% coherence, the response tuning curve strongly favored the higher-coherence component. Figure 2B shows the population-averaged direction tuning curves when the higher-coherence component moved at the clockwise side of the two component directions. At directions where the component response elicited by the higher-coherence component was greater than that elicited by the lower-coherence component, the response to the bi-directional stimuli was strongly biased toward the response elicited by the higher-coherence component (Fig. 2B). When the higher-coherence component moved in the PD of the neuron (i.e. VA direction of 45°), the bi-directional response was significantly greater than the average of the two component responses (one-tailed paired t-test, $p < 10^{-4}$). In contrast, at the other side of the tuning curve where the component response elicited by the lower-coherence component was stronger, the response to the bi-directional stimuli roughly followed the average of the component responses (Fig. 2B). We found similar results when the higher-coherence component moved at the CC side of the two component directions (Fig. 2C).

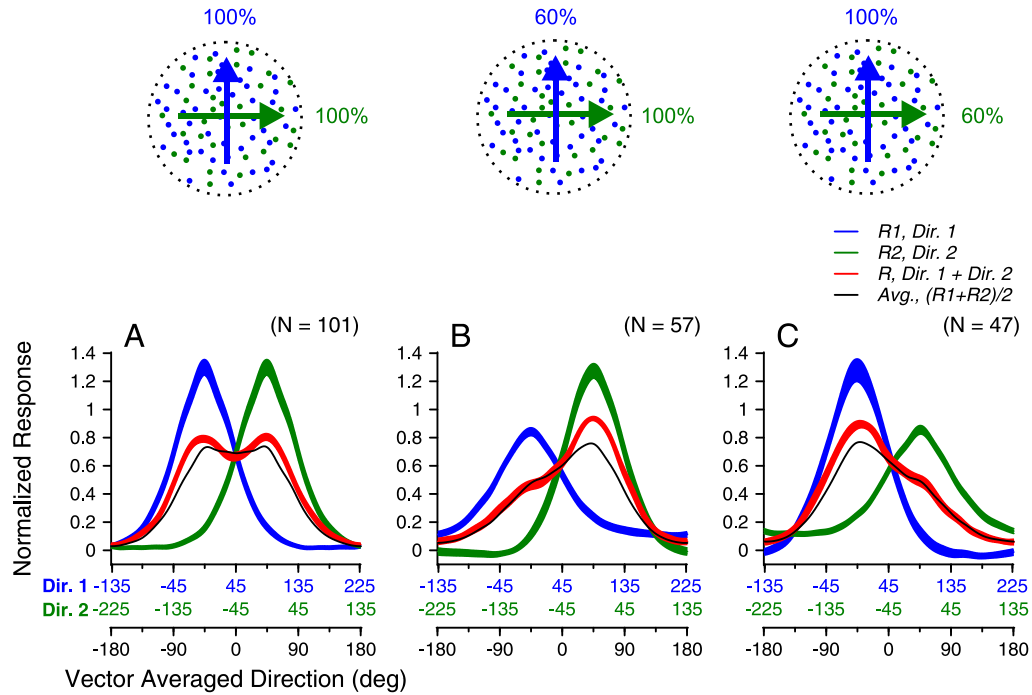


Figure 2, Population-averaged tuning curves elicited by bidirectional stimuli and the unidirectional components. VA directions of 0° was aligned with the PD for each neuron, and response tuning curves were normalized by the maximum response elicited by the bidirectional stimuli before averaging. The widths of the red, blue and green traces denote SE. **A:** both motion components moved at 100% coherence. **B:** the clockwise-side component moved at 100% coherence, whereas the CC-side component moved at 60% coherence. **C:** the CC-side component moved at 100% coherence, whereas the clockwise-side component moved at 60% coherence.

To quantify the bias of the response tuning curve elicited by bi-directional stimuli toward the response tuning elicited by the higher-coherence component, we computed the center of gravity (CG) of the response tuning curve (see Methods) and compared the CGs of the bi-directional responses with those of the component responses. Figure 3 shows the CGs of tuning curves obtained when the higher-coherence component moved either at the clockwise side (Fig. 3A1,2), or the CC side (Fig. 3B1,2) of two component directions. The CG of the bi-directional responses shifted toward the CG of the component responses elicited by the higher-coherence component more than did the CG of the averaged component responses. To quantify the extent of this shift, we calculated a *center of gravity shift index* (CGSI) (see Methods). A CGSI of greater than 0.5 indicates that the CG is closer to that of the higher-coherence component than the

lower-coherence component. The mean *CGSI* of 57 neurons shown in Figure 3A1 (mean = 0.76; std = 0.17) was significantly greater than 0.5 (one-tailed t-test, $p < 10^{-16}$), and was also significantly greater than the mean *CGSI* of the averaged component responses (mean = 0.67, std = 0.09) ($p = 1.8 \times 10^{-5}$) (Fig. 3A2). When the higher-coherence component moved at the CC side of two component directions, the mean *CGSI* of 47 neurons was also significantly greater than the mean *CGSI* of the averaged component responses ($p = 2.8 \times 10^{-4}$).

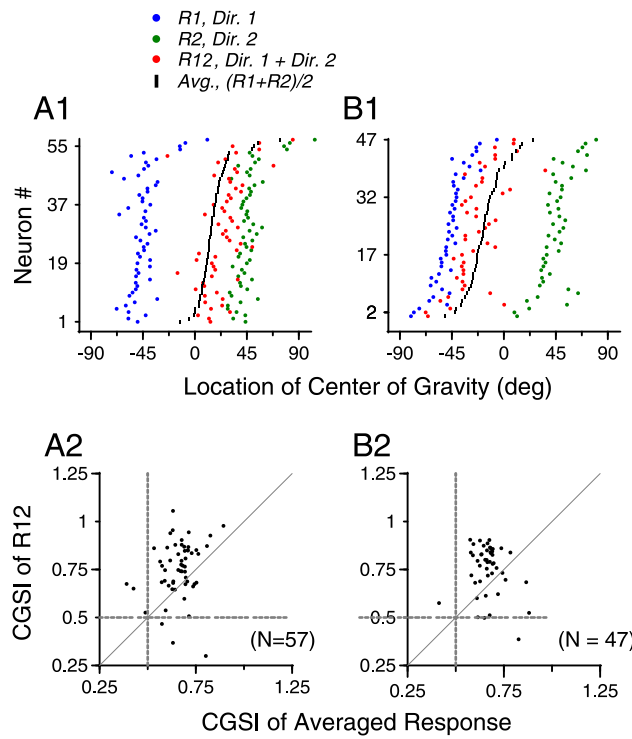


Figure 3, Center of gravity (CG) of the response tuning curve elicited by bi-directional stimuli when the constituent stimulus components moved at different levels of coherence. A. The higher-coherence component moved at the clockwise side of the two component directions. **B.** The higher-coherence component moved at the CC side of the two component directions. **A1, B1:** Each row represents data from one neuron. The green and blue dots indicate the CGs of the response tuning curves elicited by stimulus components moving at the clockwise and CC side of the two component directions, respectively. The red and black dots indicate the CGs of the response tuning curves elicited by bi-directional stimuli and those of the averaged component responses, respectively. Rows were sorted based on the CG of the averaged component responses. **A2, B2:** Scatter plots of the *CGSI* calculated based on bi-directional responses (ordinate) and the averaged component response (abscissa). One dot in A2 that had a *CGSI* for R_{12} of 1.39 and *CGSI* for the averaged responses of 0.85 is outside the axis scale and not shown.

Model fits of neuronal response tuning curves

To determine the rule that accounts for the neuronal response tuning curves elicited by bi-directional stimuli moving at the same or different levels of coherence, we fitted the data using several models. First, we fit the responses elicited by bi-directional stimuli as a weighted sum of the responses elicited by stimulus components presented alone, referred to as the linear weighted summation (LWS) model:

$$R_{12_{pred}}(\theta_1, \theta_2) = w_1 R_1(\theta_1) + w_2 R_2(\theta_2) + c, \quad (5)$$

where θ_1 and θ_2 are the two component motion directions, $R_{12_{pred}}$ is the tuning of the bi-directional response predicted by the model; R_1 and R_2 are the tuning curves of the component responses; w_1 and w_2 are the response weights for R_1 and R_2 , respectively, and c is a constant. We fixed c to zero in all data fits, except when noted.

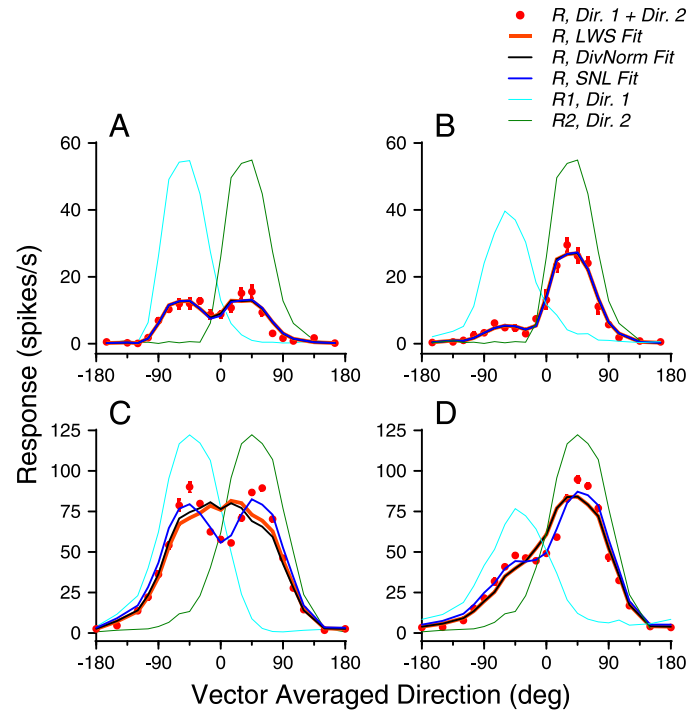


Figure 4. Model fits of the response tuning curves elicited by bi-directional stimuli of two example neurons. **A, C:** Two motion components both moved at 100% coherence. **B, D:** The clockwise-side component moved at 100% coherence and the CC-side component moved at 60% coherence. **A, B:** Example neuron 1. All three models (LWS, DivNorm and SNL) provided good fits. The fitting results of these models were almost identical. Therefore only one model fit is clearly visible. **C, D:** Example neuron 2. The LWS and DivNorm model fits were identical in D. Error bars of the bi-directional responses indicate standard errors.

Figure 4 shows the model fits for the direction tuning curves of the two example neurons shown in Figure 1. For the first example neuron, the LWS model provided good fits for the whole tuning curves, accounting for 93% of the response variance (see Methods) when both stimulus components moved at 100% coherence (Fig. 4A), and 98% of the variance when one stimulus component moved at 100% and the other moved at 60% coherence (Fig. 4B). For the second example neuron, the LWS model accounted for 87% of the response variance under the equal-coherence condition (Fig. 4C), and 94% of the variance under the different-coherence condition (Fig. 4D). Note that the LWS model failed to capture the two response peaks in Figure 4C and overestimated the responses near VA 0° in Figure 4D.

Across the neuron populations, the LWS model provided generally good fits of the data, accounting for > 81% of the response variance (see Table 1). The model accounted for the population-averaged tuning curves elicited by the bi-directional stimuli well (Fig. 5A1-C1), but with a caveat that the fits slightly overestimated the measured responses where the VA directions of the bi-directional stimuli were close to the neurons' PD. When two motion components moved at the same coherence, the mean weights of w_1 and w_2 obtained using the LWS model were 0.61 (std = 0.2) and 0.58 (std = 0.2), respectively. The two means were not significantly different ($N = 101$, paired t-test, $p > 0.1$). The pooled mean weight across w_1 and w_2 was 0.6 (std = 0.2) and was significantly smaller than 1 (one tailed t-test, $p = 2.6 \times 10^{-73}$), indicating robust subadditive summation. Furthermore, the mean weight was significantly greater than 0.5 (one tailed t-test, $p = 2.1 \times 10^{-11}$), indicating deviation from response averaging. Note that the response weights of the neurons in the population were distributed across a wide range (Fig. 5A2). For a given neuron, the bi-directional response can be closer to the weaker component response (e.g. Fig. 1C), the average of the component responses, or the stronger component response.

When two stimulus components moved at different levels of coherence, the response weights obtained from the LWS model fits were greater for the higher-coherence component than for the lower-coherence component. When the higher-coherence component moved at the clockwise side of two component directions, the mean response weight for the higher-coherence component was 0.66 (std = 0.19), significantly greater than that for the lower-coherence component of 0.45 (std = 0.21) ($N = 57$, one-tailed paired t-test, $p = 4.1 \times 10^{-7}$) (Fig. 5B2). When the higher-coherence component moved at the CC side of two component directions, the mean response weight for the higher-coherence component was also significantly greater than that for the lower-coherence component ($N = 47$, $p = 4.2 \times 10^{-5}$) (Fig. 5C2).

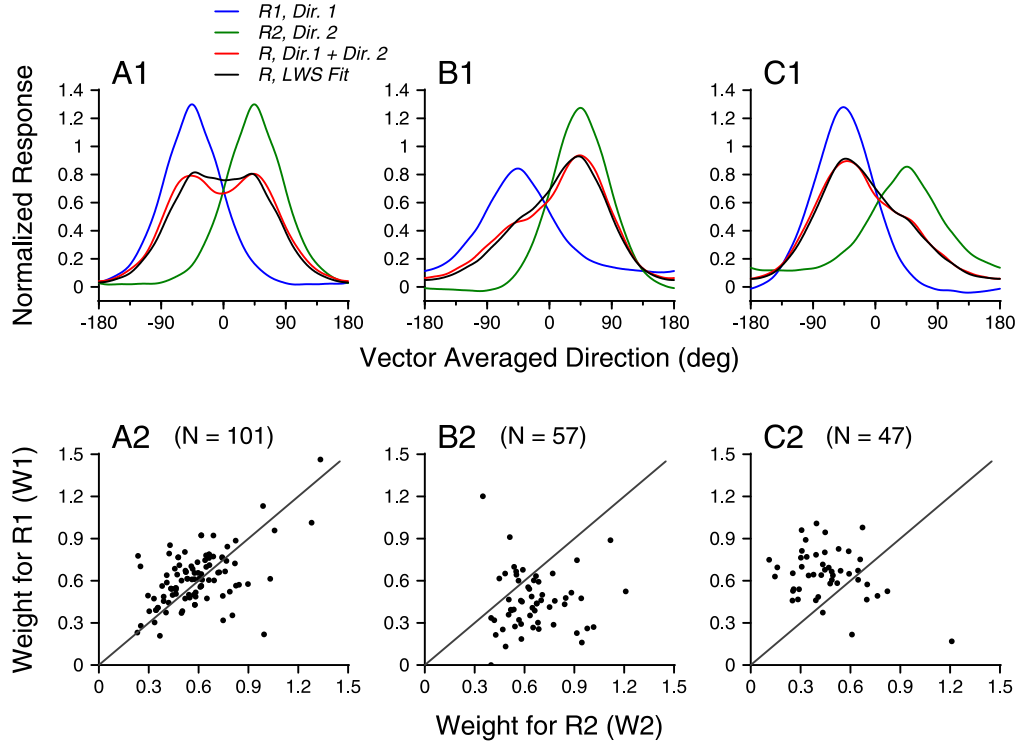


Figure 5. The LWS model fits of the population-averaged tuning curves and the response weights for the stimulus components. **A.** Both motion components moved at 100% coherence. **B.** The motion component at the clockwise side of the two component directions moved at 100% coherence and the CC-side component moved at 60% coherence. **C.** The CC-side component moved at 100% coherence and the clockwise-side component moved at 60% coherence. **A1-C1.** LWS fits of the population-averaged tuning curves elicited by bi-directional stimuli. **A2-C2.** Response weights for the two stimulus components obtained using the LWS fits.

Although the LWS model could account for the responses elicited by the bi-directional stimuli reasonably well, the model failed to capture certain salient features of the direction tuning curves for some neurons (e.g. see Fig. 4C and 5A1). We next asked whether allowing nonlinear interactions between the responses evoked by the stimulus components helped to improve the data fit. We fitted the bi-directional responses using a linear weighted sum of the component responses plus a multiplicative interaction term between the component responses, referred to as the summation plus nonlinear interaction (SNL) model:

$$R_{12_{pred}}(\theta_1, \theta_2) = w_1 R_1(\theta_1) + w_2 R_2(\theta_2) + b R_1(\theta_1) R_2(\theta_2) + c, \quad (6)$$

where b is a coefficient determining the sign and strength of the nonlinear interaction between the component responses and c is a constant. We fixed c to zero in all data fits, except when noted.

We found that the SNL model provided significantly better fits than the LWS model for 49 of 101 neurons under the equal-coherence condition (sequential F-test, $P < 0.01$) and better for 41 of 104 neurons under direction-coherence condition pooled across the two stimulus configurations (sequential F-test, $P < 0.01$; Fig.6 A1 and B1). The improvement of the goodness of fit by the SNL model for one example neuron is illustrated in Fig.4 C and D.

Table 1. Percentage of variance accounted for by different models

Stimulus Configuration Motion Coherence		Models (free parameters) Mean $\pm$ STD (%)						
Clockwise Side	Counter- clockwise side	LWS (w_1, w_2)	SNL (w_1, w_2, b)	LWS_C (w_1, w_2, c)	SNL_C (w_1, w_2, b, c)	CohNorm (n, σ)	DivNorm (n, α)	NNL (n, α, b)
100% ($n = 101$)	100% ($n = 101$)	81.6 $\pm$ 16.6	88.4 $\pm$ 9.0	84.9 $\pm$ 12.4	90.4 $\pm$ 7.2	76.6 $\pm$ 20.1	78.7 $\pm$ 17.5	85.8 $\pm$ 10.2
100% ($n = 57$)	60% ($n = 57$)	85.2 $\pm$ 13.2	90.9 $\pm$ 8.1	87.7 $\pm$ 10.8	91.9 $\pm$ 7.7	83.0 $\pm$ 17.3	83.5 $\pm$ 17.9	88.9 $\pm$ 14.6
60% ($n = 47$)	100% ($n = 47$)	86.3 $\pm$ 10.4	89.5 $\pm$ 9.4	87.8 $\pm$ 9.7	90.9 $\pm$ 8.4	83.6 $\pm$ 17.0	83.6 $\pm$ 17.4	87.6 $\pm$ 14.8

The improvement of the data fit by using the SNL model could not be explained by simply adding a constant parameter in the LWS model. When we allowed the constant c in the LWS and SNL model (Eqs. 5 and 6) to vary freely, the SNL model with the constant term (referred to as the SNL_C model) again outperformed the LWS model with the constant term (LWS_C) for 48 of 101 neurons under the equal-coherence condition and for 37 of 104 neurons under the different-coherence condition (sequential F-test, $P < 0.01$; Fig.6, A2 and B2). The significance test took into consideration that the SNL (or SNL_C) model

had one more free parameter than LSW (or LWS_C) model. When we compared the performance of LWS_C model with SNL model, in which the two models had the same number of parameters, SNL model still outperformed LWS_C model under both the equal- and different-coherence conditions ($N=101$ and 104) respectively. One-tailed paired t-test, $P < 2 \times 10^{-6}$). These results are consistent with the observation that the fitting error of the LWS model is not fixed, but rather varies with the stimulus directions and hence the component responses (see Fig. 5A1). The percentages of variance accounted for by the LWS and SNL models, with or without a constant term are shown in Table 1.

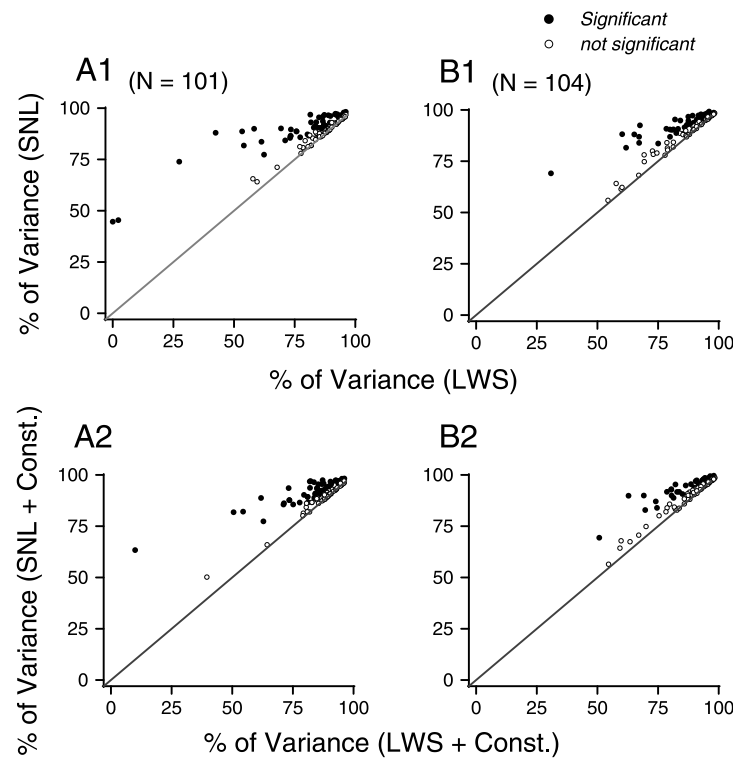


Figure 6. Comparison of the goodness-of-fit between the LWS and SNL models. A. Both motion components moved at 100% coherence. **B.** One motion component moved at 100% coherence and the other moved at 60% coherence. Data from the two stimulus configurations (i.e. the higher-coherence component moved either at the clockwise or CC side of the two component directions) were pooled together. **A1, B1.** The constant C in the LWS and SNL model was fixed at 0. **A2, B2.** The constant c in the LWS and SNL model was a free parameter.

Figure 7 shows the fitting results for the population data using the SNL model without the constant. The model fits captured population-averaged tuning curves, even where the LWS model overestimated the bi-directional responses (comparing Fig. 5A1-C1 with Fig. 7A1-C1). From the SNL fits, we obtained weights w_1 and w_2 for the component responses (see Eq. 6). When the two stimulus components both moved at 100% coherence, the mean weight was 0.67 (std = 0.29, N = 202, pooled across w_1 and w_2) (Fig. 7A2), which was significantly smaller than 1 (one-tailed t-test, $p = 2.4 \times 10^{-39}$) indicating subadditive summation. The weights were greater than 0.5 ($p = 2.1 \times 10^{-14}$), indicating deviation from response averaging. When the two stimulus components moved at different coherence levels, the mean response weight was significantly larger for the higher-coherence component than for the lower-coherence component (Fig. 7B2, C2, $p < 10^{-6}$). When the higher-coherence component moved at the clockwise side of two component directions, the mean response weights for the higher- and lower-coherence components were 0.78 (std = 0.4) and 0.5 (std = 0.26, N = 57), respectively. When the higher-coherence component moved at the CC side, the mean response weights for the higher- and lower-coherence components were 0.77 (std = 0.35) and 0.48 (std = 0.2, N = 47), respectively.

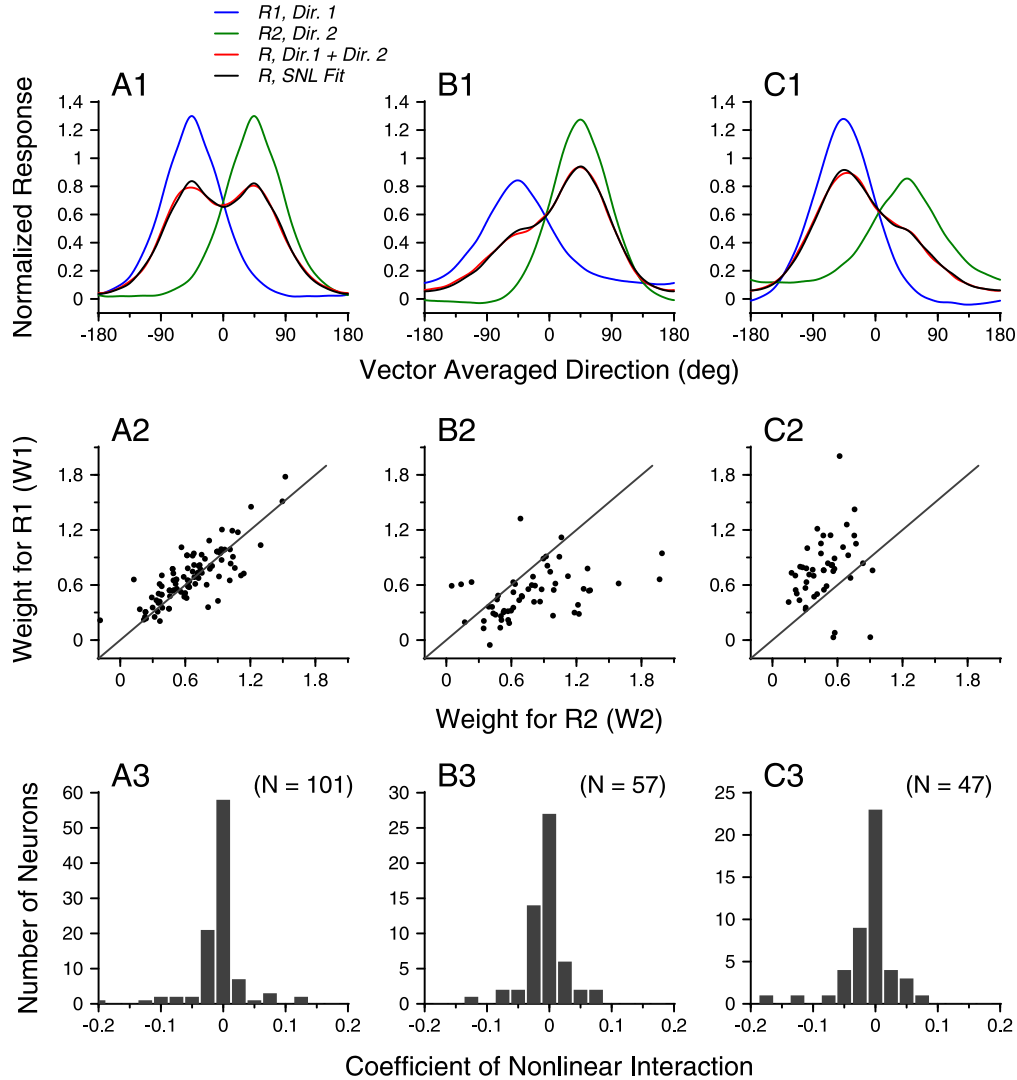


Figure 7. The SNL model fits of the population-averaged tuning curves and the response weights for the stimulus components. A. Both motion components moved at 100% coherence. **B.** The clockwise-side component moved at 100% coherence and the other motion component moved at 60% coherence. **C.** The CC-side component moved at 100% coherence and the other motion component moved at 60% coherence. **A1-C1.** SNL fits of the population-averaged tuning curves elicited by bi-directional stimuli. **A2-C2.** Response weights for the two stimulus components obtained using the SNL fits. **A3-C3.** Distribution of the nonlinear interaction coefficient of the SNL fits.

Importantly, the nonlinear multiplicative interaction between the component responses determined by the SNL fits tended to have a negative coefficient. Under the equal-coherence condition, SNL fits to the

responses of 71 (out of 101) neurons had negative nonlinear interaction coefficients (i.e. parameter b in Eq. 6). The mean coefficient was -0.016 (std = 0.09, $N = 101$), which was significantly negative (one-tailed t-test, $p = 0.04$). Under the two different-coherence conditions, the mean nonlinear interaction coefficients were also significantly negative (one-tailed t-test, $p < 0.05$). The distributions of the interaction coefficient are shown in Figure 7A3-C3.

Data fitting using both the LWS and SNL models showed subadditive summation and the response weights varied with the relative coherence levels of the motion components. We asked whether these results could be accounted for by a divisive normalization model (Carandini and Heeger 2012). We first fitted the data using a model after the “contrast normalization” model that was used to describe the response of V1 neurons elicited by gratings with different luminance contrasts (Busse et al. 2009; Carandini et al. 1997). We referred to this model as the “coherence normalization” model (CohNorm):

$$R_{12_{pred}}(\theta_1, \theta_2) = \frac{h_1^n}{\left(\sqrt{h_1^2 + h_2^2}\right)^n + \sigma^n} R_1(\theta_1) + \frac{h_2^n}{\left(\sqrt{h_1^2 + h_2^2}\right)^n + \sigma^n} R_2(\theta_2) \quad (7)$$

, where h_1 and h_2 are the coherence levels of the two motion components; n is a positive exponent; and σ is a positive constant, representing the semi-saturation coherence. The CohNorm model accounted for 76.6% of the response variance under the equal-coherence condition, and 83% of the variance under the different-coherence conditions (Table 1). The median exponent n was 1.3 and 0.8 under the equal- and different-coherence condition, respectively.

In the divisive normalization model proposed by Heeger and colleagues, the response of a single neuron is determined by the linear input that the neuron receives, divisively normalized by the summed activity of a pool of neurons (Heeger 1992; Simoncelli and Heeger, 1998; see Carandini and Heeger, 2011 for a review). In the denominator of Eq. 7, the activity of the neuron pool elicited by the bi-directional stimuli is represented as proportional to the combined motion coherence of the two stimulus

components $\left(\sqrt{h_1^2 + h_2^2}\right)^n$, as in the contrast normalization model (Carandini et al., 1997; Busse et al., 2009). Since the signal strength (motion coherence or luminance contrast) has a positive value and the exponent n is also positive, any possible multiplicative terms between the signal strengths of the two stimulus components after expanding the term $\left(\sqrt{h_1^2 + h_2^2}\right)^n$ would have only positive coefficients. However, as we have shown above, allowing a multiplicative interaction between the component responses provides significantly better fit of the responses elicited by the bi-directional stimuli and the coefficient of the multiplicative interaction is often negative. Motivated by these observations, we used a modified normalization model to fit our data.

In the modified divisive normalization model, we represented the response of the neuron pool in the denominator evoked by the bi-directional stimuli as a weighted sum of the population responses evoked by each stimulus component plus a multiplicative interaction term between the component responses. We assumed that the population neural response evoked by a stimulus component was proportional to a power-law transformation of the stimulus strength. The model has the form:

$$R_{12pred}(\theta_1, \theta_2) = \frac{h_1^n}{p_1 * h_1^n + p_2 * h_2^n + \alpha * h_1^n * h_2^n} R_1(\theta_1) + \frac{h_2^n}{p_1 * h_1^n + p_2 * h_2^n + \alpha * h_1^n * h_2^n} R_2(\theta_2) \quad (8)$$

,where h_1 and h_2 are the signal strengths of the two stimulus components; n is a *positive* exponent that transforms the signal strength into a measure of neural activity; p_1 and p_2 are two positive parameters governing the relative contribution of the population neural response evoked by each stimulus component to the population response evoked by the bi-directional stimuli; α is the coefficient of the interaction term that can be positive or negative, which allows either a facilitatory or suppressive effect of the multiplicative

interaction on the population response. The parameters p_1 , p_2 allow the response weights for the two stimulus components to have different values under the equal-coherence condition. Note that, in Experiment one, h_1 and h_2 were fixed and therefore the interaction term $\alpha h_1^n h_2^n$ is essentially the same as a single constant. This normalization model fitted the bi-directional responses under the equal-coherence condition as well as the LWS model, accounting for on average 81.6% (std = 16.6, N = 101) of the variance.

We found that setting both p_1 and p_2 to 1 still provided good fits of the data. The simplified normalization model, referred to as the DivNorm model can be expressed as:

$$R_{12_{pred}}(\theta_1, \theta_2) = \frac{h_1^n}{h_1^n + h_2^n + \alpha h_1^n h_2^n} R_1(\theta_1) + \frac{h_2^n}{h_1^n + h_2^n + \alpha h_1^n h_2^n} R_2(\theta_2) \quad . \quad (9)$$

The fits of the DivNorm model for the responses of the two example neurons are illustrated in Figures 4, together with the LWS and SNL model fits. Under the equal-coherence condition and when the motion coherence was 100% (i.e. one), the exponent n became irrelevant. The DivNorm model essentially had just one free parameter α . Nevertheless, this one-parameter model accounted for on average 78.7% of the variance of the responses elicited by the bi-directional stimuli. The goodness-of-fit of the DivNorm model was slightly better than that of the CohNorm model (Table 1) and the difference was significant (one-tailed paired t-test, $p = 0.019$). Under the different-coherence condition, both n and α were free parameters; the simplified model accounted for 83.5% of the response variance, not significantly different from that of the CohNorm model fits (Table 1).

Given that the SNL model provided significantly better fit than the LWS model, we added a nonlinear interaction term to the DivNorm model. The resultant model is referred to as the normalization plus nonlinear interaction (NNL) model (Eq. 10). Again, the parameter b is a coefficient determining the sign and strength of the nonlinear interaction between the component responses. The goodness-of-fit of the

NNL model was better than the LWS and DivNorm models and was slightly worse than the SNL model (Table 1).

$$R_{12_{pred}}(\theta_1, \theta_2) = \frac{h_1^n}{h_1^n + h_2^n + \alpha * h_1^n * h_2^n} R_1(\theta_1) + \frac{h_2^n}{h_1^n + h_2^n + \alpha * h_1^n * h_2^n} R_2(\theta_2) + b * R_1(\theta_1) R_2(\theta_2). \quad (10)$$

For all the models, it was not necessary to vary the response weights as a function of the stimulus direction to account for the whole direction tuning curve. A fixed set of weights could well account for the bi-directional responses across motion directions, suggesting that the neural mechanism underlying the response weights can be invariant to the stimulus direction.

Relationship between bi-directional responses and the component responses across multiple levels of motion coherence

To further understand the rule governing the responses elicited by multiple stimuli and how the response weights were determined, we examined the relationship between bi-directional responses and the component responses across multiple levels of motion coherence. In this experiment, we set the direction of one stimulus component at the PD of the neuron (referred to as the PD component) and the other 90° away from the PD. We varied the motion coherence of the PD component from 60% to 100% and fixed the motion coherence of the orthogonal motion component at 100%.

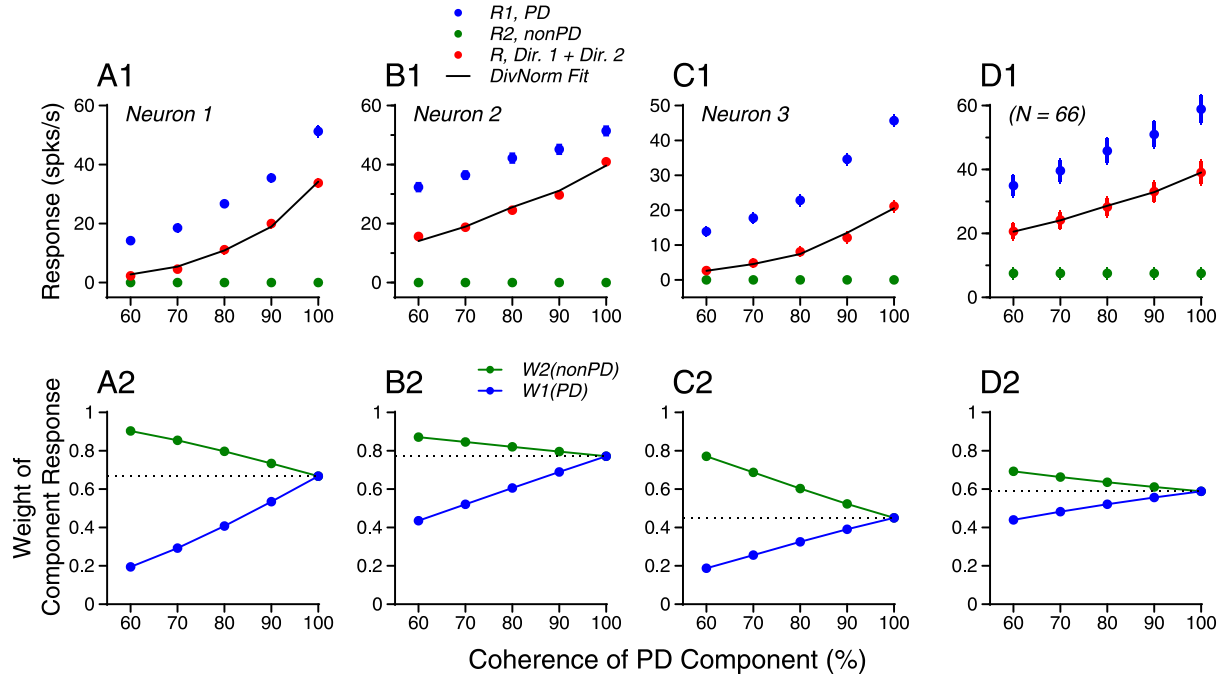


Figure 8. Responses of three example neurons and population averaged responses elicited by bi-directional stimuli as a function of the coherence level of the PD component ($R1$). The coherence level of the orthogonal motion component ($R2$) was fixed at 100%. **A1-C1.** Firing rates of bi-directional responses and the corresponding component responses of three example neurons. **D1.** Population averaged responses. Error bars indicate standard errors. Solid lines are DivNorm model fits. **A2-D2.** Corresponding to A1-D1, response weights for the two stimulus components obtained using the simplified DivNorm model fits.

Figure 8 shows the results from three example neurons. As the coherence level of the PD component increased, the response to the bi-directional stimuli changed its magnitude *relative to the component responses*. For the first neuron, when the motion coherence of the PD component was 60%, the bi-directional response was very close to the response elicited by the orthogonal motion component that had a higher coherence; as the motion coherence of the PD component increased, the bi-directional response became closer to the response elicited by the PD component (Fig. 8A1). For the second neuron, the bi-directional response was near the average of the two component responses when the coherence of the PD component was 60% and the bi-directional response became closer to the response elicited by the PD component as its coherence level increased (Fig. 8B1). For the third neuron, the bi-directional response was

near the response elicited by the orthogonal motion component when the coherence of the PD component was 60% and the bi-directional response was close to the average of two component responses when the PD coherence was 100% (Fig. 8C1). Despite these differences across individual neurons, a consistent trend was that the response to the PD component appeared to have an increasing contribution to the bi-directional response as the motion coherence of the PD component increased. This trend can also be seen from the population-averaged responses across 66 neurons (Fig. 8D1). Because the LWS and SNL models both assumed that the response weights were fixed across stimulus conditions and therefore were inappropriate for fitting the results of this experiment, we fitted the responses using the normalization model, which allowed the response weights to vary with the motion coherence.

We found that MT responses to the bi-directional stimuli across multiple coherence levels could be well accounted for by the DivNorm model (Eq. 9), in which the response weight for a motion component is expressed as the following (Eqs. 11, 12):

$$w_{PD} = \frac{h_{PD}^n}{h_{PD}^n + h_{orthogonal}^n + \alpha \bullet h_{PD}^n \bullet h_{orthogonal}^n} , \quad (11)$$

$$w_{orthogonal} = \frac{h_{orthogonal}^n}{h_{PD}^n + h_{orthogonal}^n + \alpha \bullet h_{PD}^n \bullet h_{orthogonal}^n} , \quad (12)$$

where w_{PD} and $w_{orthogonal}$ are the response weights and h_{PD} and $h_{orthogonal}$ are the motion coherence levels for the PD and orthogonal motion components, respectively. The DivNorm model provided excellent fits for the bi-directional responses of the three example neurons and the population averaged responses, accounting for greater than 98% of the response variance (Fig. 8A1-D1). As the motion coherence of the PD component increased from 60% to 100%, the response weight for the PD component increased progressively (Fig. 8A2-D2).

Across the population of 66 neurons, the DivNorm model provided good fits for bi-directional responses as the motion coherence was varied. On average, the model accounted for 80.7% (std = 29.5) of the response variance. The population mean of the response weight for the PD component varied significantly as the motion coherence of the PD component changed (one-way ANOVA, $F = 8.1$, $p = 3.1 \times 10^{-6}$). The mean response weight for the PD component, averaged across the population of 66 neurons, progressively increased from 0.39 to 0.60, as the motion coherence of the PD component increased from 60 to 100% (Fig. 9A). The median value of the exponent n of the DivNorm model fits was 1.04, suggesting a near-linear transformation between stimulus motion coherence and the magnitude of MT population response, which is consistent with the population averaged firing rate shown as a function of the motion coherence of the PD component (blue dots in Figure 8D1). Since different levels of motion coherence were randomly interleaved across trials, the change of the response weights occurred on the time scale of a single experimental trial. To determine how the response weight for the PD component changed with the motion coherence on a neuron-by-neuron basis, we fitted the response weights across five coherence levels for each neuron using linear regression. Figure 9B shows the distribution of the slope of the linear fit. The mean fitted slope was significantly positive (mean = 0.51, std = 0.39) (one-tailed t-test, $p = 4.2 \times 10^{-16}$), indicating that the response weight for the PD component increased with its coherence.

For comparison, we also fitted the data using the CohNorm model (Eq. 7). The CohNorm model accounted for 72.1% (std = 36.0) of the response variance, significantly smaller than the mean of 80.7% of the DivNorm model (one-tailed paired t-test, $p = 0.0015$). When we allowed the DivNorm and CohNorm model to have an additional constant parameter that can be either positive or negative, as the parameter c in the LWS_C and SNL_C model (Eqs. 5,6), the variance accounted for by the DivNorm and CohNorm model increased to 87.3% and 82.3%, respectively. The variance accounted for by the DivNorm model was still significantly larger than the CohNorm model (one-tailed paired t-test, $p = 0.0026$).

The normalization model makes a specific assumption regarding how the response weights are determined by the relative levels of motion coherence (Eqs. 11, 12). To verify whether the response weight for a stimulus component increased with its motion coherence, we used an approach that does not make any assumption regarding how the response weights change with the motion coherence. We calculated the response weight at each coherence level based on the relative magnitudes of the bi-directional response and the two component responses (Eq. 13):

$$R_{12} = \frac{R_2 - R_{12}}{R_2 - R_1} R_1 + \frac{R_{12} - R_1}{R_2 - R_1} R_2, \quad (13)$$

where R_{12} is the bi-directional response; R_1 and R_2 are the responses elicited by the PD and the orthogonal motion component, respectively. Therefore, the response weights are: $w_{PD} = \frac{R_2 - R_{12}}{R_2 - R_1}$, $w_{orthogonal} = \frac{R_{12} - R_1}{R_2 - R_1}$. The response weights were simply determined by how close the bi-directional response was to one of the component responses, relative to the distance between the two component responses. If the bi-directional response was closer to one component response, that motion component had a higher weight than the other component. Note that the two response weights sum to 1. Since the right-hand side of the Equation 13 can be rearranged as $\frac{R_2 R_1 - R_{12} R_1 + R_{12} R_2 - R_1 R_2}{R_2 - R_1} = \frac{R_{12}(R_2 - R_1)}{R_2 - R_1} = R_{12}$, Equation 13 always holds as long as the two component responses R_1 and R_2 are not identical.

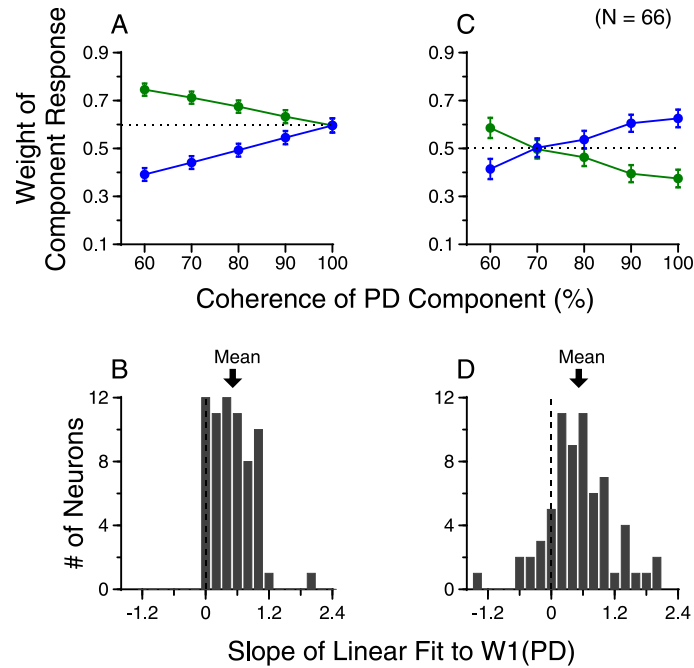


Figure 9. Response weights across multiple levels of motion coherence of a population of neurons. A, C. Averaged response weights as a function of the coherence level of the PD component. Error bars indicate standard errors. **B, D.** The distribution of the slope of the linear fit of response weights for the PD component across coherence levels. **A, B.** Weights were obtained by using the DivNorm fits. **C, D.** The response weights were calculated at each coherence level based on the relative magnitudes of the bi-directional and the component responses.

Figure 9C shows the averaged response weights across the neuron population obtained using Equation 13. The mean weights for the PD component varied significantly as the motion coherence of the PD component changed (one-way ANOVA, $F = 4.7$, $p = 0.001$). The mean weight increased progressively from 0.41 to 0.63 as the motion coherence of the PD component increased from 60 to 100%, very similar to those obtained using the DivNorm model fits. We also fitted the response weights for the PD component across five coherence levels for each neuron using linear regression. The mean slope across the neuron population was 0.52, which was significantly positive (one-tailed t-test, $p = 8.7 \times 10^{-10}$) (Fig. 9D), and almost the same as that obtained using the DivNorm model. These findings confirmed the results obtained using the DivNorm model, showing that the response weight for a stimulus component progressively increased with the coherence level of that stimulus component.

Direction tuning of neuronal responses to bi-directional stimuli moving at different luminance levels

To test the generality that the response weights changed with the relative signal strengths, we next varied the luminance levels of the stimulus components and characterized the direction tuning curves elicited by the bi-directional stimuli. The luminance of the stimulus component moving at the clockwise side of the two component directions was 40 cd/m². The luminance of the stimulus component at the CC side was either 10 or 2.5 cd/m². Figure 10 shows the results from two example neurons. The tuning curve of the bi-directional response was biased toward the tuning curve of the higher-luminance component (i.e. the right side in the plot). The response bias was stronger when the luminance difference between the two stimulus components was larger (Fig. 10B). For these two neurons, the peak response elicited by the higher-luminance component was slightly weaker than that elicited by the lower-luminance component of 2.5 cd/m² (Fig. 10B). Nevertheless, this did not prevent the bi-directional response from showing a strong bias toward the response elicited by the higher-luminance component, suggesting that the response bias was not determined by the component responses of a *single neuron*. Across the neuron population, the mean peak firing rates to the stimulus components of 2.5, 10 and 40 cd/m² were 68.5 Hz (N=30), 72.1 and 75.8 Hz (N=38), respectively, showing a modest increase with the luminance. The normalized and population-averaged tuning curve elicited by the bi-directional stimuli also biased toward the response tuning elicited by the higher-luminance component (Fig. 11). The bias was stronger when the luminance difference between the two stimulus components was larger.

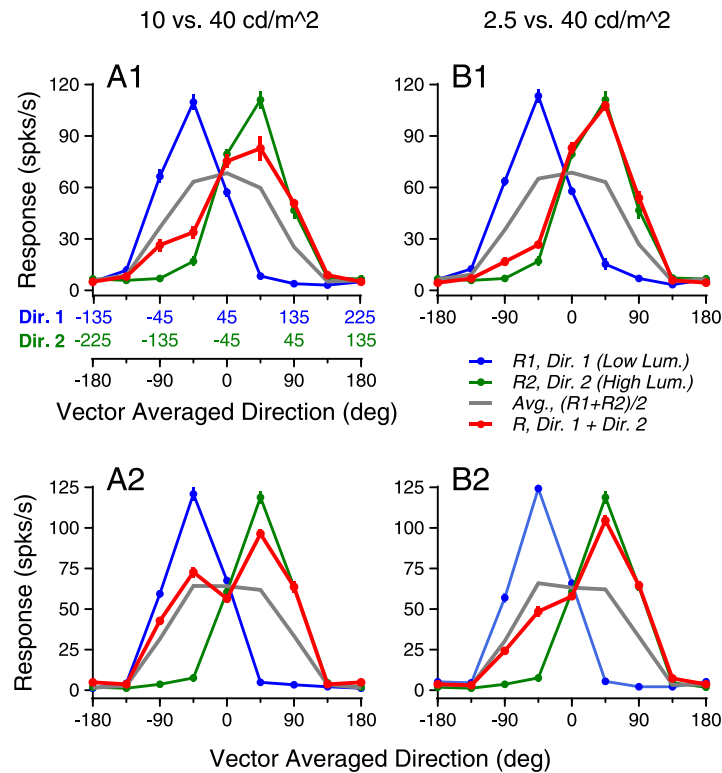


Figure 10. The direction tuning curves of two example neurons elicited by bi-directional stimuli and the stimulus components that had different luminance levels. The higher luminance of the two stimulus components was 40 cd/m^2 . The left and right columns show results when the lower luminance was 10 and 2.5 cd/m^2 , respectively. The higher-luminance component always moved at the clockwise side of the two component directions. The convention of the x-axis is the same as in Figure 1. **A1, B1.** Response tuning curves of one example neuron. **A2, B2.** Responses of another example neuron.

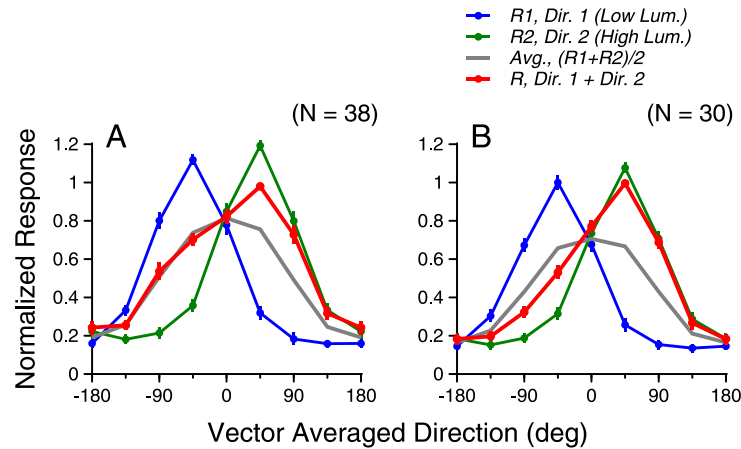


Figure 11. Population averaged tuning curves elicited by bi-directional stimuli and the stimulus components that had different luminance levels. Before averaging, VA direction of 0° was aligned with the PD for each neuron and neuronal responses were normalized to the maximum bi-directional response. The higher luminance of the two stimulus components was 40 cd/m^2 and the higher-luminance component moved at the clockwise side of the two component directions. The lower luminance of the two stimulus components was 10 cd/m^2 in **A**, and 2.5 cd/m^2 in **B**.

The tuning curves of neuronal responses elicited by the bi-directional stimuli were well accounted for by a weighted sum of the component responses (Fig. 12). The LWS model accounted for $\geq 91\%$ of the variance for the two example neurons and the population-averaged response. The LWS model accounted for on average 93.0% (std = 6.6) of the variance when the lower luminance was 10 cd/m^2 , and 94.2% (std = 4.8) when the lower luminance was 2.5 cd/m^2 .

We also fitted the data using the DivNorm model, replacing motion coherence h in Equation 9 with the luminance (L) of the stimulus component. The DivNorm model gave rise to almost identical fitting results as the LWS model, accounting for 92.9% (std = 6.8) of the response variance. The median values of the exponent n of the DivNorm model fits were 0.46 and 0.45 when the lower luminance of the two stimulus components was 10 and 2.5 cd/m^2 , respectively. The compressive power-law nonlinearity (i.e. $n < 1$) is consistent with the modest increase of the firing rate of MT neurons as the stimulus contrast increases when

the contrast is high, as of our stimuli. We also fitted the data by replacing motion coherence h in the CohNorm model (Eq. 7) with the luminance (L) and obtained the same fitting results as the DivNorm model.

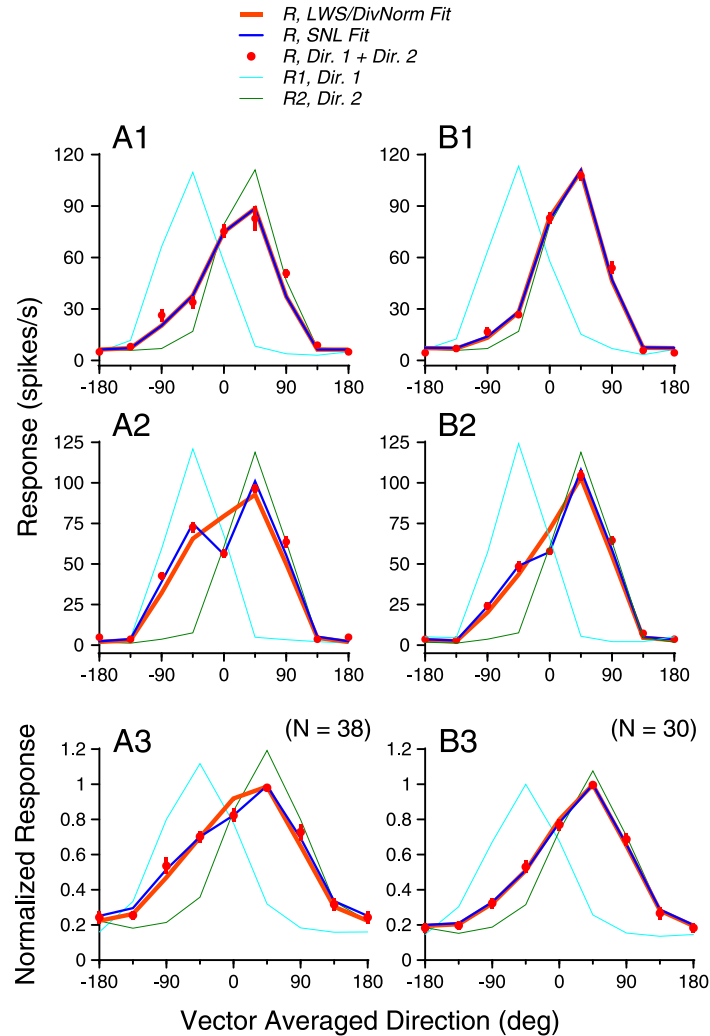


Figure 12. Model fits of response tuning curves elicited by the bi-directional stimuli moving at different luminance levels. The left and right columns show results when the luminance levels of the two stimulus components were 10 vs. 40 cd/m^2 , and 2.5 vs. 40 cd/m^2 , respectively. The DivNorm model fits gave the same results as the LWS fits. **A1, B1.** Example neuron #1. The three models, LWS, DivNorm and SNL, provided almost identical fits. **A2, B2.** Example neuron #2. The SNL model provided better fits than the LWS/DivNorm models. Including response nonlinearity increased the variance accounted for from 91.0% (LWS/DivNorm) to 98.8% (SNL) when the lower luminance was 10 cd/m^2 , and increased the variance from 96.8% (LWS/DivNorm) to 99.5% (SNL) when the lower luminance was 2.5 cd/m^2 . **A3, B3.** Model fits of the population-averaged tuning curves. Error bars of the bi-directional responses indicate standard errors.

Using the SNL model that allowed nonlinear interactions between the component responses improved the goodness-of-fit (e.g. Fig. 12A2, B2). On average, the SNL model accounted for 96.5% (std = 4.2) of the variance when the lower luminance was 10 cd/m², and 96.1% (std = 3.6) of the variance when the lower luminance was 2.5 cd/m².

The response weights for the higher-luminance component were significantly greater than those for the lower-luminance component (Fig. 13). Using the LWS fits, the median response weight for the higher-luminance component was 0.74, significantly different from that for the lower-luminance component of 0.39 (N = 38, signed-rank test, $p = 9.8 \times 10^{-8}$) when the luminance levels of the two stimulus components were 10 and 40 cd/m², respectively. As the luminance difference between the two motion components increased, the response weight for the higher-luminance component further increased and the weight for the lower-luminance component decreased. When the luminance levels of the stimulus components were 2.5 and 40 cd/m², respectively, the median response weight for the higher-luminance component was 0.86, significantly different from that for the lower-luminance component of 0.27 (N=30, signed-rank test, $p = 1.7 \times 10^{-6}$). We found the same results when the response weights were obtained using the SNL fits (signed-rank test, $p < 1.8 \times 10^{-6}$). These results indicate that MT neurons weighted the component response elicited by the higher-luminance component more strongly. Again, a fixed set of response weights could account for the bi-directional responses across all stimulus directions when the two stimulus components differed in their luminance, suggesting that the neural mechanism underlying the response weights can be invariant to the stimulus direction.

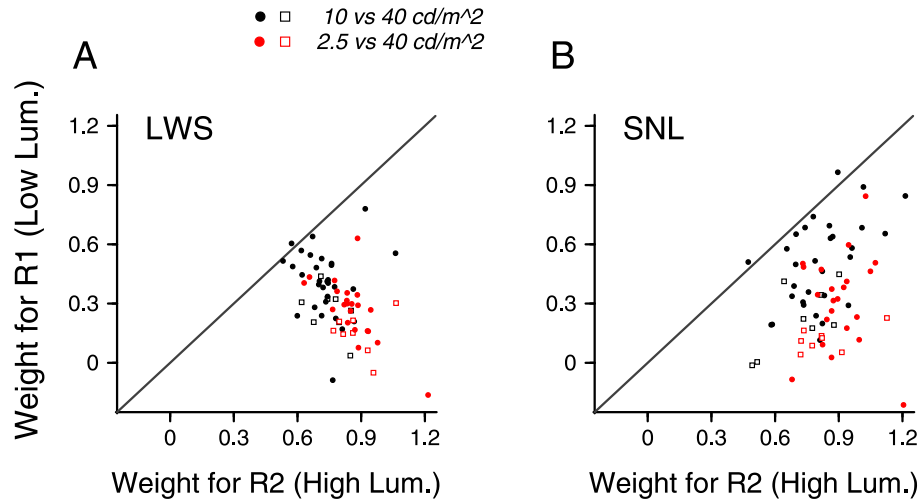


Figure 13. Response weights for the two motion components that had different luminance levels. A. Response weights obtained using the LWS fits. **B.** Response weights obtained using the SNL fits. When the lower luminance was 10 cd/m^2 , the median response weights for the higher- and lower-luminance components obtained using the SNL fits were 0.79 and 0.43, respectively. When the lower luminance was 2.5 cd/m^2 , the median response weights for the higher- and lower-luminance components were 0.87 and 0.23, respectively. Solid circles show results from monkey BJ; open squares show results from monkey RG.

2.4 Discussion

We have examined how overlapping moving stimuli that have the same or different signal strengths interact within the RFs of neurons in the extrastriate area MT. Our principle finding is that, across signal strengths and motion directions, neuronal responses elicited by multiple stimuli moving in different directions can be well accounted for by a sub-linearly weighted sum of the neuronal responses elicited by the stimulus components, plus a nonlinear interaction term between the component responses that typically has a suppressive effect on the neuronal response. Importantly, rather than always weight each stimulus

component equally, MT neurons weight the stimulus component that has the higher signal strength more strongly, regardless of whether the signal strength is defined by motion coherence or luminance contrast.

Weighting component responses according to signal strengths

Our results are consistent with previous findings made in V1 using oriented gratings and in the dorsal medial superior temporal area (MSTd) using multisensory stimuli. Responses of V1 neurons elicited by overlapping gratings follow a weighted sum of the component responses and the response weight is stronger for the stimulus component that has a higher luminance contrast (Busse et al., 2009; MacEvoy et al., 2009). In MSTd, responses of multisensory neurons elicited by optical flow and vestibular self-motion cues also follow a weighted sum of the component responses and the response weights change with the relative reliabilities of the visual and vestibular cues (Morgan et al., 2008; Fetsch et al., 2011). Our study provides new evidence supporting a general rule governing how neurons respond to multiple stimuli within their RFs. It appears that neurons weight the responses elicited by the constituent stimuli according to their signal strengths, weighting the stimulus component that has a higher signal strength more strongly. This scheme of selective weighting helps to make the stimulus that has a higher signal strength more salient and suppress the stimulus that has a weaker signal strength, implementing a type of stimulus competition useful for noise reduction and image segmentation.

When multiple stimuli are integrated to generate a single-valued behavioral output, weighting according to the signal strength would lead to a behavior that strongly favors the stimulus component having a higher signal strength. For multisensory integration, weighting multimodal cues according to their signal strengths gives rise to a perceived heading direction that is shifted toward the more reliable cue and such a scheme may provide a near optimal solution for multisensory cue integration (Morgan et al., 2008; Fetsch et al., 2011). In the case of smooth pursuit (see Lisberger, 2010 for a review), when two pursuit targets

moving in different directions have the same luminance, the initiation of pursuit follows the vector-averaged direction of the two targets (Lisberger and Ferrera, 1997). However, when two overlapping patches of moving random-dots have different levels of luminance, the initiation of pursuit is strongly biased toward the direction of the brighter random-dot patch, showing an effect of winner-take-all (Niu and Lisberger, 2011). Our finding that the direction tuning curves of MT neurons elicited by overlapping stimuli are strongly biased toward the response elicited by the higher-luminance component provides a likely physiological basis for the pursuit behavior reported by Niu and Lisberger (2011). Since we have found that response weighting also depends on relative motion coherence, our results predict that when overlapping pursuit targets move at different levels of coherence, the initiation of pursuit should bias toward the direction of the higher-coherence component.

Divisive normalization and population neural response

Response normalization in the extrastriate area MT has been suggested previously (Simoncelli and Heeger, 1998; Britten and Heuer, 1999; Heuer and Britten, 2002; Lee and Maunsell, 2009; Ni et al. 2012). Our study using multiple stimuli with different signal strengths provide new constraints for models of response normalization. We have shown that in area MT the changes of response weights with relative signal strengths, defined either by motion coherence or luminance contrast can be accounted for as the signal strengths of individual stimulus components divisively normalized by a combination of the signal strengths of all stimulus components. Our results are consistent with the findings that the changes of response weights in V1 with relative stimulus contrasts can be accounted for by a contrast normalization model (Busse et al., 2009; Carandini et al., 1997), and the changes of response weights of multisensory

neurons in MSTd with relative reliabilities of visual and vestibular cues can also be explained by a divisive normalization model (Ohshiro et al., 2011).

Our results show that, for a given MT neuron, its response weights to two stimulus components are not determined by the single neuron's response magnitudes elicited by the stimulus components, but rather by the relative signal strengths of the stimulus components. For example, the response magnitude of an MT neuron elicited by a low-coherence stimulus component moving in the PD of the neuron is typically greater than the neuron's response elicited by a high-coherence component moving in the orthogonal direction. However, the neuron may still show larger response weight for the high-coherence orthogonal component than the low-coherence PD component (e.g. see Fig. 8A1, C1). Moreover, the tuning curve of a MT neuron elicited by bi-directional stimuli tends to show a strong bias toward the stimulus component that had a higher luminance contrast, even when the single neuron's peak responses elicited by the stimulus components with different luminance contrasts are similar (e.g. see Fig. 10). In comparison, the population-averaged responses of MT neurons are correlated with power-law transformed signal strengths. In our data, the population-averaged response of MT neurons increases modestly with the stimulus luminance, which is consistent with previous reports of MT contrast response functions within a range of high luminance contrast (Sclar et al., 1990; Cheng et al., 1994; Heuer and Britten, 2002). The population-averaged response of MT neurons also increases robustly with the motion coherence (Britten et al., 1992, 1993; and our results shown in Fig. 8D1). These findings are consistent with the idea that the response weight is determined by the activity of a population of neurons that reflects the signal strength of the visual stimuli.

In the normalization models proposed by Heeger (1992), Simoncelli and Heeger (1998) and Ohshiro et al. (2011), the linear inputs that a neuron receive are divisively normalized by the summed activity of a population of neurons, referred to as the normalization pool. Unlike those models, we did not directly compute the summed activity of a population neurons as the denominator for response normalization. Instead, we used a combination of the signal strengths of stimulus components in the

denominator to represent the population neural response. Our model is similar to the contrast normalization model (Carandini et al., 1997; Busse et al., 2009), with a notable difference. The denominator in our model does not reflect the "overall" or "averaged" signal strength of multiple stimuli, which may not be easily defined when multiple stimulus components can be segmented perceptually. We used a linear combination of the power-law transformed signal strengths plus a nonlinear, multiplicative interaction term to represent the activity of a population of neurons evoked by multiple, simultaneously presented stimuli. Our choice of the model was motivated by the success of the SNL model in describing MT neuronal responses elicited by bi-directional stimuli. Because the nonlinear interaction term can have either a facilitatory or suppressive effect on the population neural response, our DivNorm model has more flexibility in capturing the population neural response and it provided better goodness-of-fit than the CohNorm model when fitting the data obtained using stimuli with multiple levels of motion coherence.

The idea that the response weights are determined by the population neural responses makes a testable prediction: the response weight for a stimulus component should increase (or decrease) when the population neural response evoked by that stimulus component is enhanced (or suppressed). A recent study using optical imaging to record population neural responses from slices of rodent superior colliculus and dual-site electrical stimulations provided some support of this prediction (Vokoun et al., 2014).

Nonlinear interaction between component responses

The success of the SNL and NNL model in fitting our data suggests that multiplicative interaction between the component responses may play an important role in shaping the response tuning for multiple stimuli. Because the product of component responses depends on the component directions, multiplicative interaction can therefore exert a *tuned* effect on MT responses. Importantly, we have found that the

nonlinear interaction often has a suppressive effect on neuronal responses and the suppression is most prominent when both component responses are strong due to multiplicative nonlinearity. Multiplicative computation has been proposed in a variety of neural models and are thought to be important for sensory processing (Hassenstein and Reichardt, 1956; Barlow and Levick, 1965; Sun and Frost, 1998; Pena and Konishi, 2001; Gabbiani et al., 2004). Possible neural implementation of multiplicative computation has been demonstrated before (Gabbiani et al., 2002), although the neural mechanism underlying multiplication and how multiplication engages suppression are not yet clear.

The SNL model was previously used to fit the responses of multisensory neurons in MSTd (Morgan et al., 2008) and the responses of MT neurons that integrate velocity and disparity gradient cues of 3-D surface orientation (Sanada et al., 2012). To our knowledge, the suppressive effect of the nonlinear interaction has not been reported before. Although the previous studies showed that the SNL model provided better fit than the LWS model, the improvement was modest, explaining an additional 1% of the response variance for the MSTd data (Morgan et al., 2008) and significantly improved data fitting for only 14% of neurons for the MT data (Sanada et al., 2012). In our study, the benefit of the SNL model was the largest under the equal-coherence condition, explaining an additional 6% of the response variance and improving the data fit significantly for about half of the neurons. A possible explanation for the discrepancy between our results and the previous studies is that, in our study, both component responses can be strong when the VA direction is close to a neuron's PD under the equal-coherence condition, giving rise to a large product. It is also possible that the nature of the stimuli may contribute to the discrepancy: our stimuli are consistent with perceptual segmentation (Braddick et al., 2002), whereas in previous studies (Morgan et al., 2008; Sandada et al., 2012) different cues were integrated to form a single percept of heading direction or surface orientation, at least when different cues are congruent.

Tuned normalization and normalization pool

It has been suggested that divisive normalization may be *tuned* to visual stimuli, such that different stimuli may have different contributions to normalization (Carandini et al., 1997; Rust et al., 2006; Ni et al., 2012). In our data, it may seem that the population-averaged response tuning curves are consistent with the idea of *tuned* normalization: At some stimulus directions, the bi-directional response is closer to the stronger component response, and at other stimulus directions, the bi-directional response roughly follows the average of the component responses (e.g. see Fig. 2). However, our model fitting showed that using a weighted sum of the component responses with a *fixed* set of weights plus a nonlinear interaction term could well account for MT responses across different motion directions. To explain our data, it was not necessary to assume a mechanism that gives rise to different response weights at different stimulus directions. Our findings of fixed response weights across motion directions are consistent with the idea that the normalization pool includes neurons tuned to all motion directions and therefore the neural activity of the normalization pool is invariant to the stimulus direction (Carandini et al., 1997; Simoncelli & Heeger, 1998).

Consistent with previous reports (Qian and Andersen, 1994; Ni et al., 2012), we have also found that bi-directional responses of MT neurons are quite variable and can follow the average of the component responses, the stronger (or weaker) component response, or anywhere in-between. This phenomenon can be explained by adding a parameter in the denominator of a normalization model that adjusts the relative contributions of different stimuli to normalization (see Eq. 2 in Ni et al., 2012; also see Carandini et al., 1997). However, we have found that the broad spectrum of the response weights across MT neurons can be explained by varying the sign and magnitude of the coefficient α of the nonlinear interaction term in the denominator (see our Eqs. 8-10). Although our results do not rule out tuned normalization, they suggest another possibility involving nonlinear interactions between component responses to set the response weights differently across neurons.

Depending on the nature of visual stimuli, how multiple stimuli interact within neurons' RFs to influence neuronal responses may be more complicated than documented in this study. Krekelberg and van Wezel (2013) recorded responses of MT neurons using random-dot stimuli moving in opposite directions and found that the *speed tuning curves* of many MT neurons elicited by the bi-directional stimuli cannot be accounted for by a normalization model. It is possible that the interaction between stimulus components across motion speeds may differ from that across motion directions. Other factors such as adaptation (Patterson et al., 2014), and whether multiple stimuli give rise to integrated or segmented perception (Stoner and Albright, 1992) may also influence the relationship between the responses elicited by multiple stimuli and the stimulus components. Understanding the neural computation underlying how multiple stimuli interact across a wide range of stimulus and perceptual conditions is likely to provide important insights into the operations of cortical circuits and sensory processing.

References

- Albright TD.** Direction and Orientation Selectivity of Neurons in Visual Area MT of the Macaque. *J Neurophysiol* 52: 1106-1130, 1984.
- Born RT, and Bradley DC.** Structure and function of visual area MT. *Annu Rev Neurosci* 28: 157-189, 2005.
- Braddick OJ, Wishart KA, and Curran W.** Directional performance in motion transparency. *Vision Res* 42: 1237-1248, 2002.
- Britten KH, and Heuer HW.** Spatial summation in the receptive fields of MT neurons. *J Neurosci* 19: 5074-5084, 1999.
- Britten KH, Shadlen MN, Newsome WT, and Movshon JA.** The Analysis of Visual Motion: a Comparison of Neuronal and Psychophysical Performance. *J Neurosci* 12: 4745-4765, 1992.
- Britten KH, Shadlen MN, Newsome WT, and Movshon JA.** Responses of Neurons in Macaque MT to Stochastic Motion Signals. *Visual Neurosci* 10: 1157-1169, 1993.
- Busse L, Wade AR, and Carandini M.** Representation of Concurrent Stimuli by Population Activity in Visual Cortex. *Neuron* 64: 931-942, 2009.

- Carandini M, and Heeger DJ.** Normalization as a canonical neural computation. *Nat Rev Neurosci* 13: 51-62, 2012.
- Carandini M, Heeger DJ, and Movshon JA.** Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17: 8621-8644, 1997.
- Cheng K, Hasegawa T, Saleem KS, and Tanaka K.** Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *J Neurophysiol* 71: 2269-2280, 1994.
- Churchland AK, Huang X, and Lisberger SG.** Responses of neurons in the medial superior temporal visual area to apparent motion stimuli in macaque monkeys. *J Neurophysiol* 97: 272-282, 2007.
- Draper NR, Smith HS.** Extra sums of squares and tests for several parameters being zero. In: *Applied Regression Analysis* (3rd ed.), New York: Wiley, p. 149-165, 1998.
- Fetsch CR, Pouget A, DeAngelis GC, and Angelaki DE.** Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci* 15: 146-154, 2011.
- Heeger DJ.** Nonlinear model of neural responses in cat visual cortex. In: *Computational Models of Visual Processing*, edited by. Landy MS, Movshon, JA. Cambridge, MA: MIT Press, p. 119-133, 1991.
- Heeger DJ.** Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9: 181-197, 1992.
- Heeger DJ, Simoncelli EP, and Movshon JA.** Computational models of cortical visual processing. *P Natl Acad Sci USA* 93: 623-627, 1996.
- Heuer HW, and Britten KH.** Contrast dependence of response normalization in area MT of the rhesus macaque. *J Neurophysiol* 88: 3398-3408, 2002.
- Huang X, and Lisberger SG.** Circuit mechanisms revealed by spike-timing correlations in macaque area MT. *J Neurophysiol* 109: 851-866, 2013.
- Judge SJ, Richmond BJ, and Chu FC.** Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res* 20: 535-538, 1980.
- Kukkonen H, Rovamo J, Tiippana K, and Nasanen R.** Michelson contrast, RMS contrast and energy of various spatial stimuli at threshold. *Vision Res* 33: 1431-1436, 1993.
- Lisberger SG.** Visual Guidance of Smooth-Pursuit Eye Movements: Sensation, Action, and What Happens in Between. *Neuron* 66: 477-491, 2010.
- Lisberger SG, and Ferrera VP.** Vector averaging for smooth pursuit eye movements initiated by two moving targets in monkeys. *J Neurosci* 17: 7490-7502, 1997.
- MacEvoy SP, Tucker TR, and Fitzpatrick D.** A precise form of divisive suppression supports population coding in the primary visual cortex. *Nat Neurosci* 12: 637-645, 2009.
- Maunsell JHR, and Vanessen DC.** Functional-Properties of Neurons in Middle Temporal Visual Area of the Macaque Monkey .1. Selectivity for Stimulus Direction, Speed, and Orientation. *J Neurophysiol* 49: 1127-1147, 1983.
- Morgan ML, DeAngelis GC, and Angelaki DE.** Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* 59: 662-673, 2008.

- Moulden B, Kingdom F, and Gatley LF.** The Standard-Deviation of Luminance as a Metric for Contrast in Random-Dot Images. *Perception* 19: 79-101, 1990.
- Newsome WT, and Pare EB.** A Selective Impairment of Motion Perception Following Lesions of the Middle Temporal Visual Area (Mt). *J Neurosci* 8: 2201-2211, 1988.
- Ni AM, Ray S, and Maunsell JHR.** Tuned Normalization Explains the Size of Attention Modulations. *Neuron* 73: 803-813, 2012.
- Niu YQ, and Lisberger SG.** Sensory versus motor loci for integration of multiple motion signals in smooth pursuit eye movements and human motion perception. *J Neurophysiol* 106: 741-753, 2011.
- Nover H, Anderson CH, and DeAngelis GC.** A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J Neurosci* 25: 10049-10060, 2005.
- Ohshiro T, Angelaki DE, and DeAngelis GC.** A normalization model of multisensory integration. *Nat Neurosci* 14: 775-782, 2011.
- Peli E.** Contrast in Complex Images. *J Opt Soc Am A* 7: 2032-2040, 1990.
- Qian N, and Andersen RA.** Transparent Motion Perception as Detection of Unbalanced Motion Signals .2. Physiology. *J Neurosci* 14: 7367-7380, 1994.
- Recanzone GH, Wurtz RH, and Schwarz U.** Responses of MT and MST neurons to one and two moving objects in the receptive field. *J Neurophysiol* 78: 2904-2915, 1997.
- Ramachandran R, and Lisberger SG.** Normal performance and expression of learning in the vestibulo-ocular reflex (VOR) at high frequencies. *J Neurophysiol* 93: 2028-2038, 2005.
- Rust NC, Mante V, Simoncelli EP, and Movshon JA.** How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9: 1421-1431, 2006.
- Sanada TM, Nguyenkim JD, and Deangelis GC.** Representation of 3-D surface orientation by velocity and disparity gradient cues in area MT. *J Neurophysiol* 107: 2109-2122, 2012.
- Sclar G, Maunsell JH, and Lennie P.** Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res* 30: 1-10, 1990.
- Simoncelli EP, and Heeger DJ.** A model of neuronal responses in visual area MT. *Vision Res* 38: 743-761, 1998.
- Snowden RJ, Treue S, Erickson RG, and Andersen RA.** The Response of Area Mt and V1 Neurons to Transparent Motion. *J Neurosci* 11: 2768-2785, 1991.

Chapter 3

Distributed and dynamic neural encoding of multiple motion directions of transparently moving stimuli in macaque cortical area MT

This chapter has been published as:

Xiao J and Huang X. Distributed and dynamic neural encoding of multiple motion directions of transparently moving stimuli in macaque cortical area MT. *J Neuroscience*, 35(49), 16180-16198

Abstract

Segmenting visual scenes into distinct objects and surfaces is a fundamental visual function. To better understand the underlying neural mechanism, we investigated how neurons in the middle-temporal cortex (MT) of macaque monkeys represent overlapping random-dot stimuli moving transparently in slightly different directions. It has been shown that the neuronal response elicited by two stimuli roughly follows the average of the responses elicited by the constituent stimulus components presented alone. In this scheme of response pooling, the ability to segment two simultaneously presented motion directions is limited by the width of the tuning curve to motion in a single direction. We found that, although the population-averaged neuronal tuning showed response-averaging, subgroups of neurons showed distinct patterns of response tuning and were capable of representing component directions that were separated by a small angle, less than the tuning width to unidirectional stimuli. One group of neurons preferentially represented the component direction at a specific side of the bi-directional stimuli, weighting one stimulus component more strongly than the other. Another group of neurons pooled the component responses

nonlinearly and showed two separate peaks in their tuning curves, even when the average of the component responses was unimodal. We also showed for the first time that the direction tuning of MT neurons evolved from initially representing the vector-averaged direction of slightly different stimuli to gradually representing the component directions. Our results revealed important neural processes underlying image segmentation and suggest that information about slightly different stimulus components is computed dynamically and distributed across neurons.

3.1 Introduction

Natural scenes often contain multiple entities. The ability to segregate visual scenes into distinct objects and surfaces, referred to as image segmentation, is fundamental to vision. Although a great deal has been learned about the neural mechanisms underlying image segmentation, how the visual system segments two entities that differ only slightly remains an important, open question.

Visual motion provides a potent cue for image segmentation. When two stimuli overlap in space and move in different directions, the primate visual system can segment them into distinct, transparent surfaces based on visual motion cues alone. The extent of "separation" between two transparently moving stimuli can be conveniently controlled by an angular difference of motion directions. In primates, neurons in extrastriate middle-temporal (MT) cortex are direction selective (Maunsell and van Essen, 1983). Area MT is also important for image segmentation pertinent to visual motion signals (Britten, 1999, 2003; Born and Bradley, 2005; Allman et al., 1985; Snowden et al., 1991; Stoner and Albright, 1992; Qian and Andersen, 1994; Born et al., 2000; Huang et al., 2007, 2008).

MT neurons are broadly tuned to motion in a single direction, with a mean tuning width of $\sim 100^\circ$ (Albright, 1984). Broad direction-tuning can be useful for integrating motion signals (Braddick, 1993; Rust

et al., 2006), but it can also limit the ability to distinguish two "slightly different" directions that are separated by an angle less than the tuning width to unidirectional stimuli. Previous studies have examined neural mechanisms underlying transparent motion (Snowden et al., 1991; Qian and Andersen, 1994; Treue et al., 2000; Rosenberg et al., 2008; Krekelberg and van Wezel, 2013; McDonald et al., 2014). However, it remains unclear how MT neurons encode visual stimuli moving transparently in slightly different directions. It has been shown that the response of an MT neuron elicited by two stimuli presented simultaneously tends to follow the average of the responses elicited by the stimulus components presented alone (Qian and Andersen, 1994; Recanzone et al., 1997). Given the broad tuning of MT neurons to unidirectional stimuli, averaging the component responses elicited by two slightly different directions would give rise to a unimodal tuning curve to the bi-directional stimuli. The response peak of the tuning curve is reached when the vector-averaged direction is aligned with a neuron's preferred direction. Indeed it has been shown that, when visual stimuli moving transparently in two directions separated by less than 90° , the population-averaged tuning curve of MT neurons contains only a single response peak (Treue et al., 2000; McDonald et al., 2014). Such a scheme of response averaging would make the segmentation of slightly different directions challenging. In contrast, humans can segment transparently moving stimuli separated by an angle much smaller than the tuning width of MT neurons (Braddick et al., 2002). We have also reported in an abstract form that macaque monkeys have similar ability as humans to segment transparently moving stimuli (Gaudio and Huang, 2011).

We hypothesized that, although on average MT neurons appear to perform a linear response averaging, different subgroups of neurons may be informative about slightly different component directions by selectively pooling the response elicited by one of the stimulus components, and by performing nonlinear operations that emphasize the differences between the stimulus components. We trained two monkeys to perform either a fixation task or a perceptual discrimination task and presented overlapping random-dot stimuli moving transparently in two different directions. We characterized the tuning curves of MT neurons

to the bi-directional stimuli, and the time course of the response tuning. Our results confirmed the hypotheses described above and further showed that the neural representation of component directions developed gradually over time. Our findings revealed important neural processes underlying image segmentation, and provide new insights into how the visual system segments two stimuli even when the stimulus separation is smaller than the tuning width to a single stimulus.

3.2 Materials and Methods

Two adult male rhesus monkeys (*Macaca mulatta*) were used in the neurophysiological experiments. Experimental protocols were approved by the local *IACUC* and followed the NIH *Guide for the Care and Use of Laboratory Animals*. Procedures for surgical preparation and electrophysiological recording were routine and similar to those described previously (Huang et al., [2008](#); Huang and Lisberger, 2009). During sterile surgery with the animal under isoflurane anesthesia, a head post and a recording cylinder were implanted to allow recording from neurons in cortical area MT. Eye position was monitored using a video-based eye tracker (EyeLink, SR Research) at a rate of 1000 Hz.

We used tungsten electrodes (1~3M Ω , FHC) for electrophysiological recordings from neurons in area MT. We identified area MT by its characteristically large portion of directionally selective neurons, small receptive fields (RFs) relative to those of neighboring medial superior temporal cortex (area MST), and its location at the posterior bank of the superior temporal sulcus. Electrical signals were amplified and single units were identified with a real-time template matching system and an offline spike sorter (Plexon).

Visual stimuli and experimental procedure

Stimulus presentation, the behavioral paradigm, and data acquisition were controlled by a real-time data acquisition program (<https://sites.google.com/a/srscicomp.com/maestro/>). Visual stimuli were presented on a 25" CRT monitor at a viewing distance of 63 cm. Monitor resolution was 1024×768 pixels

and the refresh rate was 100 Hz. Visual stimuli were generated by a Linux workstation using an OpenGL application that communicated with an experimental control computer. The output of the video monitor was measured with a photometer (LS-110, Minolta) and was gamma corrected.

Visual stimuli were achromatic random-dot patterns, presented within a stationary, circular aperture that was 7.5° across. Each dot was a square of 2 pixels, extending 0.08° on a side. The dot density of a single random-dot pattern was 3.4 dots/deg^2 , and all dots of a random-dot pattern moved in the same direction at the same speed (i.e. had a motion coherence of 100%). The luminance levels of the dots and the background were 15.3 and 1.9 cd/m^2 , respectively. The bi-directional stimuli contained two overlapping random-dot patterns translating in different directions. Each random-dot pattern is referred to as a "stimulus component". In the main experiment, the angle separation between the two component directions was 60° . In a subset of the experiments, four direction separations (DSs) of 45° , 60° , 90° and 135° were randomly interleaved. At each DS, we varied the vector-averaged (VA) direction of the bi-directional stimuli across 360° to characterize the response tuning, typically in an even step of 15° . In some experiments, we used a fine sampling step of 10° within $\pm 90^\circ$ of the recorded neuron's preferred direction (PD) and a step of 45° beyond.

We also used plaid stimuli consisting of superimposed sinusoidal gratings to characterize the pattern- and component-direction selectivity of MT neurons. The plaids and gratings were presented within a circular aperture that was 7.5° across. The component gratings were separated by 135° in orientation. Grating was presented at 50% contrast, with a mean luminance of 15.3 cd/m^2 . The spatial frequency was 0.8 cycles/deg and the temporal frequency was between 4 to 16 cycles/s. The plaids and grating were sampled in a step of 22.5° or 15° . Gratings were first turned on and stationary for 200 ms, and then drifted for 500 ms.

In each experiment, we first characterized the direction selectivity of a neuron by interleaving trials of $30 \times 27^\circ$ random-dot patches moving at $10^\circ/\text{s}$ in different directions at 45° steps. Directional tuning and the PD of the neuron was evaluated on-line using matlab (MathWorks). We next characterized the speed tuning of the neuron using random-dot patches moving at different speeds of 1, 2, 4, 8, 16, 32, 64 or $128^\circ/\text{s}$ in the PD. The speed tuning curve was fitted using a cubic spline and the speed that gave rise to the highest firing rate in the fitted tuning curve was taken as the preferred speed (PS) of the neuron. The mean PS of our neuron population was $24^\circ/\text{s}$ (std= $16^\circ/\text{s}$). We then mapped the RF of the neuron by recording responses to a series of $5 \times 5^\circ$ patches of random dots that moved in the PD and at the preferred speed of the neuron. The location of the patch was varied randomly to tile the screen in 5° steps without overlap and to cover an area of $35 \times 25^\circ$. The raw map of the RF was interpolated, and the location yielding the highest firing rate was taken as the center of the RF. In the following experiments, testing stimuli were centered on the RF. The mean RF size of our neuron population was 12.8° in diameter (std = 6.6°). The RF size was calculated as the square root of the area where the baseline-subtracted neural activity exceeded the half-maximal response (after Womelsdorf et al., 2006). If the RF only occupied a small portion of a mapping patch, the area of the whole patch would be considered as part of the RF.

Behavioral paradigms

The experiments were conducted while the monkeys performed either a fixation task or a discrimination task. All visual stimuli were presented in individual *trials* while the animals fixated on a spot of light within a $1.5^\circ \times 1.5^\circ$ window to receive juice rewards. In the fixation paradigm, visual stimuli were first turned on but remained stationary for 200 ms before moving for 1000 ms. This allowed us to separate the neuronal response elicited by stimulus motion from stimulus onset. The animals maintained fixation for another 250 ms after the stimulus offset. Trials containing a single stimulus component of the bi-directional stimuli were randomly interleaved with trials of the bi-directional stimuli. Trials that sampled all tested motion directions once were grouped randomly into “a block of trials”. A new block of trials was

initiated only after the previous block was successfully completed. Each stimulus direction was repeated an average of 10 times (std = 3.6).

In the discrimination paradigm, the monkeys were trained to distinguish the bi-directional stimulus that had a DS of 60° from a unidirectional stimulus. The unidirectional stimulus contained two overlapping random-dot patterns that moved in the *same* direction and had the same dot density as the bi-directional stimulus. We trained two monkeys on two variants of the task. For monkey GE, a bi-directional (or a unidirectional) stimulus was centered on a neuron's RF and the corresponding unidirectional (or bi-directional) stimulus was presented at the other half of the visual field, symmetric to the RF location relative to the fixation spot. Visual stimuli moved for 1500 ms. Following the stimulus offset, two spots of light were presented at the two stimulus centers. Once the fixation spot turned off, the monkey was required to make a saccadic eye movement to the spot of light at the center of the bi-directional stimuli to receive juice rewards. In half of the trials, the bi-directional stimuli were placed on the RFs and in the other half of the trials the unidirectional stimuli were placed on the RFs. All trials were randomly interleaved. In 41 of a total of 48 experiments, visual stimuli started to move as soon as they were turned on. In the remaining 7 experiments, visual stimuli remained stationary for 200 ms before moving.

For monkey BJ, only one stimulus, either bi-directional or unidirectional, was centered on the RF and presented in a given trial. Trials containing bi-directional and unidirectional stimuli were randomly interleaved. In all 37 experiments, visual stimuli were turned on and remained stationary for 200 ms and then moved for 1000 ms. Following the stimulus offset, two reporting targets were turned on. The monkey was required to make a saccadic eye movement to the target located at the right (or left) side of the fixation spot when a bi-directional (or unidirectional) stimulus was presented in a given trial to receive juice rewards.

Analysis of response tuning

We calculated the firing rate for each unidirectional stimulus and each VA direction of the bi-directional stimuli based on the spike count during the 1000-ms motion interval and averaged the response across repeated trials. Based on the experiments in which the motion onset was separated from the stimulus onset in time, we found that the transient neuronal response elicited by the stimulus onset lasted on average less than 150 ms. In a subset of the experiments of the discrimination paradigm, motion onset was not separated from the stimulus onset and the stimulus duration was 1500 ms. For these experiments, we calculated the firing rate based on the response from 150 to 1150 ms after the stimulus onset.

We constructed the response tuning curves to unidirectional stimuli and to bi-directional stimuli and fitted the raw direction tuning curves using cubic splines at a resolution of 1° . For each VA direction, we determined the responses elicited by the bi-directional 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 bi-directional 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 bi-directional response and averaged the aligned, normalized tuning curves across cells.

When the sampling step of the VA direction in a block of trials was 15° , the unidirectional stimuli sampled all possible component directions across different VA directions. However, when a neuron was tested with the sampling steps of a mixture of 10° and 45° , the unidirectional stimuli did not fully sample the component directions. For those neurons, we first used a cubic-spline to fit the neuronal responses elicited by the unidirectional stimuli and then resampled the tuning curve to obtain the responses elicited by the component directions.

We fitted the response tuning curves elicited by the bi-directional stimuli using a linear weighted summation (LWS) model (Eq. 1), and a summation plus nonlinear interaction (SNL) model (Eq. 2) (Sanada et al., 2012; Xiao et al., 2014) by minimizing the sum of squared error.

$$R_{pred} = w_1 R_1 + w_2 R_2 + C , \quad (1)$$

$$R_{pred} = w_1 R_1 + w_2 R_2 + b R_1 R_2 , \quad (2)$$

R_{pred} is the model predicted response to the bi-directional stimuli, R_1 and R_2 are the measured component responses elicited by the two unidirectional motion components, and w_1 and w_2 are the weights for the component responses, respectively. C in the LWS model is a constant and b in the SNL model is referred to as the "nonlinear interaction coefficient" that determines the sign and strength of the multiplicative interaction between the component responses.

We also fitted the responses to the bi-directional stimuli using a power-law summation (PWS) model (Eq. 3) after Britten and Heuer (1999). We allowed the response weights w_1 and w_2 for the two stimulus components to be different. The parameter n is a positive exponent and C is a constant.

$$R_{pred} = (w_1 R_1^n + w_2 R_2^n)^{1/n} + C , \quad (3)$$

To evaluate the goodness-of-fit of each model, we computed the percentage of variance (PV) accounted for by a model fit:

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

where SSE is the sum of squared errors between a model fit and the data, and SST is the sum of squared differences between the data and the mean of data (after Morgan et al., 2008). When occasionally SSE exceeded SST and gave rise to a negative PV, we forced the PV to be zero.

For each neuron, we also fitted its responses to the unidirectional stimuli first by a cubic spline at a step of 1° and then fitted the spline-fitted tuning curve using the von Mises function, similar to a circular Gaussian function:

$$R(\theta) = ae^{b[\cos(\theta-\theta_c)-1]} + C, \quad (5)$$

where θ is the motion direction of the unidirectional stimulus, θ_c is the direction where the tuning curve reaches its peak, a and b determine the magnitude and bandwidth of the tuning curve, respectively and C is a positive constant. First fitting the tuning curve using a spline at a fine step allowed the following Gaussian-like fit to accurately capture the response tuning, especially when the tuning curve's bandwidth was narrow. The full width at the half-height of the fitted tuning curve by Equation 5 was taken as the "tuning width" for a given neuron. For a neuron to be selected for further data analyses, we required the goodness of fit (PV) of a neuron's unidirectional tuning curve by Equation 5 to be greater than 90%. A small number of neurons that had irregular or multimodal tuning curve to the unidirectional stimuli were rejected by this criterion (see Results).

To calculate the skewness of a neuron's response tuning curve to a unidirectional stimulus, we used a measure of Pearson's first skewness coefficient, defined as: (mean-mode)/standard deviation. The mean, mode and standard deviation were calculated from the unidirectional tuning curve of each neuron elicited by a random-dot stimulus.

Classification of response tuning curves

We classified neurons into different classes based on the tuning curves in response to the bi-directional stimuli. To determine whether a tuning curve contained only a single peak or at least two peaks, we first located the global peak, a candidate second peak, and a "trough" in between based on the spline-fitted and smoothed tuning curve of the trial-averaged responses. The smoothing was done using a second-

order, seven-point Savitzky-Golay filter. To qualify as a candidate second peak, a response at a given VA direction had to be a local maximum within the neighborhood of $\pm 10^\circ$ along the spline-fitted tuning curve. A candidate trough was determined as the minimum between the global peak and a candidate second peak and had to be within $\pm 40^\circ$ from VA direction of 0° when the DS of the bi-directional stimuli was 60° . More generally, at a DS that was not greater than 135° , a candidate trough had to be within an angle range of $\pm 2 \times \text{DS} / 3$, centered on VA direction 0° . If the response at the candidate trough was significantly smaller than the global peak and the candidate second peak, the tuning curve was considered as containing two peaks. If these criteria were not met after searching through all candidate second peaks, the tuning curve was considered as having one peak.

We used the bootstrap method (Efron and Tibshirani, 1994) to determine whether a candidate trough was significantly smaller than the global peak or a candidate second peak. Specifically, for each spline-fitted but not smoothed tuning curve determined by a single trial across different VA directions, the responses at the locations of the global peak, a candidate second peak and a trough were taken and the difference between a peak and the trough was calculated. If the difference between a peak and a trough was greater than 1.5 times of the standard error of the mean difference (i.e. 87% of confidence interval), the trough was considered as significantly smaller than the peak. The standard error of the mean difference was estimated by bootstrapping (200 times) the difference between the responses at the peak and trough locations from individual trials. Note that the peak and trough locations were determined by the trial-averaged tuning curve and they were not necessarily peaks and trough for a tuning curve based on a single set of trials.

To determine whether a trial-averaged tuning curve of the responses elicited by the bi-directional stimuli was significantly biased toward one motion component (i.e. showed "side-bias"), we used the SNL model to fit each tuning curve determined by a single trial across different VA directions. Across the repeated trials, we obtained trial-by-trial model fits of the response weights for the two components w_1 and

w_2 . We conducted a permutation test (50,000 times) to determine whether w_1 was significantly greater than w_2 ($p < 0.05$), or vice versa.

Time course analysis of response tuning

To compute the time course of the response tuning to the bi-directional stimuli, for each neuron we calculated the tuning curve using trial-averaged firing rates within a 50-ms time window, sliding at a 10-ms step. The responses were normalized by the maximum firing rate across all time windows and averaged across neurons. *For the time course analysis, we excluded the experiments in which the motion onset of the visual stimuli was not separated from the stimulus onset.* When analyzing the timecourse of the response tuning from neurons that showed side-bias, we pooled the results from neurons that showed side-bias to the component directions at the *clockwise side* (C-side) (i.e. *Dir. 2* in the cartoon illustration in Fig. 1) and *counter-clockwise side* (CC-side) (i.e. *Dir. 1*) together. To do so, we first horizontally flipped the tuning curves to the bi-directional and unidirectional stimuli of neurons that showed side-bias to the C-side, along the axis of VA direction 0° , and then averaged the flipped tuning curves together with the tuning curves from the neurons that showed side-bias to the CC-side.

Analysis of pattern and component direction selectivity

We used the methods of Movshon et al. (1985) and Smith et al. (2005) to quantify the pattern- and component-direction selectivity of MT neurons. The pattern prediction was determined by the responses to gratings drifting in the pattern directions and the component prediction was the sum of the responses elicited by component gratings. We calculated the partial correlations R_{pp} and R_{pc} for the pattern and component predictions, respectively (Eqs. 6, 7).

$$R_{pp} = \frac{(r_p - r_c r_{pc})}{\sqrt{(1 - r_c^2)(1 - r_{pc}^2)}} \quad , \quad (6)$$

$$R_{pc} = \frac{(r_c - r_p r_{pc})}{\sqrt{(1 - r_p^2)(1 - r_{pc}^2)}} \quad , \quad (7)$$

in which r_p and r_c are the correlations between the neuronal responses to plaid and the pattern and component prediction respectively, and r_{pc} is the correlation between the two predictions. We converted each value of R_{pp} and R_{pc} to a Z-score designated as Z_p and Z_c respectively using Fisher's r -to- Z transformation (after Smith et al., 2005). For a neuron to be judged as pattern-direction selective, Z_p had to exceed Z_c a value of 1.28 (or zero if Z_c was negative), equivalent to a probability of 0.9, and vice versa for a neuron to be judged as component-direction selective. Otherwise, the cell was considered unclassified.

3.3 Results

To gain a better understanding of the fundamental neural process of image segmentation, we asked the question of how simultaneously presented and slightly different motion directions are represented by neurons in extrastriate area MT. To address this question, we first presented overlapping random-dot stimuli moving in two directions separated by 60° and characterized the tuning curves of MT neurons in response to the bi-directional stimuli. We next used bi-directional random-dot stimuli that had different DSs of 45° , 60° , 90° and 135° to determine whether the tuning curves showed a consistent trend across DSs. We also characterized the time courses of the tuning curves in response to the bi-directional stimuli. Finally, we compared the tuning properties of MT neurons in response to our random-dot stimuli with the pattern- and component-direction selectivity characterized by plaid stimuli. We recorded from 290 neurons in area MT of two macaque monkeys as they performed either a simple fixation task or a perceptual discrimination

task. A majority of the recorded neurons ($N = 267$, 92%, 131 from monkey GE and 136 from monkey BJ) passed our selection criteria for the direction tuning curve to a unidirectional stimulus and were included in our data set (see Materials and Methods).

MT direction tuning to bi-directional stimuli separated by 60°

We set the DS between two stimulus components to 60° and varied the VA direction of the bi-directional stimuli to characterize the direction tuning curve. In this experiment, the monkeys performed a fixation task. The data sample of this experiment contained 202 neurons (107 from GE and 95 from BJ). The mean eccentricity of the RFs was 7.1° (std = 3.7°). We chose to use a DS of 60° for two reasons. First, at this DS, human subjects can reliably segment the two component directions of our stimuli (Gaudio and Huang, 2012). Second, for a majority of MT neurons, the *average* of the responses elicited by two motion components separated by 60° contains only a single response peak, because the tuning curves of MT neurons elicited by unidirectional stimuli have a mean width of $\sim 100^\circ$ (Albright 1984). The peak response occurs when the VA direction is aligned with a neuron's PD, which imposes a challenge for the neural encoding of two separate component directions.

Figure 1 shows the direction tuning curves of four representative MT neurons. The neuron shown in Figure 1A had a tuning curve elicited by the bi-directional stimuli (R_{12} , shown in red) roughly following the average of the component responses (R_{avg} , shown in gray). However, we also found that many MT neurons showed tuning curves that significantly deviated from the response average. For the second neuron shown in Figure 1B, the response elicited by the bi-directional stimuli was the strongest when the motion component at the C-side of the two directions (i.e. *Dir. 2*) was near the neuron's PD, but *not* when the CC-side component (i.e. *Dir. 1*) was near the PD. We refer to this response bias toward the stimulus component

at a specific side of two motion directions as the “side-bias”. The third neuron had a side-bias toward the CC-side of the two component directions (Fig. 1C). About 40% of the neurons showed the side-bias. Lastly, the fourth neuron showed two separate response peaks reached when either stimulus component moved in a direction near the neuron’s PD, although the average of the component responses only contained a single peak located when the VA direction was at the PD (Fig. 1D). About 20% the neurons showed this type of tuning curve.

We fitted the response tuning curves using a LWS model (see Materials and Methods, Eq. 1) and a SNL model (Eq. 2). Each model has three free parameters. The SNL model accounted for on average, 92.4% of the response variance (std = 12.0%), whereas the LWS model accounted for 90.8% of the response variance (std = 11.8%). The SNL model provided significantly better fit than the LWS model ($N = 202$, one-tailed paired t-test, $p < 10^{-6}$) (Fig. 2A). Figure 2D-F demonstrates this by comparing the SNL and LWS model fits to three of the example neurons shown in Figure 1. We therefore used the SNL model to fit the data in the rest of our analyses.

We obtained the response weights w_1 and w_2 for the two component responses respectively using the SNL model fit (see Eq. 2). When averaged across all neurons in the sample, the mean w_1 and w_2 were 0.64 and 0.63, respectively (std = 0.23 for both), suggesting sublinear summation. The two response weights were not significantly different from each other (pair t-test, $p > 0.8$). The distributions of the two response weights w_1 and w_2 are shown in Figure 2C. However, many individual neurons ($N = 94$, 47%) had significantly different response weights for the two stimulus components (permutation test, $p < 0.05$) (Fig. 2B), consistent with the observation that the tuning curves of many MT neurons showed the side-bias.

To classify the response tuning curves elicited by the bi-directional stimuli, we used an algorithm to determine whether a tuning curve contained a single response peak or two separate peaks based on a bootstrap method. We also classified neurons as “side-biased” if the response weights for the two stimulus components were significantly different, based on a permutation test (see Materials and Methods). Figure

3 shows the response tuning curves averaged across the whole population and subgroups of neurons. The tuning curve averaged across all neurons in the population is very similar to the average of the component responses, albeit being slightly broader (Fig. 3A). The tuning curve contains only a single response peak, located when the VA direction was in the PD. Among 202 neurons, 85 neurons (42%) showed a single response peak and no side-bias. The response tuning curves of these neurons were similar to the average of the component responses (Fig. 3B). We referred to these neurons as "averaging" cells. In contrast, the response tuning curves of other subgroups of neurons appear to be informative about the component directions. Another 79 neurons (39%) showed a single response peak and side-bias to one side of the two component directions (Fig. 3C and D). On average, these neurons were most active when a component direction at a specific side of the two motion directions was near the PD, but not when the other component direction was near the PD. In other words, these neurons showed selectivity not only to the motion direction of a stimulus component, but also to which side the component direction was situated relative to the other component direction.

Another 38 neurons (19%) showed two response peaks (Fig. 3E). The peaks were located when either component direction was near the PD. In other words, these neurons were informative about the direction of a stimulus component, regardless of which side the component direction was situated relative to the other stimulus component. In comparison with the average of the component responses R_{avg} (shown in gray), the two response peaks appear to be shaped by facilitation at the outer flanks of the tuning curve and suppression near VA 0° (Fig. 3E). Consistent with the observation that suppression occurred when the VA direction was near the PD, for neurons that showed two response peaks, the mean value of the nonlinear interaction coefficient b of the SNL model fit (see Eq. 2) was significantly smaller than zero (see Table 1) (one-tailed t-test, $p < 0.001$). This indicates that the multiplicative interaction between the component responses had a suppressive effect on the neuronal response. In contrast, for neurons that showed a single

response peak, with or without a side-bias, the mean values of b (see Table 1) were not significantly different from 0 (student t-test, $p > 0.2$).

Notably, neurons that showed two response peaks in the tuning curves had a narrower tuning width to unidirectional stimuli than other subgroups of neurons (one-tailed two-sample t-test, $p < 0.002$, after Bonferroni correction for multiple comparisons) (see Table 1). The mean tuning width of the side-biased neurons was marginally smaller than that of the averaging neurons (one-tailed two-sample t-test, $p = 0.02$).

The SNL model outperformed the LWS model due to the inclusion of a nonlinear interaction term between the component responses. We asked whether a model involving a different type of response nonlinearity could account for the response tuning curves to the bi-directional stimuli. We fitted our data using a nonlinear summation model (after Britten and Heuer, 1999). The PWS model (see Eq. 3 in Materials and Methods) fitted the data well and accounted for on average 93.2% of the response variance ($N = 201$, $\text{std} = 10.6\%$). The model was unable to fit the response from one neuron. Note that the PWS model has one additional free parameter than the LWS and SNL model. Across the neuron population, the median of the exponent parameter n of the PWS model fit was 1.5. The median n of the neurons that showed two response peaks was 3.7 and significantly greater than that of the averaging neurons and the side-biased neurons (one-tailed Wilcoxon rank-sum test, $p < 0.001$, after Bonferroni correction) (see Table 1). This indicates that neurons showing two response peaks are more likely to perform a soft MAX-like operation (Riesenhuber and Poggio, 1999; Britten and Heuer 1999) than other subgroups of neurons. The median n of the side-biased neurons was also significantly greater than that of the averaging neurons (one-tailed Wilcoxon rank-sum test, $p < 0.01$).

Table 1. Parameters of the model fits and tuning widths of subgroups of neurons

	# of cells and percentage	Nonlinear interaction coefficient b of SNL model fit (mean $\pm$ std)	Exponent n of PWS model fit (median, # of cells)	Tuning width to unidirectional stimuli (mean $\pm$ std)
All cells	202 (100%)	-0.007 ± 0.058	1.5 (N = 201)	$99^\circ \pm 22^\circ$
Averaging cells	85 (42%)	-0.004 ± 0.028	1.2 (N = 84)	$105^\circ \pm 18^\circ$
Side-biased cells	79 (39%)	0.001 ± 0.079	1.7	$99^\circ \pm 19^\circ$
Two-peaked cells	38 (19%)	-0.03 ± 0.052	3.7	$85^\circ \pm 27^\circ$

Comparison of response tuning curves across different angular separations

As characterized above, some MT neurons showed side-bias in the response tuning curves to the bi-directional stimuli separated by 60° . We asked whether the side-bias was consistent across different angular separations. In this experiment, we randomly interleaved experimental trials of four angular separations of 45° , 60° , 90° and 135° and varied their VA directions to characterize the response tuning curves. The monkeys performed a fixation task and our data sample included 96 neurons.

We found that MT neurons showed consistent side-bias across different angular separations (Figure 4). We classified side-biased neurons based on their responses to bi-directional stimuli separated by 60° . For neurons showing side-bias to one side of two component directions at the DS of 60° (Fig. 4A2, B2), their tuning curves tended to bias to the same side at other angular separations (Fig. 4A1, A3, A4 and B1, B3, B4). Figure 4C shows the response weights obtained using the SNL-model fit. We pooled together the response weights of 44 neurons (46% of the data sample) that showed side-bias to the C-side (Fig. 4A) and the CC-side (Fig. 4B) at 60° DS. Although the "biased-side" was only determined by the response tuning

to the DS of 60° , at the DSs of 45° , 90° and 135° , the mean response weight for the component direction at the "biased-side" defined at 60° DS was significantly greater than the response weight for the other stimulus component (one-tailed paired t-test, $p < 0.00012$, after Bonferroni correction) (Fig. 4C).

Relationship between the shapes of tuning curves to bi-directional and unidirectional stimuli

For neurons that showed side-bias in the response tuning to the bi-directional stimuli (i.e. R_{12}) separated by 60° (red curves in Fig. 3C and D), the mean component response averaged across the same sub-population of neurons (i.e. R_{avg} , gray curves in Fig. 3C and D) had the response peak slightly shifted toward the same side as that of R_{12} . We asked whether the side-bias of R_{12} was linked to a shift of peak location in R_{avg} . We found that the peak location of R_{12} of the side-biased neurons in response to the bi-directional stimuli separated by 60° was correlated with that of R_{avg} (Spearman's $\rho = 0.54$, $p < 10^{-6}$, $N = 79$) (Fig. 5A). However, the side-bias in R_{12} cannot be explained simply by the bias in R_{avg} . For some neurons, as shown in the second and fourth quadrants, the peak locations of R_{12} and R_{avg} were at opposite sides of VA direction 0° (Fig. 5A). When we determined the "biased-side" by the response weights of the SNL model fit, the shift of the peak location in R_{12} toward the biased-side (mean = 17°) was significantly greater than the shift of peak location to the same side in R_{avg} (mean = 5°) (one-tailed t-test, $p < 10^{-5}$).

The peak location of R_{avg} is related to the shape, specifically the skewness of the tuning curve to the unidirectional component. The peak location of R_{avg} would shift to the right (or left) side of VA direction 0° , if a neuron's tuning to unidirectional stimuli has a positive (or negative) skewness. Since the peak location of R_{12} was correlated with that of R_{avg} , we asked whether the side-bias in R_{12} of a neuron was linked to the skewness of its tuning curve to unidirectional stimuli. We found that the difference between the response weights for the two stimulus components obtained from the SNL model fit of R_{12} had a weak but significant correlation with the skewness of the tuning curve to unidirectional stimuli (Spearman's $\rho = 0.38$,

$p < 0.001$, $N = 79$) (Fig. 5B). Despite this correlation, the side-bias in the response tuning to the bi-directional stimuli cannot be explained simply by a neuron's tuning property to unidirectional stimuli. For some neurons, as shown in the second and fourth quadrants of Figure 5B1, the skewness of the tuning curve to unidirectional stimuli mismatched the side-bias in R_{12} . Furthermore, for neurons that had a symmetric tuning curve to unidirectional stimuli (e.g. the neuron shown in Fig. 5B3 and Fig. 1B), they nevertheless showed side-bias to either the C-side or CC-side (Fig. 5B1).

Time course of response tuning to bi-directional stimuli

We made a novel finding that the side-bias and two response peaks in MT response tuning to the bi-directional stimuli evolved over time. When the angular separation between the two component directions was small, the tuning curve initially followed the average of the component responses and gradually changed to better represent the constituent component directions. To measure the time course of the response tuning, we characterized the direction tuning curves using a time window of 50 ms, sliding at a step of 10 ms.

Figure 6 shows the time course of the response tuning curve elicited by the bi-directional stimuli separated by 60° . For neurons that showed side-bias (79 out of 202 neurons), the initial response tuning had a symmetric single peak, located near the VA direction 0° (see the black curve in Fig. 6A). Over time, the response peak gradually shifted toward one side of the tuning curve (Fig. 6A, B). In contrast, the average of the component responses elicited by the two component directions (i.e. R_{avg}) had a roughly symmetric single peak throughout the motion response period with only a slight bias (Fig. 6C). Taking the difference between the response tuning curves elicited by the bi-directional stimuli (i.e. R_{12}) and R_{avg} revealed strong facilitation at the biased side and suppression at the other side of the tuning curve (Fig. 6D).

For neurons that showed two response peaks (38 out of 202 neurons), the initial response tuning also had a single, symmetric peak near VA 0° (Fig. 6E, F), similar to the average of the component responses (Fig. 6G). Over time, the response elicited by the bi-directional stimuli split into two separate peaks (Fig. 6E, F), whereas the average of the component responses remained a single peak (Fig. 6G). The difference between the bi-directional response and the average of the component responses indicated response facilitation at the two outer flanks of the tuning curve and response suppression near VA 0° (Fig. 6H). For both groups of neurons that showed the side-bias and two response peaks, the facilitatory and suppressive effects occurred after a delay following the onset of the neuronal response (Fig. 6D, H).

The peak locations in the response tuning curves of the side-biased neurons and, those showing two separate peaks were roughly aligned with the directions of the stimulus components (Fig. 7). For neurons that showed side-bias, the peak location of R_{12} was initially near VA direction 0° and progressively shifted to near -30° over a period of 80 to 100 ms (Fig. 7A, red curve). At VA direction -30° , the two component directions were 0° (i.e. the PD of the neuron) and -60° , respectively. In other words, these neurons responded most strongly when one component direction was near the PD. Following the initial, progressive shift, the peak location fluctuated to some extent, but returned to VA -30° at times. In contrast, the peak location of R_{avg} (Fig. 7A, black curve) had a much smaller bias and the fluctuation started in the very beginning of the neural response.

For neurons that showed two response peaks, the peaks did not separate until 30 to 40 ms after the response onset and it took another 40 to 50 ms for the two peaks to separate to the farthest angles of near VA directions of $\pm 40^\circ$ (Fig. 7B, red and blue curves). Following this initial, farthest separation, the two response peaks returned and remained near VA directions of $\pm 35^\circ$ (i.e. a DS of $\sim 70^\circ$). The overshoot of the peak separation from the veridical 60° separation may be related to the perceptual phenomenon of direction repulsion reported by human subjects (Marshak and Sekuler, 1979). When human subjects viewed visual

stimuli similar to those used in our neurophysiological experiments, they also reported direction repulsion (Gaudio and Huang, 2012).

At other angular separations, the response tuning for neurons with side-bias also shifted gradually from initially following the average of the component responses to later being biased toward one side of the tuning curve. Moreover, the side-bias appeared to develop slightly slower at the smaller angular separation. Figure 8 shows the results of the neurons that displayed side-bias at the DS of 45, 60, 90 and 135°, respectively. At the DS of 45°, the average of the component responses contained a single response peak centered near VA 0° throughout the motion response period (Fig. 8A3). The bi-directional responses initially showed a single response-peak centered near VA 0° (see the gray, black, green and pink curves in Fig. 8A1), but over a period of ~50 ms, the response peak shifted to a side (Fig. 8A1-A2). At the DS of 90°, the average of the component responses showed a broad but roughly symmetric tuning curve (Fig. 8C3). The bi-directional response of the same group of neurons initially showed a broad and symmetric tuning curve peaked near VA 0°. Over a period of ~40 ms, the response peak shifted ~45° in the VA direction and was reached when one component direction was near a neuron's PD (Fig. 8C1-C2). At the DS of 135°, the average of the component responses contained two separate but symmetric peaks (Fig. 8D3). The bi-directional response tuning showed two symmetric peaks in the early response. However, over a period of ~30 ms, one response peak evolved to be higher than the other peak and the whole response tuning curve was biased toward one side (Fig. 8D1-D2). The temporal development of the side-bias was delayed relative to the response onset and the transition occurred slightly earlier at the larger angular separations (Fig. 8A1-D1, A2-D2).

Neuronal response tuning obtained during a perceptual discrimination task

The results reported so far were obtained while the monkeys performed a simple fixation task and viewed the visual stimuli passively. We asked whether MT neurons showed similar patterns of response tuning when monkeys performed a behavioral task to actively discriminate bi-directional stimuli from unidirectional stimuli (see Methods and Materials). The DS of the bi-directional stimuli was set to 60°. For monkey GE, a bi-directional stimulus and a unidirectional stimulus moving in the VA direction were presented simultaneously at locations symmetric to the fixation spot. One of the two stimuli was centered on a neuron's RF. After viewing the moving stimuli for 1.5 seconds, the monkey was required to make a saccadic eye movement to the location of the bi-directional stimulus to receive a juice reward (referred to as Task I) (Fig. 9A). The behavioral performance of monkey GE was on average 83% correct (std = 7.1%) across 48 recording sessions during which 51 neurons were recorded. The corresponding discriminability index d' was 1.9. As in the fixation task, when the monkey performed this task we found that some MT neurons showed side-bias in their response tuning curves (Fig. 10A1, B1), and some neurons showed tuning curves containing two separate peaks (Fig. 10C1) although the average of the component responses had only a single peak.

To perform Task I, however, the monkey may shift its attention back and forth between the RF stimulus and the other stimulus at the opposite side of the visual field. To better control the spatial allocation of attention, we trained a second monkey BJ on a modified task, in which only one stimulus, either bi-directional or unidirectional, was presented in a given experimental trial and centered on the RF. After viewing the moving stimuli for 1 second, the monkey was required to make a saccadic eye movement to one of two reporting targets according to whether the stimulus was bi-directional or unidirectional (referred to as Task II) (Fig. 9B). Across 37 recording sessions during which 44 neurons were recorded, monkey BJ correctly identified the bi-directional stimuli at a rate of 90% (std = 3.8%) and correctly identified the unidirectional stimuli at a rate of 92% (std = 5.9%). Again, we found that some MT neurons showed side-

bias (Fig. 10A2, B2) or two response peaks (Fig. 10C2) in their tuning curves while the monkey performed this task. Since we found similar results across Tasks I and II, we pooled the data to calculate the population-averaged tuning curves. MT neurons showed similar patterns of response tuning curves when the monkeys performed the perceptual discrimination tasks (Fig. 10A3-C3) as when they performed the fixation task (Fig. 3C-E). Notably, a higher percentage of neurons showed two response peaks when the monkeys performed the discrimination tasks (29% of 95 neurons) than when they performed the fixation task (19% of 202 neurons). Because the correct rates of the animals' performance were high and our algorithm classifying tuning curves into different subgroups required equal number of trials at different VA directions, to characterize the tuning curves obtained during the discrimination tasks, we used all the experimental trials, including those that had correct and incorrect behavioral reports.

When the monkeys performed the perceptual discrimination task, the time course of MT response tuning to the bi-directional stimuli also evolved from initially following the average of the component responses to later showing the side-bias or two response peaks, as found using the fixation paradigm. Figure 11 shows the results from monkey BJ while it performed Task II, in which the stimulus motion onset was separated in time from the stimulus onset (see Materials and Methods).

Stimulus discrimination using a classifier of SVM

To evaluate whether the population of recorded neurons contained sufficient information to discriminate a bi-directional stimulus from a unidirectional stimulus, and to discriminate between two bi-directional stimuli that had different angular separations, we used the SVM to classify different visual stimuli (see Materials and Methods). We assumed that, for each neuron in our data set, there was a family of “cloned” neurons that had the same tuning curve but different PDs evenly spanning 360° (Fig. 12A, B). The inclusion of the cloned neurons in the population allowed an unbiased representation of all motion

directions and the conversion of the direction tuning curve of a single neuron (Fig. 12A) to the responses of a population of the cloned neurons elicited by a given stimulus (Fig. 12C).

The classifier was capable of discriminating a bi-directional stimulus of 60° DS from a unidirectional stimulus moving at the same VA direction (Fig. 12D). Based on the responses of all 202 neurons in the data set shown in Figure 3, the discrimination performance of the classifier, measured in d' , was 1.3. The performances of the classifier based on the responses of the two-peaked neurons ($N = 38$) and the side-biased neurons ($N = 79$) were better than the averaging neurons ($N = 85$) (Fig. 12D). The d' based on the responses of the averaging neurons merely reached 1 (Fig. 12D, solid green bar), whereas the d' based on the side-biased neurons and the two-peaked neurons was 1.7 and 2.8, respectively (Fig. 12D, solid blue and red bars), similar to the behavioral performance of the two monkeys. This difference in classification was not due to different pool sizes of three subgroups of neurons. We randomly picked 38 neurons from 85 averaging neurons, and from 79 side-biased neurons and repeated the procedure 100 times. The averaged classification performance based on 38 sub-sampled neurons was similar to that based on all averaging neurons or side-biased neurons (Fig. 12D, open green and blue bars). The mean d' based on 38 sub-sampled averaging neurons was significantly smaller than that based on 38 side-biased neurons (t-test, $N = 100$, $p < 10^{-74}$), and both were significantly smaller than the d' value of 2.8 based on 38 two-peaked neurons (t-test, $p < 10^{-84}$). These results support the idea that the two-peaked neurons and the side-biased neurons carry more information about the bi-directional stimuli of 60° DS than the averaging neurons.

The classifier was also capable of discriminating a bi-directional stimulus of 60° DS from another bi-directional stimulus moving at the same VA direction (Fig. 12E), based on the neuronal responses of 96 neurons shown in Figure 4. As expected, the classification performance increased as the difference between the DSs of two stimuli increased from 15° to 75° . Among the three subgroups of neurons, classification based on the two-peaked neurons gave the best performance. For the most difficult discrimination between

DS 60° and DS 45°, the classification based on all 96 neurons and the averaging neurons was poor and had a d' of 0.51 and 0.33, respectively. In contrast, the classification based on the two-peaked neurons had a d' of 1.4. The d' based on the side-biased neurons was 0.74, which was better than the d' based on the averaging neurons (Fig. 12E).

The two-peaked neurons classified based on the tuning curves to the bi-directional stimuli of 60° DS provided good classification between a unidirectional stimulus and a bi-directional stimulus with various DSs, suggesting that this group of neurons were informative about bi-directional stimuli in general (Fig. 12F). Classification based on the side-biased neurons was better than the averaging neurons (Fig. 12F). Note that, at DS 135°, all side-biased neurons showed two response peaks, which may explain why the d' value based on the side-biased neurons was the largest. At DS 45°, although not all 13 two-peaked neurons classified at DS 60° showed two response peaks, they nevertheless supported reliable discrimination between DS 45° and 0°, giving a d' value of 1.4 (Fig. 12F). The d' based on the side-biased neurons was 0.75, better than the d' of 0.38 based on the averaging neurons (Fig. 12F).

Based on the response tuning curves to DS 45°, 42 of the 96 neurons were classified as the side-biased neurons with a single response peak, and 7 neurons were classified as the two-peaked neurons. When discriminating between DS 45° and 0°, the classifier gave the largest d' of 0.90 based on the 42 side-based neurons. The d' based on the 7 two-peaked neurons was smaller, and had a value of 0.58, which may be caused by the small sample size of the two-peaked neurons at DS 45°. In comparison, the d' based on the 47 averaging neurons was 0.41. These results suggest that, at a DS of 45° or smaller, the side-biased neurons may be important for representing the bi-directional stimuli.

Comparison of direction tuning curves elicited by overlapping random-dot stimuli and plaid stimuli

It has been well established that, when tested with overlapping sinusoidal gratings (i.e. plaid stimuli) drifting in widely different directions, some MT neurons are selective to the pattern-motion direction of the plaid, whereas some other neurons are selective to the directions of the component gratings (Movshon et al., 1985; Rodman and Albright, 1989; Smith et al., 2005; Rust et al., 2006). MacDonald et al. (2014) recently showed that pattern cells in area MT of marmosets represented component directions of transparently moving random-dot stimuli that had a large DS of 120°, whereas component cells tended to represent the VA direction.

We examined the relationship between the pattern- and component-direction selectivity to plaid and the types of response tuning curves elicited by overlapping random-dot stimuli moving in slightly different directions. One hundred and two (102) neurons were tested with both plaid stimuli that had a large DS of 135° and random-dot stimuli that had a DS of 60°. Among them, 99 neurons were tested with component gratings moving at the same speed as the random-dot patterns. The classification of neurons as pattern- or component-direction selective relies on the "pattern prediction" and the "component prediction" to be significantly different from each other (see Materials and Methods), which only holds true when the DS of plaid stimuli is large. We therefore chose to use plaid stimuli that had a DS of 135° rather than 60°. The monkeys performed a fixation task when tested with the plaid stimuli, and performed either a fixation task or the perceptual discrimination tasks as mentioned earlier when tested with the random-dot stimuli. Since we found that the response tuning curves elicited by the random-dot stimuli were qualitatively similar when the monkeys performed the fixation task (60 neurons recorded) and the perceptual discrimination tasks (42 neurons recorded), we pooled the results of these neurons together.

We did not find a significant relationship between the properties of the response tuning curves to the random-dot stimuli and the pattern/component selectivity to the plaid stimuli. Neurons that showed each

of the three types of tuning curves: side-bias, two response peaks and response-averaging to random-dot stimuli that had a DS of 60° could be either pattern-selective or component-selective to our plaid stimuli. Across the population of 102 neurons, the "averaging" neurons included a higher percentage of pattern cells than did the side-biased and two-peaked neurons. Whereas the two-peaked neurons included a slightly higher percentage of component cells than did the side-biased and averaging neurons (Fig. 13 and Table 2). Examining the difference between Z-transformed pattern correlation (Z_p) and component correlation (Z_c) (see Methods and Materials), revealed that the median value of $Z_p - Z_c$ was positive and the largest for the averaging neurons, indicating that these neurons tended to be more pattern-selective. Whereas the median value of $Z_p - Z_c$ was negative for the two-peaked neurons, indicating that these neurons tended to be more component-selective (Table 2). However, the median values of $Z_p - Z_c$ were not significantly different between any two subgroups of averaging, side-biased and two-peaked neurons (Wilcoxon rank-sum test, $p > 0.18$). We found the same result when we constrained our data set to the 99 neurons that were tested with the plaid and random-dot stimuli moving at the same speed.

We also tested 46 neurons with both the plaid stimuli and random-dot stimuli that had the same DS of 135° . We did not find a significant relationship between the tuning properties to the random-dot stimuli and the pattern/component selectivity, even when the DS was matched. In response to the random-dot stimuli, most of these neurons (45/46) showed two response peaks because the DS was large. Twenty-three (23) of the two-peaked neurons followed the average of the component responses and 22 neurons showed side-bias. Only one of the 46 neurons showed a single response-peak. This cell was classified as an averaging neuron. Based on the responses to the plaid stimuli, the median values of $Z_p - Z_c$ were not significantly different between the two-peaked averaging neurons ($N = 23$) and the two-peaked side-biased neurons ($N=22$) (Wilcoxon rank-sum test, $p = 0.59$) (see Table 2).

Smith, Majaj and Movshon (2005) have shown that it takes longer for the pattern-direction selectivity of the pattern cells to develop than the component-direction selectivity of the component cells.

Our data obtained using the plaid stimuli confirmed this finding. Among 102 neurons in this data set, the pattern-selectivity of 33 pattern cells emerged later than the component-selectivity of 25 component cells. A neuron showing a longer build up of the pattern-selectivity (consistent with motion integration) depended on whether or not the neuron was a pattern-cell, regardless of the neuron's tuning property to the random-dot stimuli. All three types of tuning curves characterized based on the response to the random-dot stimuli: two-peaked, side-biased and averaging neurons showed a longer build-up of pattern-selectivity with the plaid stimuli, as long as they were also pattern cells (results not shown). Interestingly, for neurons that were pattern cells and also showed two response peaks or side-bias to the bi-directional random-dot stimuli with a small DS, the temporal evolution of the response tuning to bi-directional stimuli was stimulus-dependent and switched from a gradual buildup of segmentation (as shown in Fig. 6) to a buildup of integration when the visual stimuli changed from random dots to plaids. This adaptive change of the temporal property of response tuning is akin to the stimulus-dependent change of surround antagonism and integration found in area MT (Huang et al., 2007; 2008). Future study is needed to understand the neural mechanism underlying such adaptive change of direction tuning over time.

Table 2. Relationship between the properties of response tuning to random-dot stimuli (RDS) and plaid stimuli

Plaid (135°) RDS (60°)	Number of neurons	Pattern- Selective	Component- Selective	Unclassified	Zp - Zc (Median)	Zp - Zc (Mean)	Zp - Zc (Std)
All	102	32%	25%	43%	0.25	0.36	2.78
Averaging	30	43%	17%	40%	0.47	0.96	2.57
Side-bias	48	27%	25%	48%	0.14	0.087	2.76
Two-peak	24	29%	33%	38%	-0.24	0.15	3.04
Plaid (135°) RDS (135°)	Number of neurons	Pattern- Selective	Component- Selective	Unclassified	Zp - Zc (Median)	Zp - Zc (Mean)	Zp - Zc (Std)
All	46	39%	26%	35%	0.22	0.46	3.47
Averaging (1-peak)	1	0	0	100%	0.50	0.50	-
Averaging (2-peak)	23	39%	17%	44%	0.23	0.74	2.76
Side-bias (2-peak)	22	41%	36%	23%	0.07	0.16	4.20

3.4 Discussion

We found that many neurons in area MT were capable of representing component directions of transparently moving stimuli, even when the angular difference between two directions was smaller than the tuning width to unidirectional stimuli. We also discovered that the neural representation of component directions developed over time. The tuning curves of some neurons initially followed the average of the component responses, and later showed side-bias or two response peaks. The early neuronal responses better represented the VA direction of slightly different stimuli, whereas the late responses were informative about the component directions.

Previous studies show 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 (van Wezel et al., 1996; Qian and Andersen, 1994; Recanzone et al., 1997; Britten and Heuer, 1999; Zoccolant et al., 2005). Consistent with the model of response normalization (Carandini and Heeger, 2011), the response weight is greater for the stimulus component that has a stronger signal strength, as found in area V1 (Busse et al., 2009; MacEvoy et al., 2009), MT (Xiao et al., 2014) and MST (Morgan et al., 2008). In the current study, the stimulus components had the same signal strength. Although, as expected, the population-averaged response weights for the two stimulus components with a DS of 60° were identical, we found for many neurons, the response weight for one stimulus component was significantly greater than the other. This *unequal pooling* of the component responses allows a neuron to selectively represent the direction at a specific side of two motion vectors. To the best of our knowledge, this finding establishes for the first time the selectivity of MT neurons for the side relationship of two motion directions.

Ni et al. (2012) suggest that response normalization is tuned, meaning that different visual stimuli contribute differently to normalization (also see Carandini et al., 1997; Rust et al., 2006). Tuned normalization can explain why for some MT neurons the response to the bi-directional stimulus follows the average of the component responses, whereas for others the response follows the stronger (or weaker) component response or anywhere in between. At its current form, however, tuned normalization cannot explain the side-bias found in our study. Taking the neuron in Figure 1B as an example, this neuron showed response averaging when the component direction at the CC-side was closer to the PD, and winner-take-all when the C-side component was closer to the PD. The two component stimuli contributed equally to normalization at one side of the tuning curve but contributed differently at the other side of the tuning curve. The extent of tuned normalization itself is tuned to the visual stimuli.

We speculate that recurrent interactions among MT neurons may be involved in shaping the side-bias, because the side-bias developed over time and emerged later in the tuning curve. Feedback connections from higher-order areas may also be involved. Area MT receives feedforward inputs from direction-selective neurons in V1 (Movshon and Newsome, 1996). If feedforward connections between V1 neurons and a target MT neuron have an asymmetric distribution of the synaptic weights in relation to whether the PD of a V1 neuron is at the C-side or CC-side of the PD of the MT neuron (illustrated in Fig. 14A), the MT neuron's response to the bi-directional stimuli would show side-bias. However, it is unlikely that feedforward connections alone can explain the time course of the side-bias. Alternatively, the side-bias may arise due to asymmetric recurrent connections between MT neurons (Fig. 14B). It is also possible that a slight asymmetry in the feedforward connections is amplified by recurrent interactions (Fig. 14C). Future studies are needed to understand the neural circuit mechanisms underlying the side-bias, possibly involving an attractor network (Knierim and Zhang, 2012).

Some MT neurons showed two response peaks to the bi-directional stimuli even when the average of the component responses was unimodal. This result suggests that stimulus components interact nonlinearly within the RF. One possible mechanism involves response suppression proportional to the product of the component responses (Xiao et al., 2014). Another possibility involves a soft Max-like operation (Riesenhuber and Poggio, 1999; Lampl et al., 2004), in which the neuronal response elicited by two stimuli is close to the stronger component response. The suppressive mechanism involving multiplicative interaction and the soft Max-like operation may work synergistically to allow neurons to represent slightly-different stimulus components.

Treue et al. (2000) previously investigated how bi-directional random-dot stimuli were represented by neurons in area MT. Recently, McDonald et al. (2014) studied the response tuning of MT neurons in marmosets using similar stimuli. Consistent with these studies, we found that the tuning curve averaged across all neurons in response to the bi-directional stimuli with a small DS roughly followed the average of the component responses, and showed a symmetric, unimodal shape (Fig. 3A). It is unclear whether the individual neurons in these previous studies also showed the side-bias or two response-peaks that deviated from response averaging. Moreover, these previous studies did not examine the time course of the tuning curve. In the study of Treue et al. (2000), the random-dot stimuli moved on a circular path (Schoppmann and Hoffmann, 1976). Although this method of stimulus presentation is efficient for measuring direction tuning, the constant change of stimulus direction could make it difficult to reveal the time course of response tuning to the bi-directional stimuli.

Perceptually, lowering luminance contrast benefits motion integration rather than segmentation (Murakami and Shimojo, 1993). Response normalization in area MT is also contrast dependent (Heuer and

Britten, 2002). The response tuning to the bi-directional random-dot stimuli may become less supportive of segmentation when luminance contrast is reduced. Furthermore, the perceived angle separation between two component directions of random-dot stimuli varies with motion coherence (Gaudio and Huang, 2012). As the coherence level is lowered, the perceived angle shifts from repulsion to attraction, consistent with a change from motion segmentation to integration. The response tuning to the bi-directional stimuli therefore may also depend on the coherence level of the random dots. Future studies are needed to test these hypotheses.

Attention can bias the neuronal response elicited by multiple stimuli in the RF (Reynolds et al., 1999; Ferrera and Lisberger, 1997; Treue and Martinez-Trujillo, 1999; Recanzone and Wurtz, 2000; Li and Basso 2005). Wannig et al. (2007) showed that attention directed to one of two transparently moving surfaces could alter the responses of MT neurons in favor of the direction of the attended surface. In that study, attention was cued to one of two already segregated surfaces. In contrast, the two slightly-different stimuli in our experiments had not been rendered in advance as separate surfaces, and no visual cue was given for attention selection. It is unlikely that our finding of the side-bias was due to an attentional bias. For the side-bias to be caused by attention, attention had to be directed to the stimulus component at a specific side (e.g. the C-side) of the two component directions, across different VA directions. Without an attention cue, such a specific and consistent attention selection is unlikely to occur. The fact that we found similar results when the animals performed a passively-viewing fixation task, and performed two variants of a perceptual discrimination task, also suggests that the side-bias was not caused by an attentional bias. Under our experimental conditions, the visual system may have to solve the problem of segmentation first, at least at a primitive level, before attention can be directed to one of the stimulus components. Although attention selection may occur during the later portion of the stimulus presentation, it is unlikely to be specific to one side of the bi-directional stimuli.

A recent theoretical study shows that having heterogeneous response weights and response nonlinearity in "stimulus mixing" benefits the neural coding of multiple stimuli (Orhan and Ma, 2015). Our findings of the side-biased and two-peaked neurons in area MT provide experimental evidence that the visual system uses these strategies to encode slightly different stimuli. The existence of the side-biased neurons toward the stimulus component at the C- or CC-side, and the two-peaked neurons suggest that information regarding slightly-different moving stimuli is distributed across sub-populations of neurons in area MT. The visual system may take the distributed neural code of multiple stimuli into consideration to fully utilize such information.

Previous studies show that direction tuning curves of MT neurons undergo dynamic changes during the process of *motion integration* related to the solution of the aperture problem (Pack and Born, 2001), and the emergence of pattern-motion selectivity (Smith et al., 2005). Here we show that, during the process of *motion segmentation*, the direction tuning curves of the side-biased and two-peaked neurons evolve over time to better represent the component directions. Our results suggest that segmenting different stimuli is a dynamic process and may involve recurrent interactions within the neuronal network. Taken together, our findings put new constraints on neural models of visual motion processing and have implications for understanding the neural mechanisms underlying image segmentation in general.

References

- Albright TD.** Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol* 52: 1106-1130, 1984.
- Allman J, Miezin F, and McGuinness E.** Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci* 8: 407-430, 1985.
- Born RT, and Bradley DC.** Structure and function of visual area MT. *Annu Rev Neurosci* 28: 157-189, 2005.
- Born RT, Groh JM, Zhao R, and Lukasewycz SJ.** Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. *Neuron* 26: 725-734, 2000.
- Braddick O.** Segmentation versus integration in visual motion processing. *Trends Neurosci* 16: 263-268, 1993.
- Braddick OJ, Wishart KA, and Curran W.** Directional performance in motion transparency. *Vision Res* 42: 1237-1248, 2002.
- Britten KH.** How are moving images segmented? *Curr Biol* 9: R728-730, 1999.
- Britten KH.** The middle temporal area: motion processing and the link to perception. In: *The visual neurosciences*, edited by Chalupa LM, and Werner JS. London: The MIT press, 2003, p. 1203-1216.
- Britten KH, and Heuer HW.** Spatial summation in the receptive fields of MT neurons. *J Neurosci* 19: 5074-5084, 1999.
- Busse L, Wade AR, and Carandini M.** Representation of concurrent stimuli by population activity in visual cortex. *Neuron* 64: 931-942, 2009.
- Carandini M, and Heeger DJ.** Normalization as a canonical neural computation. *Nat Rev Neurosci* 13: 51-62, 2011.
- Efron B, and Tibshirani RJ.** *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall/CRC, 1994.
- Ferrera VP, and Lisberger SG.** Neuronal responses in visual areas MT and MST during smooth pursuit target selection. *J Neurophysiol* 78: 1433-1446, 1997.
- Fetsch CR, Pouget A, DeAngelis GC, and Angelaki DE.** Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci* 15: 146-154, 2011.
- Gaudio JL, and Huang X.** Perceptual segmentation and integration of spatially-overlapping moving features in humans and macaques. Vision Sciences Society Annual Meeting Abstract. *Journal of Vision*:11(11):760. doi: 10.1167/11.11.760, 2011.
- Gaudio JL, and Huang X.** Motion noise changes directional interaction between transparently moving stimuli from repulsion to attraction. *PLoS One* 7: e48649, 2012.
- Huang X, Albright TD, and Stoner GR.** Adaptive surround modulation in cortical area MT. *Neuron* 53: 761-770, 2007.

- Huang X, Albright TD, and Stoner GR.** Stimulus dependency and mechanisms of surround modulation in cortical area MT. *J Neurosci* 28: 13889-13906, 2008.
- Huang X, and Lisberger SG.** Noise correlations in cortical area MT and their potential impact on trial-by-trial variation in the direction and speed of smooth-pursuit eye movements. *J Neurophysiol* 101: 3012-3030, 2009.
- Knierim JJ, and Zhang K.** Attractor dynamics of spatially correlated neural activity in the limbic system. *Annu Rev Neurosci* 35: 267-285, 2012.
- Kouh M, and Poggio T.** A canonical neural circuit for cortical nonlinear operations. *Neural Comput* 20: 1427-1451, 2008.
- Krekelberg B, and van Wezel RJ.** Neural mechanisms of speed perception: transparent motion. *J Neurophysiol* 110: 2007-2018, 2013.
- Lampl I, Ferster D, Poggio T, and Riesenhuber M.** Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J Neurophysiol* 92: 2704-2713, 2004.
- Li X, and Basso MA.** Competitive stimulus interactions within single response fields of superior colliculus neurons. *J Neurosci* 25: 11357-11373, 2005.
- MacEvoy SP, Tucker TR, and Fitzpatrick D.** A precise form of divisive suppression supports population coding in the primary visual cortex. *Nat Neurosci* 12: 637-645, 2009.
- Marshak W, and Sekuler R.** Mutual repulsion between moving visual targets. *Science* 205: 1399-1401, 1979.
- Maunsell JH, and Van Essen DC.** Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol* 49: 1127-1147, 1983.
- McDonald JS, Clifford CW, Solomon SS, Chen SC, and Solomon SG.** Integration and segregation of multiple motion signals by neurons in area MT of primate. *J Neurophysiol* 111: 369-378, 2014.
- Morgan ML, Deangelis GC, and Angelaki DE.** Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* 59: 662-673, 2008.
- Movshon JA, Adelson EA, Gizzi M, and Newsome WT.** The analysis of moving visual patterns. In: *Study Group on Pattern Recognition Mechanisms*, edited by Chagas C, Gattass R, and Gross CG. Vatican City: Pontifica Academia Scientiarum, 1985, p. 117-151.
- Movshon JA, and Newsome WT.** Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J Neurosci* 16: 7733-7741, 1996.
- Orhan AE, and Ma WJ.** Neural population coding of multiple stimuli. *J Neurosci* 35: 3825-3841, 2015.
- Pack CC, and Born RT.** Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature* 409: 1040-1042, 2001.
- Qian N, and Andersen RA.** Transparent motion perception as detection of unbalanced motion signals. II. Physiology. *J Neurosci* 14: 7367-7380, 1994.
- Recanzone GH, and Wurtz RH.** Effects of attention on MT and MST neuronal activity during pursuit initiation. *J Neurophysiol* 83: 777-790, 2000.

- Recanzone GH, Wurtz RH, and Schwarz U.** Responses of MT and MST neurons to one and two moving objects in the receptive field. *J Neurophysiol* 78: 2904-2915, 1997.
- Reynolds JH, Chelazzi L, and Desimone R.** Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19: 1736-1753, 1999.
- Riesenhuber M, and Poggio T.** Hierarchical models of object recognition in cortex. *Nat Neurosci* 2: 1019-1025, 1999.
- Rodman HR, and Albright TD.** Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Exp Brain Res* 75: 53-64, 1989.
- Rosenberg A, Wallisch P, and Bradley DC.** Responses to direction and transparent motion stimuli in area FST of the macaque. *Vis Neurosci* 25: 187-195, 2008.
- Rust NC, Mante V, Simoncelli EP, and Movshon JA.** How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9: 1421-1431, 2006.
- Smith MA, Majaj NJ, and Movshon JA.** Dynamics of motion signaling by neurons in macaque area MT. *Nat Neurosci* 8: 220-228, 2005.
- Snowden RJ, Treue S, Erickson RG, and Andersen RA.** The response of area MT and V1 neurons to transparent motion. *J Neurosci* 11: 2768-2785, 1991.
- Stoner GR, and Albright TD.** Neural correlates of perceptual motion coherence. *Nature* 358: 412-414, 1992.
- Treue S, Hol K, and Rauber HJ.** Seeing multiple directions of motion-physiology and psychophysics. *Nat Neurosci* 3: 270-276, 2000.
- Treue S, and Maunsell JH.** Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382: 539-541, 1996.
- Treue S, and Martinez Trujillo JC.** Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399: 575-579, 1999.
- van Wezel RJ, Lankheet MJ, Verstraten FA, Maree AF, and van de Grind WA.** Responses of complex cells in area 17 of the cat to bi-vectorial transparent motion. *Vision Res* 36: 2805-2813, 1996.
- Wannig A, Rodriguez V, and Freiwald WA.** Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron* 54: 639-651, 2007.
- Xiao J, Niu YQ, Wiesner S, and Huang X.** Normalization of neuronal responses in cortical area MT across signal strengths and motion directions. *J Neurophysiol* 112: 1291-1306, 2014.
- Zoccolan D, Cox DD, and DiCarlo JJ.** Multiple object response normalization in monkey inferotemporal cortex. *J Neurosci* 25: 8150-8164, 2005.

Figures and figure legends

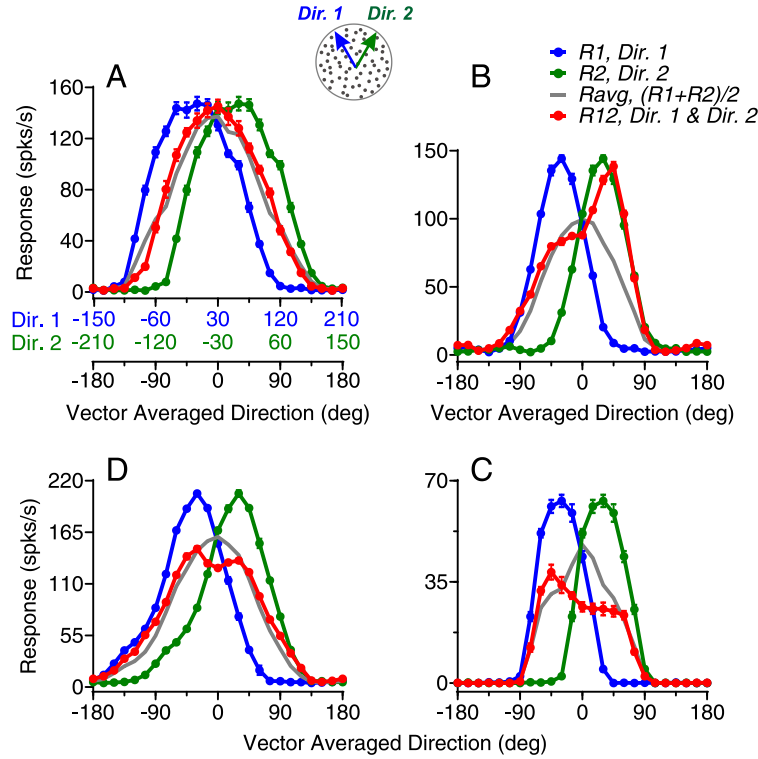


Figure 1. Direction tuning curves of four example neurons elicited by bi-directional stimuli separated by 60° and constituent unidirectional stimulus components. **A.** Cartoon illustration of the visual stimuli and the response tuning curves of an example neuron. Blue and green dots in the cartoon indicate two overlapping, *achromatic* random-dot patterns moving in two directions separated by 60°. Colors were used for illustration purposes only. The component direction shown in blue (*Dir. 1*) moved at the counter-clockwise side (CC-side) of the other component direction shown in green (*Dir. 2*). The abscissas in blue and green show unidirectional components *Dir. 1* and *Dir. 2*, respectively, of the corresponding bi-directional stimuli whose vector-averaged (VA) direction is shown by the black abscissa. The blue and green axes are shifted by 60° relative to each other. VA direction of 0° is aligned with the neuron's preferred direction (PD). For this neuron, the responses elicited by the bi-directional stimuli roughly followed the average of the responses elicited by the stimulus components. **B. C. D.** Response tuning curves of another three example neurons. Error bars indicate SE and they are sometimes smaller than the symbol size (e.g. as in **D**).

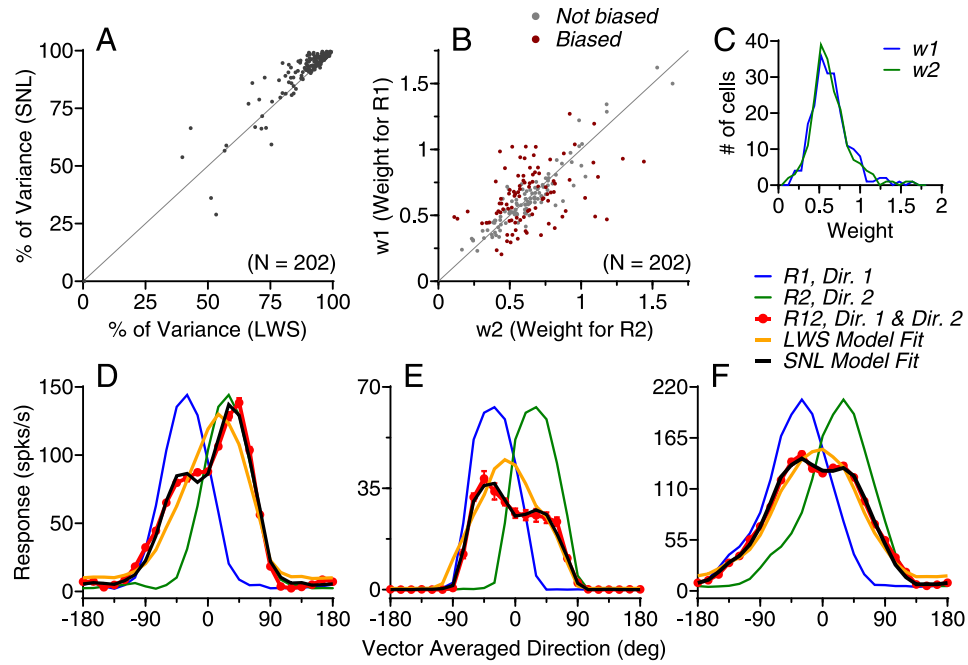


Figure 2. Model fits of the response tuning curves elicited by bi-directional stimuli separated by 60°. **A.** Percentage of the variance (PV) accounted for by the linear weighted summation (LWS) model and the summation plus nonlinear interaction (SNL) model. **B.** Weights for the component responses obtained using the SNL model fit. Each dot in A and B represents data from one neuron. Colored dots in B indicate that the two response weights are significantly different. **C.** Distribution of the response weight for each stimulus component. **D-F.** Direction tuning curves and their model fits of three example neurons. These neurons are the same as those shown in Figure 1B-D.

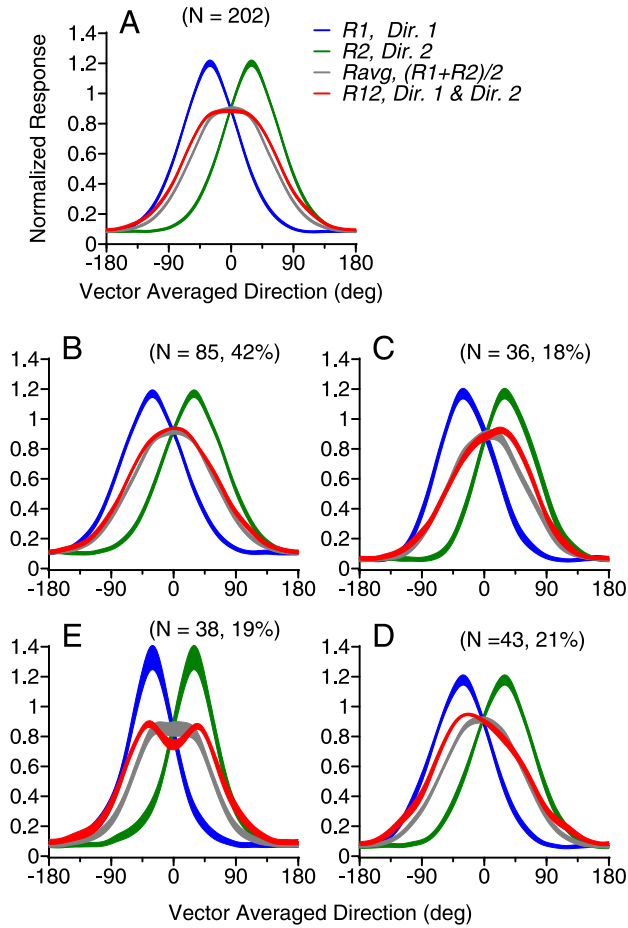


Figure 3. Population-averaged tuning curves elicited by bi-directional stimuli separated by 60° and by constituent unidirectional components. **A.** Tuning curves averaged across all neurons in the sample. **B.** Averaged across neurons that showed a single response peak and no side-bias to the bi-directional stimuli. **C.** Neurons that showed side-bias toward the stimulus component at the clockwise side (C-side) of the two component directions (i.e. *Dir. 2*). **D.** Neurons that showed side-bias toward the CC-side component (i.e. *Dir. 1*). **E.** Neurons that showed two response peaks. The convention of the abscissa is the same as that in Figure 1. Abscissas for component *Dir. 1* and *2* are not plotted for simplicity. The width of each curve indicates SE.

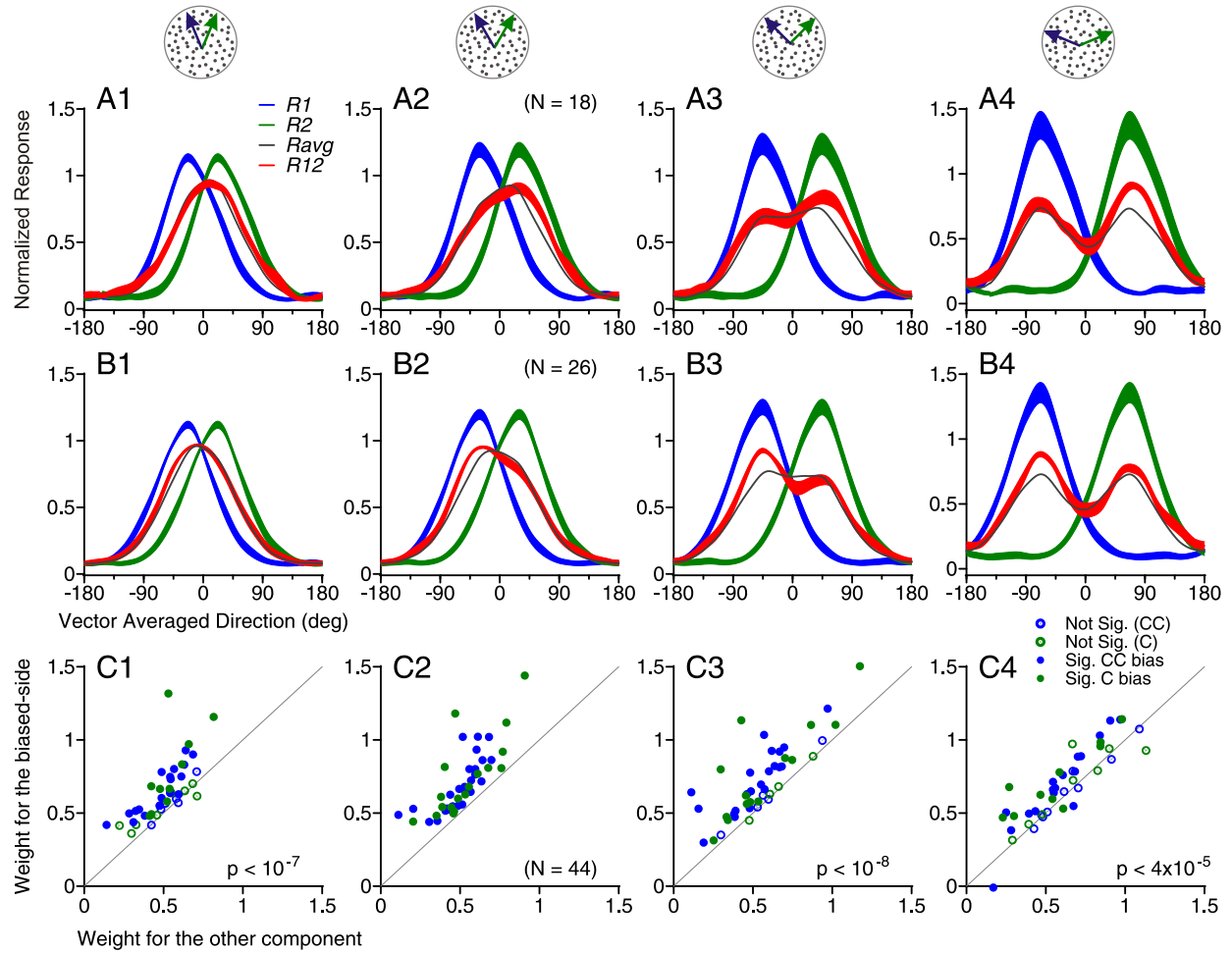


Figure 4. Consistent side-bias of response tuning curves across different angular separations. A. Averaged response tuning-curves across 18 neurons that showed side-bias to the C-side component of the bi-directional stimuli separated by 60°. **B.** Averaged response tuning curves across 26 neurons that showed side-bias to the CC-side component of the bi-directional stimuli separated by 60°. The width of each curve indicates SE. **C.** Response weights obtained using the SNL model fit for the combined 44 neurons that showed side-bias to the bi-directional stimuli separated by 60°. The “biased-side” in the ordinate was determined by the response to 60° DS. Each dot represents data from one neuron. Solid dots indicate that the response weight for one stimulus component was significantly greater than the other component (permutation test, $p < 0.05$). The mean weight for the component response to the “biased-side” defined at 60° DS was significantly greater than the weight to the other stimulus component at angular separations of 45°, 90° and 135°. p values are the results of one-tailed paired t-test. **A1-C1**, 45° DS; **A2-C2**, 60° DS; **A3-C3**, 90° DS; **A4-C4**, 135° DS. Note that we allowed the side-biased neurons to have either a single response peak or two response peaks in this analysis.

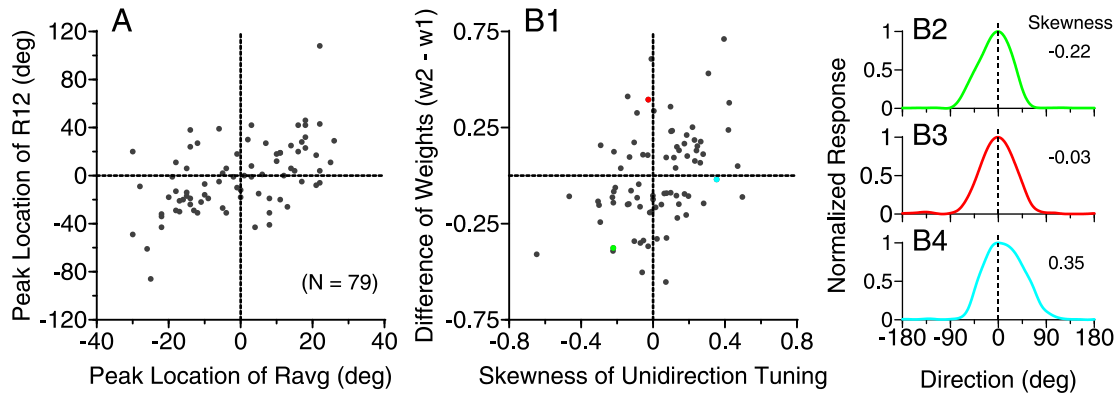


Figure 5. Relationship between the shapes of response tuning curves to bi-directional and unidirectional stimuli. The results were from 79 neurons that showed side-bias in their tuning curves to the bi-directional stimuli separated by 60° . **A.** Relationship between the peak locations of the response tuning to the bi-directional stimuli R_{12} and the average of the responses elicited by individual stimulus components $R_{avg} = (R_1 + R_2)/2$. The peak location is represented as the VA direction of the bi-directional stimuli. **B1.** Relationship between the difference of the response weights for the two stimulus components obtained from the SNL model fit of R_{12} , and the skewness of the unidirection tuning curve (see Materials and Methods). w_2 and w_1 correspond to the response weights for *Dir. 2* and *Dir. 1*, respectively (see Fig. 1 cartoon). **B2-B4.** Response tuning curves to unidirectional stimuli of three neurons that had negative (B2), near zero (B3) and positive skewness (B4). These neurons are marked in B1 using the corresponding colors. The neuron shown in B3 is the same example neuron shown in Figure 1B.

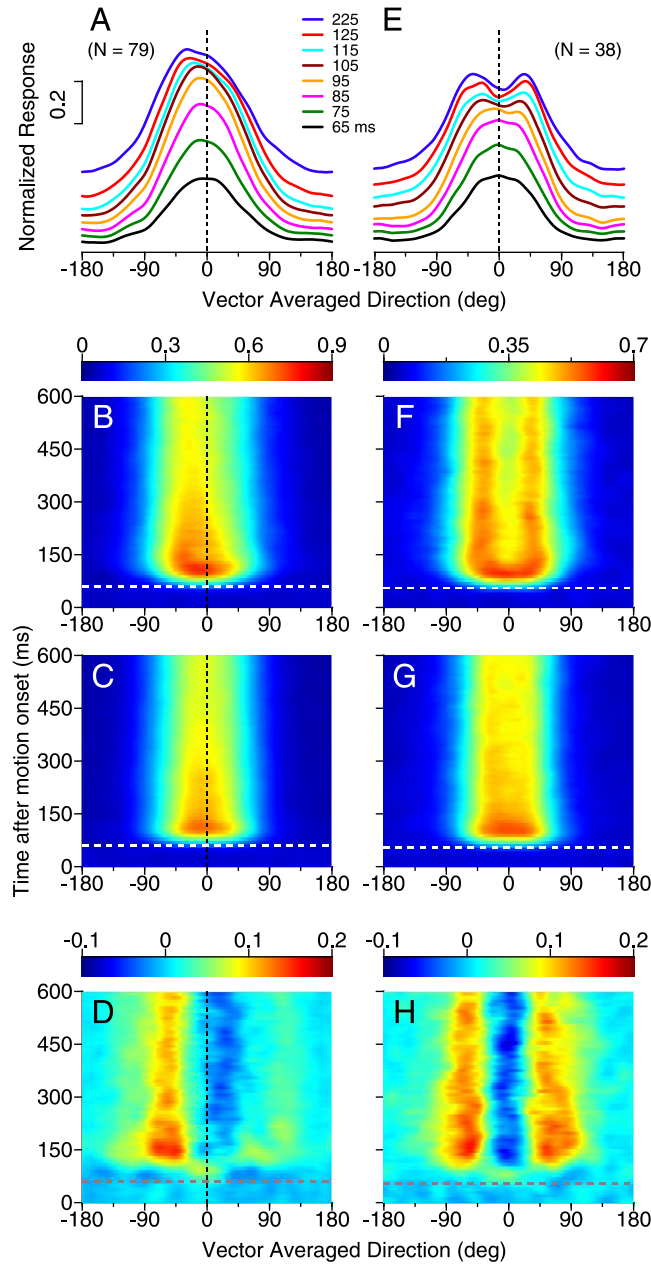


Figure 6. Time course of response tuning to bi-directional stimuli separated by 60°. **A-D.** Responses of neurons showing side-bias to bi-directional stimuli. **E-H.** Responses of neurons showing two peaks to bi-directional stimuli. **A, E.** Subpopulation-averaged response tuning curves to bi-directional stimuli at different time epochs following the stimulus motion onset. The duration of each time epoch is 50 ms. The tuning curves of later epochs are stacked on earlier ones. The number in figure legends indicates the middle point of each time epoch. **B, F.** Time course of the subpopulation-averaged response tuning to bi-directional stimuli (i.e. R_{12}). Ordinates indicate the middle point of each time epoch. **C, G.** Time course of the average

of the component responses (i.e. R_{avg}), averaged across neurons in each subpopulation. **D, H.** Difference between the response to bi-directional stimuli shown in B, F and the average of the component responses shown in C, G, respectively. In **A-C**, We pooled the results from neurons that showed side-bias to the component directions at the C-side (i.e. *Dir. 2*) and CC-side (i.e. *Dir. 1*) together by first flipping the bi-directional and unidirectional tuning curves of neurons that showed the side-bias to *Dir. 2* along the axis of VA direction 0° . The white and gray horizontal lines in B-D and F-H indicate the neuronal response onset to the stimulus motion. In D and H, the deviation from the average of the component responses emerged later than the neuronal response onset.

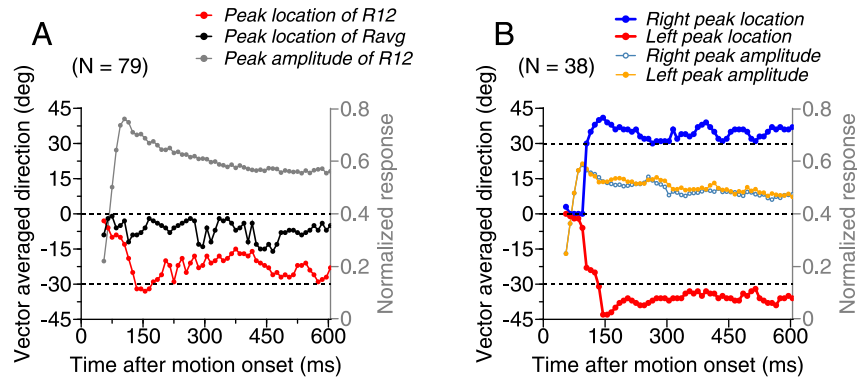


Figure 7. The temporal evolution of the response peak(s) in the tuning curve to bi-directional stimuli separated by 60° . **A.** Location and amplitude of the response peak of neurons that showed a single peak and side-bias. We pooled neurons that showed side-bias to the C-side (i.e. *Dir. 2*) and CC-side (i.e. *Dir. 1*) together as in Fig. 6A and B. **B.** Peak locations and amplitudes of neurons that showed two response peaks to bi-directional stimuli. Right side and left side refer to the peaks seen in Fig. 6E and F. The peak location and magnitude were calculated based on averaged response tuning curves across neurons in each subgroup. The tuning curve was calculated within a 50-ms time epoch, sliding at a 10-ms step. Abscissa indicates the middle point of each time epoch.

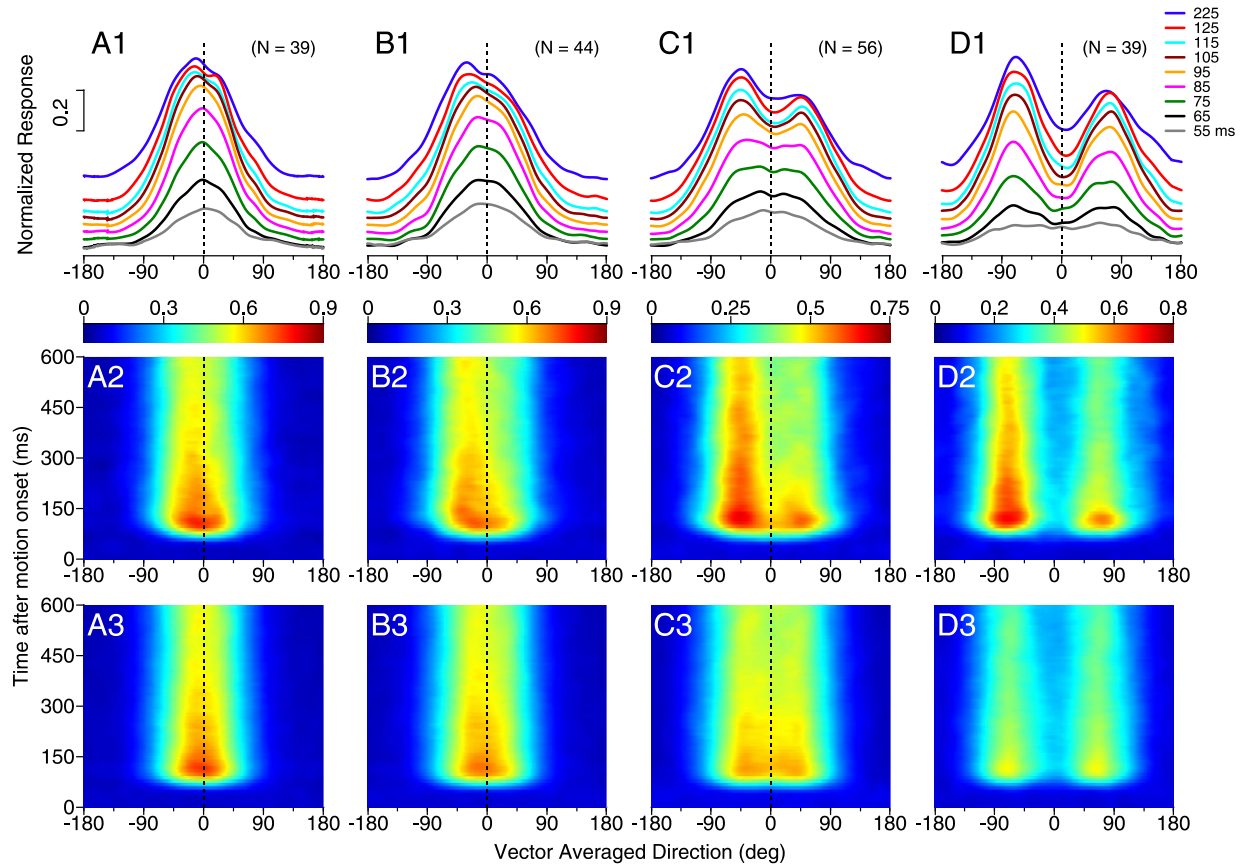


Figure 8. Temporal development of side-bias in the response tuning curves elicited by bi-directional stimuli at different angular separations. **A1-3.** Responses elicited by bi-directional stimuli separated by 45°. **B1-3.** 60° DS. **C1-3.** 90° DS. **D1-3.** 135° DS. **A1-D1.** Subpopulation-averaged tuning curves of responses at different time epochs elicited by bi-directional stimuli. The duration of each time epoch is 50 ms. **A2-D2.** Time course of the response tuning to bi-directional stimuli averaged across side-biased neurons at each DS. Classification for the side-biased neurons was based on the responses to the bi-directional stimuli at each DS. **A3-D3.** Time course of the average of the component responses. Conventions are the same as for Fig. 6A-C. The results from neurons that showed side-bias to the component directions at the C-side and CC-side were pooled together. This analysis included side-biased neurons that showed either a single response peak or two peaks at each DS.

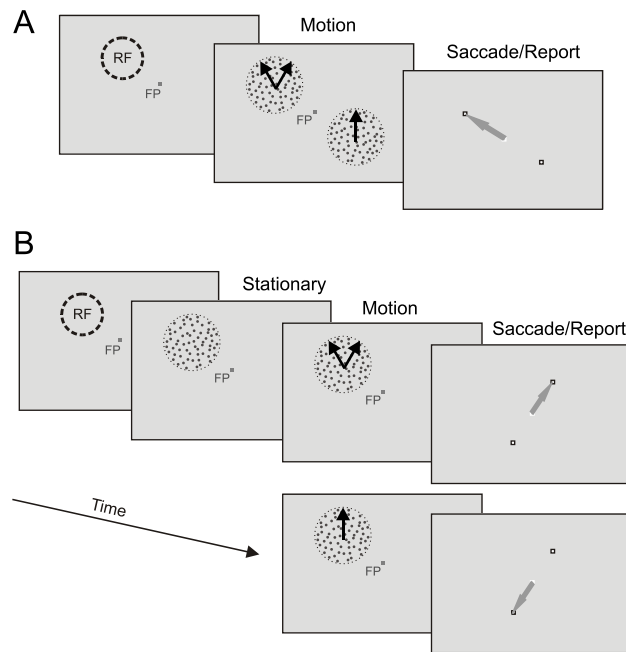


Figure 9. Illustration of two perceptual discrimination tasks. A. In Task I, a bi-directional stimulus and a unidirectional stimulus were presented simultaneously. The animal was required to make a saccadic eye movement after the stimulus offset toward the location where a bi-directional stimulus was presented. The stimulus centered on the RF was either bi-directional or unidirectional. **B.** In Task II, only one stimulus, either bi-directional or unidirectional was presented in a given trial and centered on the RF. The animal was required to make a saccadic eye movement after the stimulus offset toward one reporting target if the stimulus was bi-directional or the other target if the stimulus was unidirectional.

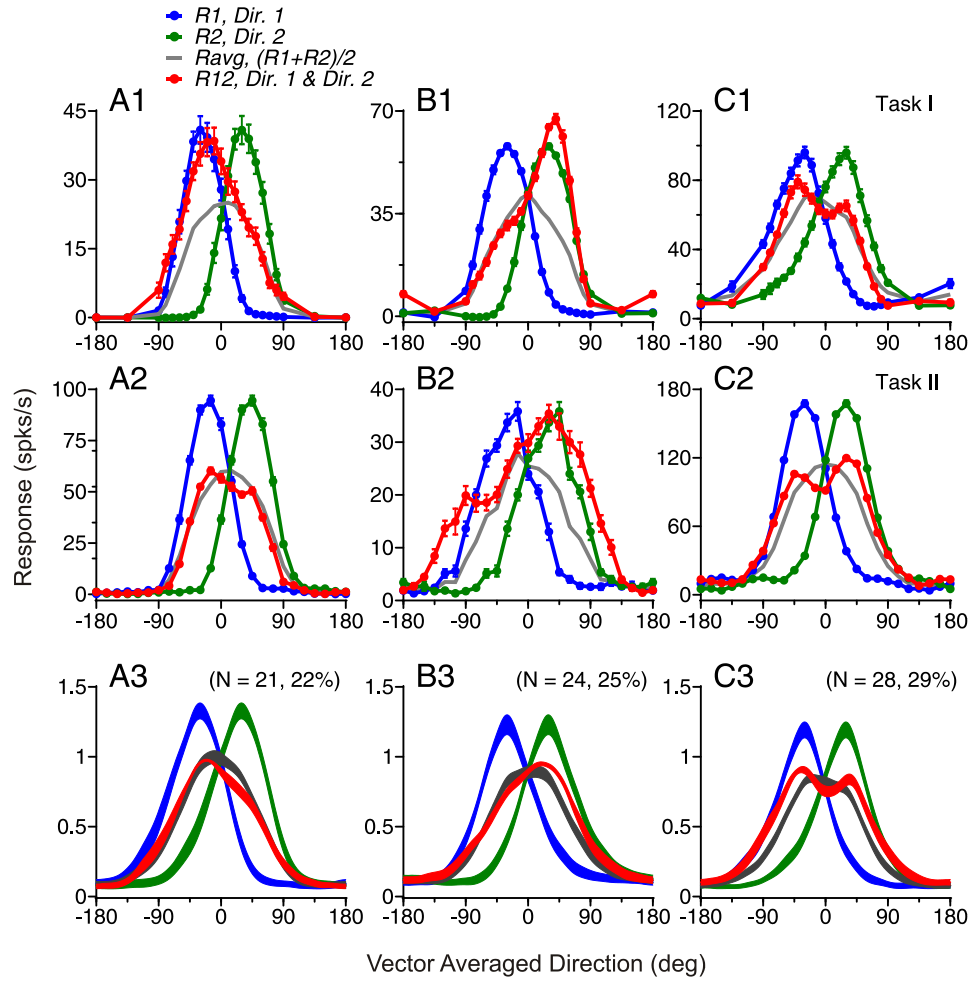


Figure 10. Direction tuning curves obtained while the animals performed the perceptual discrimination tasks. The DS of the bi-directional stimuli was 60°. **A1-C1.** Three example neurons recorded while monkey GE performed Task I. **A2-C2.** Three example neurons recorded while monkey BJ performed Task II. **A3-C3.** Direction tuning curves averaged across subpopulations of neurons from the two animals. **A.** Neurons showing side-bias to the CC-side. **B.** Neurons showing side-bias to the C-side. **C.** Neurons showing two separate peaks in the response tuning curves.

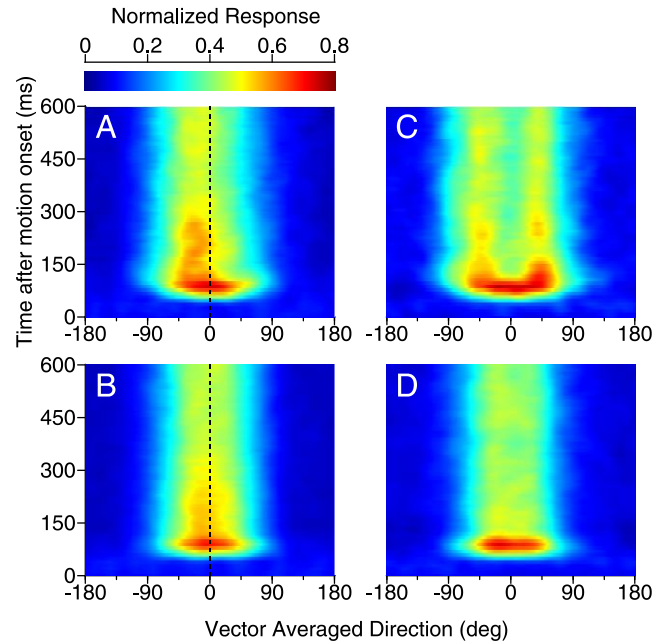


Figure 11. Time course of the direction tuning curves obtained while an animal performed a perceptual discrimination task. The DS of the bi-directional stimuli was 60° . Results were recorded while monkey BJ performed Task II, during which the onset of visual motion was separated from the stimulus onset (see Materials and Methods). **A, B.** Averaged responses from 25 neurons that showed side-bias to the bi-directional stimuli. **C, D.** Averaged responses from 14 neurons that showed two peaks in the tuning curves to the bi-directional stimuli. **A, C.** Response tuning curves to the bi-directional stimuli. Ordinates indicate the middle point of each time epoch of 50-ms. **B, D.** Average of the component responses.

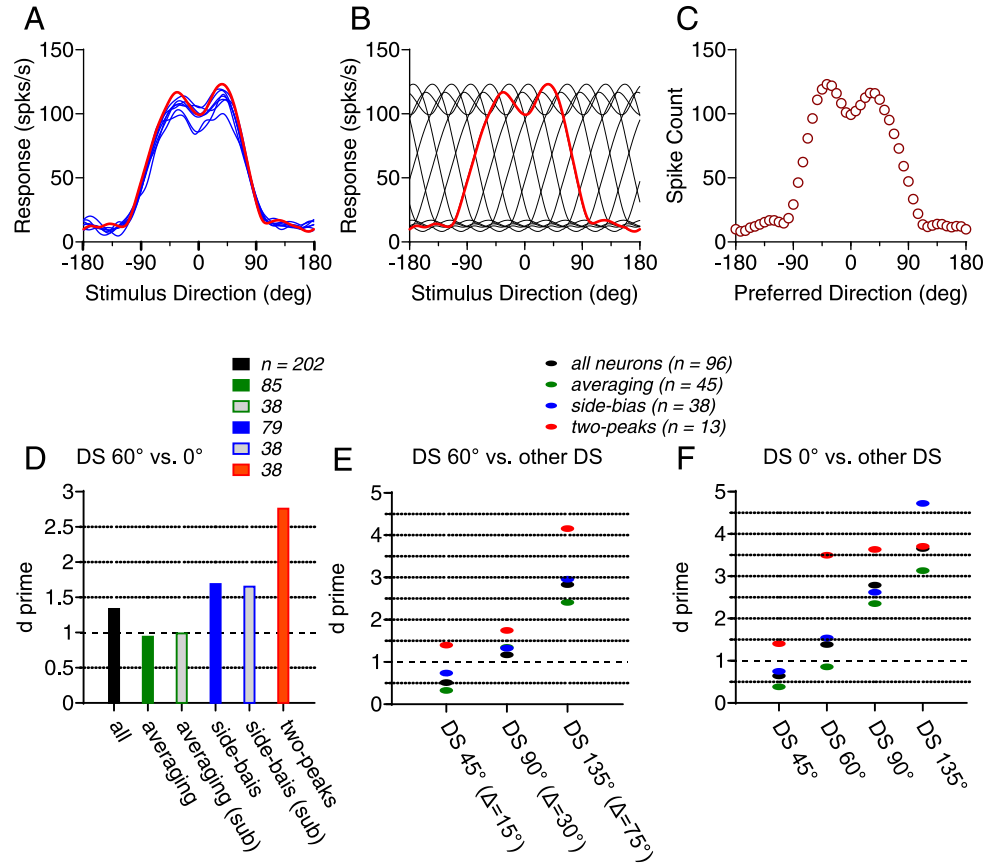


Figure 12. Stimulus discrimination using a classifier of support vector machine (SVM). A-C. The procedure of constructing population neural response based on the direction tuning curve. D-F. Discrimination performance of the classifier. A. Trial-by-trial tuning curves of one example neuron in response to the bi-directional stimuli that had a DS of 60°. A tuning curve based on one block of trials (red) was selected. Other blocks of trials are shown in blue. B. The selected single-trial tuning curve from A (red) was duplicated, and shifted it in a step of 7.5° to create the tuning curves of 48 "cloned" neurons. For clarity, only the cloned neurons that had PDs in a step of 30° were shown (black). C. A single-trial population neural response of the cloned neurons elicited by a bi-directional stimulus moving at VA direction 0°. D. Discrimination between the bi-directional stimuli of 60° DS and the unidirectional stimulus. Stimulus classification was based on the neural responses of all neurons in the data set (black) or one of three subgroups of neurons (green, blue and red). The label "sub" in the abscissa means a subset of randomly selected neurons within the groups of averaging and side-biased neurons. E. Discrimination between a bi-directional stimulus that had a DS of 60° and another bi-directional stimulus that had a DS of 0°, 45°, 90° and 135°, respectively. In this analysis, the side-biased neurons only contained a single response peak to the bi-directional stimulus of 60 DS°, so that the side-biased and two-peaked neurons did not overlap. F. Discrimination between a unidirectional and a bi-directional stimulus. The classifications of averaging, side-biased and two-peaked neurons were based on the responses to DS 60°.

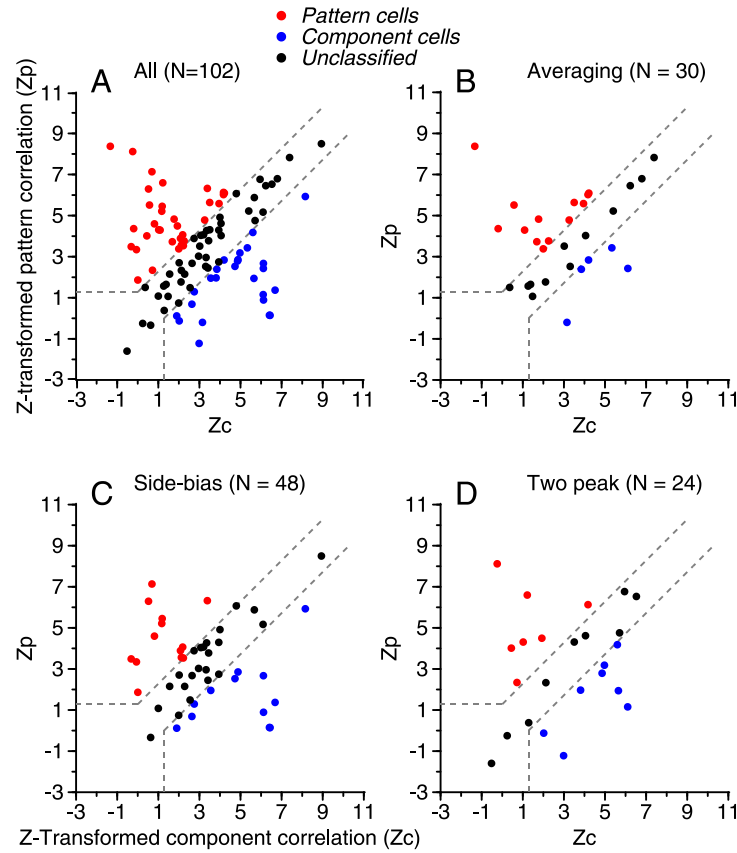


Figure 13. Z-transformed pattern and component correlation between the responses elicited by 135° plaid and pattern/component predictions. Each dot represents results from one neuron. **A.** Results from overall neuron population. **B.** Results from neurons whose responses to the bi-directional random-dot stimuli that had a DS of 60° roughly followed the average of the component responses. **C.** Neurons whose responses to the random-dot stimuli showed side-bias. **D.** Neurons whose responses to the random-dot stimuli showed two response peaks.

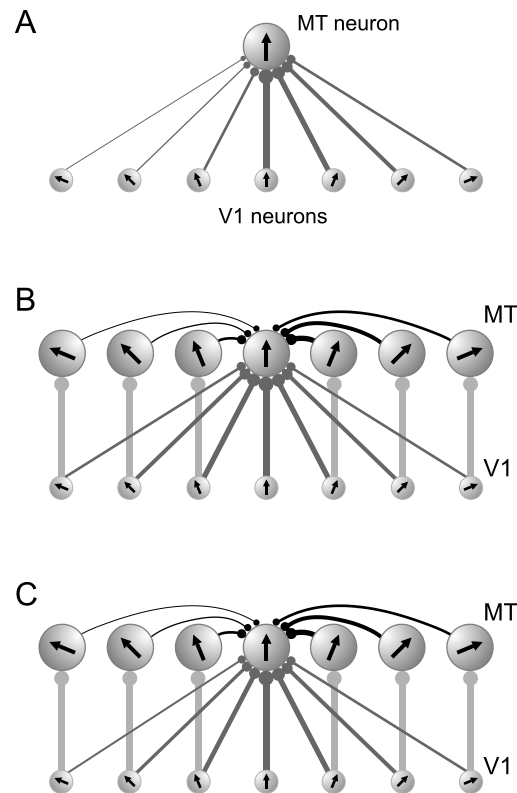


Figure 14. Illustration of possible circuit mechanisms underlying the side-bias. Each circle represents one neuron in area MT or V1. The arrow inside each circle represents the PD of that neuron. The width of the connection line and the diameter of the solid circle at the end of each line indicate the strength of the synaptic connection. **A.** Asymmetric feedforward connections between V1 neurons and a MT neuron. The connection between a V1 neuron, whose PD is at the C-side of the PD of the MT neuron is stronger than the connection between a V1 neuron whose PD is at the CC-side of the MT neuron. **B.** Symmetric feedforward connections and asymmetric recurrent connections among MT neurons. Recurrent connections between the center MT neuron and the MT neurons that have PDs at the C-side are stronger than those at the CC-side. For clarify, only one-way of the recurrent connections is shown. **C.** Slightly asymmetric feedforward connections coupled with asymmetric recurrent connections. Except for the center MT neuron, only the feedforward connections between V1 neurons and MT neurons that have matching PDs are illustrated for simplicity. Recurrent interactions between MT neurons may include both excitatory and inhibitory connections (not differentiated in the illustration).

Chapter 4

Neural representation of multiple motion directions at different depths and the effects of feature-based attention

4.1 Introduction

In motion transparency, one surface is often perceived as moving on top of the other despite no actual depth cue in the display. In addition to being the consequence of motion segmentation, the depth perception may contribute to the buildup and maintenance of motion segmentation. Early studies have shown that binocular disparity facilitates the detection of transparent motion in a variety of stimulus settings, such as random-dot stimuli, grating-plaid stimuli and motion after-effects (Hibbard & Bradshaw 1999, Qian et al 1994, Snowden & Rossiter 1999). Moreover, the addition of depth difference enables the detection of more transparently moving stimuli than without depth difference (Greenwood & Edwards 2006). These studies all demonstrate that transparent motion is more frequently observed when components are presented with different binocular disparities than when components have the same disparity. To understand the relationship between the perception of different depths and perception of transparent motion, it is critical to study the joint representation of the transparent motion and different binocular disparities.

MT neurons are selective to both motion direction and binocular disparity. A recent study (Smolyanskaya et al 2013) investigated the joint tuning of individual MT neurons for motion direction and binocular disparity and showed that MT neurons represent the motion direction and binocular disparity information in a largely separable manner -- the preferred direction does not change as the disparity varies and vice versa. However, their visual stimuli constituted only one random-dot patch moving at a randomly-

drawn depth. It is different from the scenario when two stimuli move transparently at different depths. So, it remains to be determined how the transparent motion at different depths is represented.

It has been shown that the neuronal responses evoked by an individual component stimulus moving in the PD are suppressed by adding a non-PD component stimulus. This is known as opponent suppression, which has also been observed in the neuronal representation of transparent motion stimuli (chapter 2 & 3). The study by (Bradley et al 1995) found that the opponent suppression in area MT is removed and a more salient transparent motion perception is developed when opposing directions are placed at different stereoscopic depths, suggesting that the stereopsis can assist the buildup of transparent motion perception. Later studies (Hibbard and Bradshaw 1999; Snowden and Rossiter 1999) suggested that this facilitation effect may relate to the attention modulation effect proposed by (Treue & Maunsell 1996). However, it remains to be examined whether the facilitation is related to attention modulation, and if so, how the attention modulation is implemented?

It has been well known that visual attention improves behavioral performance in segmentation task and contrast discrimination task (Carrasco 2011, Carrasco et al 2006a, Yeshurun & Carrasco 1998). However, there has been a longstanding debate about whether segmentation guides the deployment of visual attention or visual attention helps segment visual scenes. This is partly because it is difficult to isolate the bottom-up effect and top-down effect on visual segmentation. A recent study shows that the strength of top-down attention modulation is constrained by the degree to which the segmentation has been resolved by bottom-up processing (McMains & Kastner 2011). The previous theoretical study suggests that bottom-up processing alone can roughly segment scenes and feedback from top-down processing are needed for the accurate delineation of segmentation boundary (Sharon et al 2006). However, the interaction between attention modulation and stimulus-driven visual processing remains unclear.

The attention effect on neuronal activity has been extensively studied and it is known that attention increases the signal-to-noise ratio in many ways to parse the complex visual scenes. Particularly for motion

segmentation, prior studies have shown that attention modulation can enhance the neuronal response to the attended motion components in area MT (Wannig et al 2007). However, in their experiment, they cued the animal with color, a feature which MT neuron is not selective for, indicating the presence of the attention effect does not require a perceptual judgment of the motion information. Perceptual discrimination of target and distractor may have already been completed by other brain regions. Therefore, to investigate how attention is implemented through the perception of depth differences, we designed an experiment to manipulate attention based on the depth cue to examine the attention effect on transparent motion perception and its neuronal representation.

We found that MT neuron preferentially represents the near component of the transparent motion stimuli moving at different depths. In addition, a tuned feature-based attention effect was observed as visual attention was switched between near and far plane, and a significant firing rate modulation was only observed when the near component moved near the PD. Our results revealed the neural representation of transparent motion at different depths and provided evidence on how visual attention facilitates motion segmentation through adding depth differences.

4.2 Materials and Methods

One adult male rhesus monkey (*Macaca mulatta*) was used in this study so far, future work will need to collect data from another monkey. This monkey is one of the animals that we used in the previous studies (Chapter 2 and Chapter 3). All our procedures for surgical preparation and electrophysiological recording were the same, and the experimental protocols were approved by local *IACUCs* and followed *the NIH Guide for the Care and Use of Laboratory Animals*.

Visual stimuli and behavioral paradigm

Visual stimuli were achromatic random-dot patterns presented on a 25" CRT monitor at a viewing distance of 63 cm. The monitor has a resolution of 1024x768 pixels and a refresh rate of 100 Hz.

Transparent motion stimuli were constituted by two 15°- wide overlapping circular random-dot patches moving at two different depths – one at 0.1° binocular disparity ('Near depth') and the other at -0.1° binocular disparity ('Far depth') relative to the fixation point. To enhance the perception of a reference surface at the fixation plane, a stationary rectangle, random-dot pattern extending 25°x30° was present at 0° depth. The transparent motion stimuli were presented and centered within a stationary circular aperture that had a diameter of 16°. There was a 1° blank gap between the random-dot moving pattern and the stationary reference background. The binocular disparity of each component stimulus was created by overlapping two identical dot patterns with one in red and the other in green, and adjusting the horizontal offset in between this pair of dot patterns according to the specific disparity value needed. In the present study, the "Near" component stimulus was always fixed in front of the fixation plane at 0.1° depth and the "Far" component stimulus was fixed behind the fixation plane at the -0.1° depth. The animal viewed the stimuli through a pair of anaglyph glasses (*Kodak Wratten filters, nos. 25 and 58*) colored in red in front of the right eye, and green in front of the left eye so that only one set of dots was visible to each eye and a binocular disparity was perceived as a result. Crosstalk between the two eyes, as viewed through the filters, was measured as 14.5% (the luminance of red light through red glasses is measured as 13.85 cd/m², likewise, red-through-green is 2.32 cd/m², green-through –green is 14.5 cd/m², green-through –red is 1.8 cd/m²) Dots of each patch were presented at a spatial density of 0.89 dots/°² with infinite lifetime and had 100% coherence. The unidirectional stimulus contained one pair of random-dot patches moving in the same direction while the transparent motion stimuli were constituent by two pairs of random-dot patches and therefore its dot-density was twice as high as that of unidirectional stimulus. The stimuli usually moved at

a speed in the range of 8~15 degree/s. The background luminance was 0.68 cd/m² through red glass and 1.01 cd/m² through green glass

In a stereoscopic system, the crosstalk rate is defined as the amount of luminance leaking through the “wrong” filter divided by the amount of luminance passing through the “correct” filter. An early study has shown that stereoscopic depth perception can survive relatively high levels of crosstalk (10% in Shiozaki et al., 2012).

Experimental design

After an MT neuron was isolated, the selectivity to binocular disparity was determined by measuring the neuronal responses to a series of random-dot pattern moving in the preferred direction at a thirteen different depths ranges from -2.0° to +2.0° (including -2°, -1.5°, -1°, -0.75°, -0.5°, -0.25°, 0°, 0.25°, 0.5°, 0.75°, 1°, 1.5°, and 2°) and near the preferred speed (often within a range from 2~20 °/s). Likewise, the size selectivity was determined by measuring the response tuning to a series of a random-dot pattern of different sizes, ranging from 3° to 30° (including 3°, 5°, 7.5°, 10°, 15°, 20°, 25°, and 30°).

In the main paradigm, the monkey was cued to report the motion direction of the component stimulus moving at the cued surface. Every trial began with the appearance of the fixation dot, after the fixation was achieved, a static circular dot-patch was presented with a diameter of 15° at either near plane (0.1° depth) or far plane (-0.1° depth) as the attention cue for 500ms, followed by a 300ms blank period. Then, both stimulus components appeared for 500ms but remained stationary, then visual stimuli started moving for 600ms. After the motion period ended, 12 reporting targets arranged along a circle of 16° wide, were presented for 650ms and the animal was required to report the motion direction of the cued component by making an eye saccade to the corresponding target and holding his fixation within an invisible 2°x 2° target window for 350ms. In each block of trials, we randomly interleaved trials of single stimulus located

either at near plane or far plane with trials of transparent motion stimuli separated by either 60° or 120°. For single stimulus trials, the attention was always directed to the plane on which the stimulus was presented, while for the transparent motion stimuli, animal's attention would be directed to either near or far plane. Each pair of transparent motion stimuli was tested with two different configurations. The motion component at the near depth was moving in a direction that was either at the clockwise (C) side or the counter-clockwise (CC) side of the motion component at the far depth. All types of trials were tested in

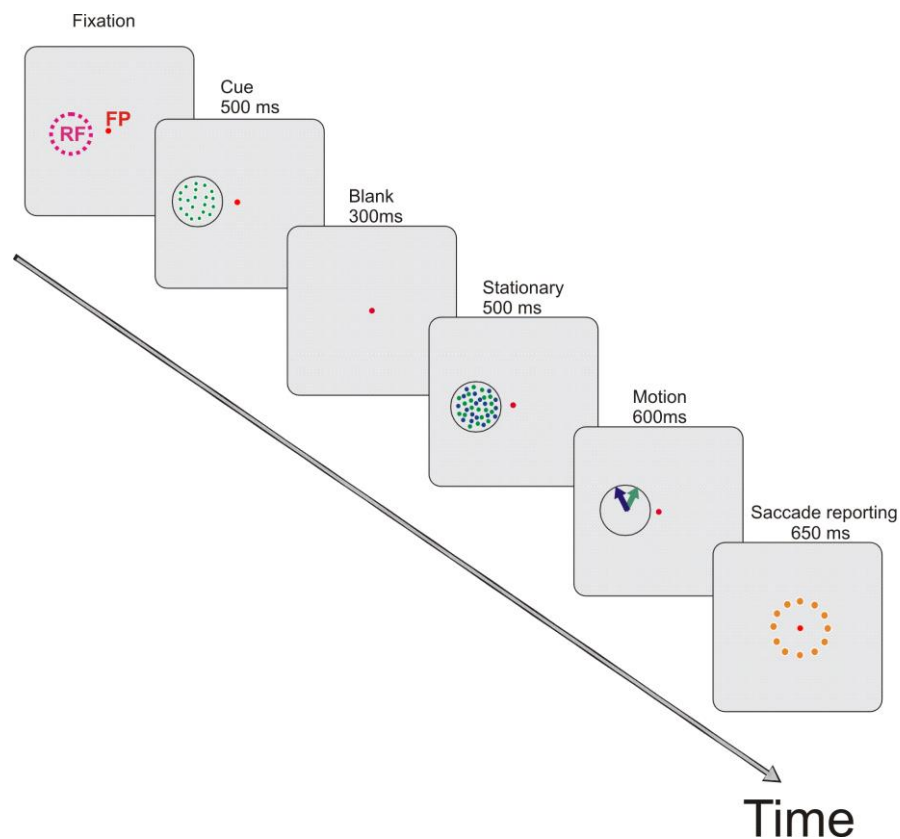


Figure 4.1. Experimental Design. The spatial configuration of the fixation-point (FP) and the receptive field (RF) of the recorded neuron. The visual stimuli and visual cue were presented over the neuron's RF. The green and blue dots were used to denote the near and the far plane, respectively. Each 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 another 300ms blank period, either a unidirectional stimulus or a transparent motion stimuli (as shown in the figure) appeared and stayed stationary for 500ms, then motion started and lasted for 600ms. The animal had to maintain fixation of FP as long as it was on. At the end of each trial, the monkey was required to make an eye saccade to report the motion direction of the stimulus on the cued plane and hold the fixation within a 2°x2° window for 350ms.

12 different motion directions around 360° separated by $30^\circ/\text{step}$. There was a total of 120 trials in one block of trials. A new block of trials was initiated only after the previous block was completed. A trial would be terminated whenever the monkey broke the fixation requirement before the end of the motion period or did not saccade to target on time (within 650ms after motion offset) or did not hold fixation long enough (350ms) after making a saccade to one target. Juice reward was delivered when a correct target was chosen at the end of each trial, but a saccade made to an incorrect target was considered as a completed trial. To balance the need for good behavioral performance and the need for strong neuronal activity, the visual stimuli were usually presented at a location covering the receptive field (RF) of the recorded cell but might not be centered with a neuron's RF when the RF is fairly peripheral.

Data analysis

The behavioral performance was measured by the correct rate of reporting the motion direction of the cued surface. Errors were split into “distractor errors” in which the animal reported the motion direction of the other uncued motion component, “VA errors” in which the animal reported the vector-averaged direction of the two motion components and the Unspecific errors in which the animal reported any other directions.

We measured the neuronal responses during the motion period. Cells were included in the analysis only if at least five repetitions were completed for each trial and the correct rate across all trials of the behavioral task was at least 49% (there is a clear underestimation of the correct rate for some recording sessions due to the lack of calibration, average correct rate is very likely to increase as calibration is added into the analysis in the future). Mean firing rates were computed by averaging the responses elicited during

the 600ms motion period associated with all the trials, for now, later we will only analyze the neuronal responses associated with correct trials.

We fitted the neuronal response with the Summation plus non-linear interaction model (or called SNL model, as shown in equation 4.1) described in chapter 2. The contribution of each component response was evaluated by the fitted weights of each component.

$$R_{12_{pred}}(\theta_1, \theta_2) = w_1 \bullet R_1(\theta_1) + w_2 \bullet R_2(\theta_2) + b \bullet R_1(\theta_1)R_2(\theta_2) \quad (4.1)$$

To characterize the attention effect, we measured the difference between the area under the bi-directional response tuning curve (AURC) when the animal attended to the near component and when the animal attended to the far component. To focus on the conditions in which neurons were well evoked, the AURC was defined as the area under the bi-directional response tuning within 30° from the direction when one of the stimulus components moved in the PD. Specifically, I analyzed the attention effect when the near component moved near the PD and when the far component moved near the PD.

To examine the time-course of the response tuning to the transparent motion 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 far component response across all time windows.

4.3 Results

To understand the neural representation of transparently moving stimuli separated by depths and the effects of feature attention in motion segmentation, we recorded from 80 MT neurons in one macaque monkey as the animal performed a motion discrimination task involved feature-based attention. We excluded neurons

that showed a weak response or were recorded during sessions when the behavioral performance was poor (averaged correct rate across all trials was less than 49%). Fifty-nine neurons remained in our data set for further data analysis.

Behavioral Performance

The monkey performed well in discriminating the motion direction of the cued component with a 69% correct rate averaged across all recorded sessions recorded for the selected 59 neurons (**Table 4.1**). When only a single stimulus component was presented, the averaged correct rate in detecting motion direction was 81% and the animal was equally good at detecting near component and detecting the far component. When two motion components were presented together, the average performance in detecting motion directions of cued component is 66%. The correct rates for the angle separations of 60° and 120° were similar, but the average correct rate of the conditions when the angle separation between the two motion components was 60° was significantly higher than that of the conditions when the angle separation was 120° (67% for DS=60°, 65% for DS=120°, paired t-test, p-value = 0.0012, N=59). This is probably due to significantly longer training time was spent on DS60° conditions. Also, the monkey's correct rate of discriminating motion direction of the near component was significantly greater than that of discriminating motion direction of the far component (**Figure 4.2**, DS=60°, paired t-test, $p < 10^{-12}$, N=59; DS=120°, paired t-test, $p < 10^{-12}$, N=59). In addition, we found that the animal's distractor error was lower when the animal attended to near component than when the animal attended to far component (**Table 4.1**, DS=60°, paired t-test, $p < 10^{-6}$, N=59; DS=120°, paired t-test, $p < 10^{-9}$, N=59). Overall, the correct rate is about 8~10 times higher than the distractor error rate (**Table 4.1**), which also indicates that the animal's attention was well controlled. There was a positive correlation between the performance of discriminating the direction of far component and the performance of discriminating the direction of the near component (**Figure 4.2B**, $R^2 = 0.62$, p-value <0.001). The slope of the linear regression between the correct rate for near component and

correct rate for the far component is less than 1 (slope =0.78), which re-confirms that animal was better at detecting the near component.

Table 4.1. Behavioral performance.

(N=59)	DS60		DS120	
	Cue Near	Cue Far	Cue Near	Cue Far
Correct rate	70.5%	64.0%	68.5%	61.9%
Distractor error	6.4%	9.6%	6.3%	10.4%
VA error	11.1%		1.1%	

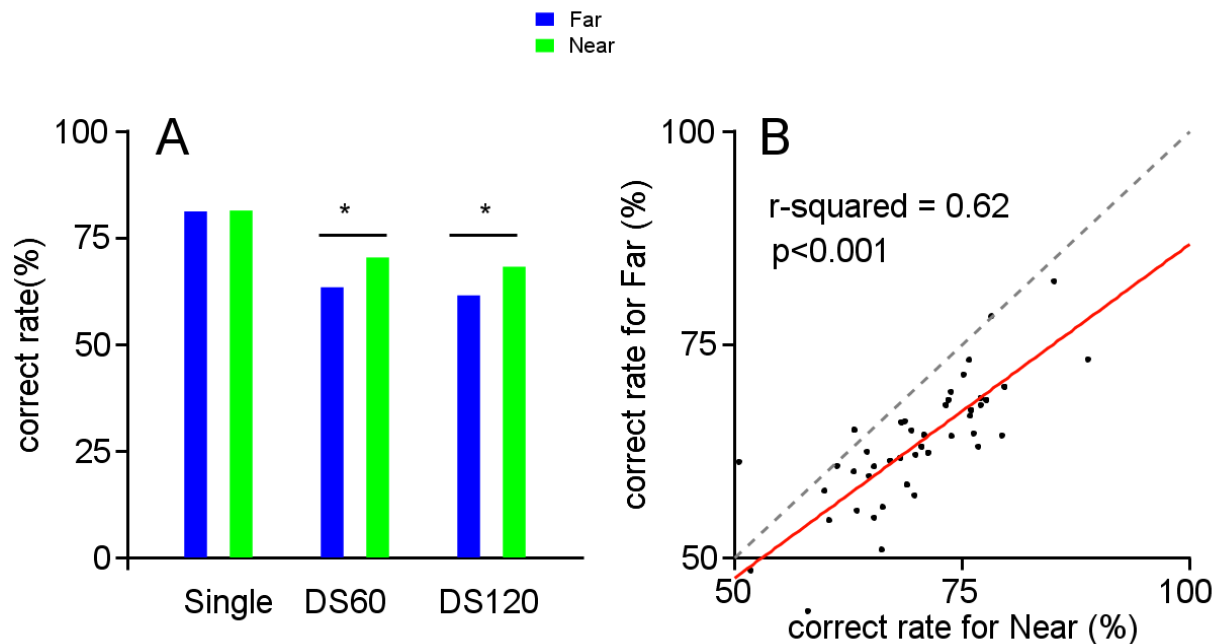


Figure 4.2. Behavioral performance. **A.** The animal had a good performance in detecting the motion of the cued component across different angle separations and attentional states. * denotes p -value < 0.001, paired t-test. **B.** Animal's performance in detecting near motion component is significantly correlated with that of detecting far motion component and significantly better than detecting far motion component. The red line is the linear regression of the correct rate for Far on the correct rate for Near. The gray line is where the monkey equally good at discriminating motion direction of near and motion direction of the far component.

Significant dominance by the Near component

A total of 59 neurons were recorded from monkey GE. We plotted the neuronal response tuning as a function of the vector-averaged direction of the two motion components as described in the previous two chapters. **Figure 4.3** shows the response tuning curves of a sample MT neuron, which preferred the far disparity and therefore has a stronger peak firing-rate in the direction-tuning curve to the far component. This plot shows that the bi-directional response tuning (shown in red and black) biased toward the tuning of the responses elicited by the unidirectional stimulus moving at the near plane (shown in green) regardless of the attentional state (red denoted attending to near, black denoted attending to far). This neuron had a stronger far component response (shown in blue). For instance, the black curve in **Figure 4.3 A** shows that the bi-directional response biased toward the near component although the animal attended to far plane and the far component response tuning was stronger. We found this bias toward near component was consistent across conditions regardless of when the near component moved at the C-side (**Figure 4.3. A, C**) or the CC-side (**Figure 4.3. B, D**) of the two directions and across conditions regardless of when the angle separation between the two components was 60° (**Figure 4.3 A, B**), or 120° (**Figure 4.3 C, D**). This near-component dominant neural representation (hereafter called near-dominance) is also shown in the population-averaged response profile as shown in **Figure 4.4**, in which the bi-directional responses tuning (shown in red and black line) is more biased toward the near component response tuning than the linear averaging of the two component response tuning (shown in gray dashed line). What is more striking is that the far disparity preferred neurons also consistently biased toward the near component (see **Appendix 1** for a detailed description of each condition).

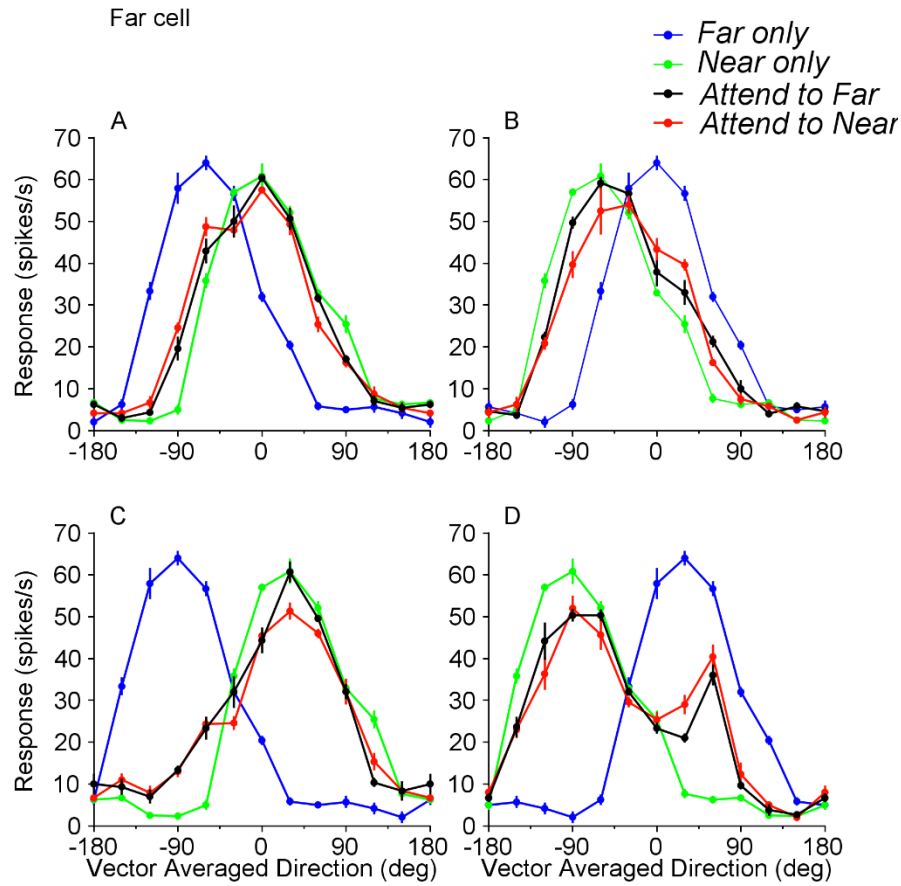


Figure 4.3. Responses from an example neuron show a strong dominance of the near component. A. In response to the transparent motion stimuli moving at different depths and separated by 60° , bi-directional responses tuning (shown in red for attending to Near and black for attending to Far) strongly biases toward the near unidirectional component response tuning (shown in green), although the Far component response tuning (shown in blue) is stronger. Similar dominance by the near component response was observed in the other stimulus configurations shown at DS 60° (B), and for the two stimulus configurations at DS 120° (C and D).

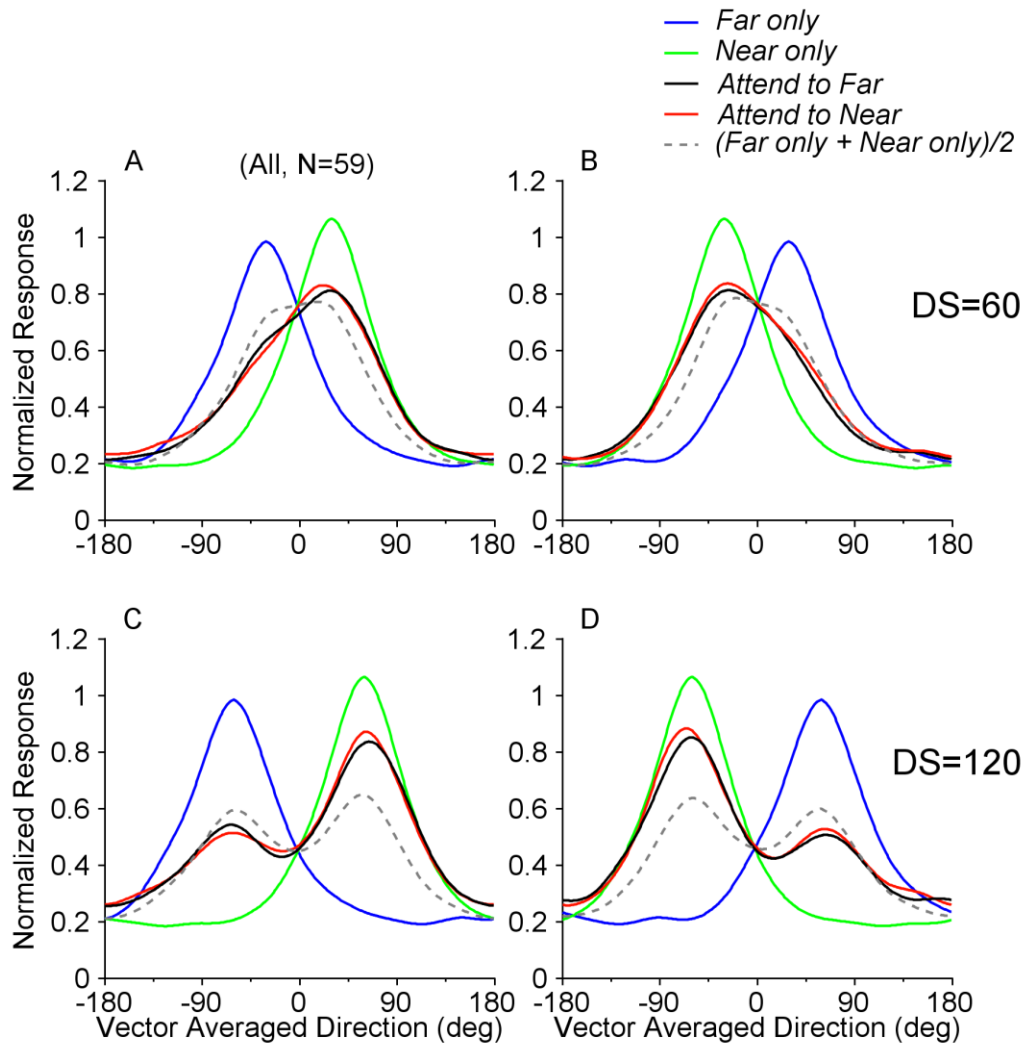


Figure 4.4 Population-averaged response tuning curves. The bi-directional response tuning is biased toward the response tuning of the near component. **A.** the transparent motion stimuli were separated by 60°, and the near component moved at the relative C-side; **B.** stimuli were separated by 60°, but the near component moved at the relative CC-side. **C.** stimuli were separated by 120°, and the near component moved at the relative C-side; **D.** stimuli were separated by 120°, and the near component moved at the relative CC-side. The gray curve is the linear averaging of the two component response tuning curves.

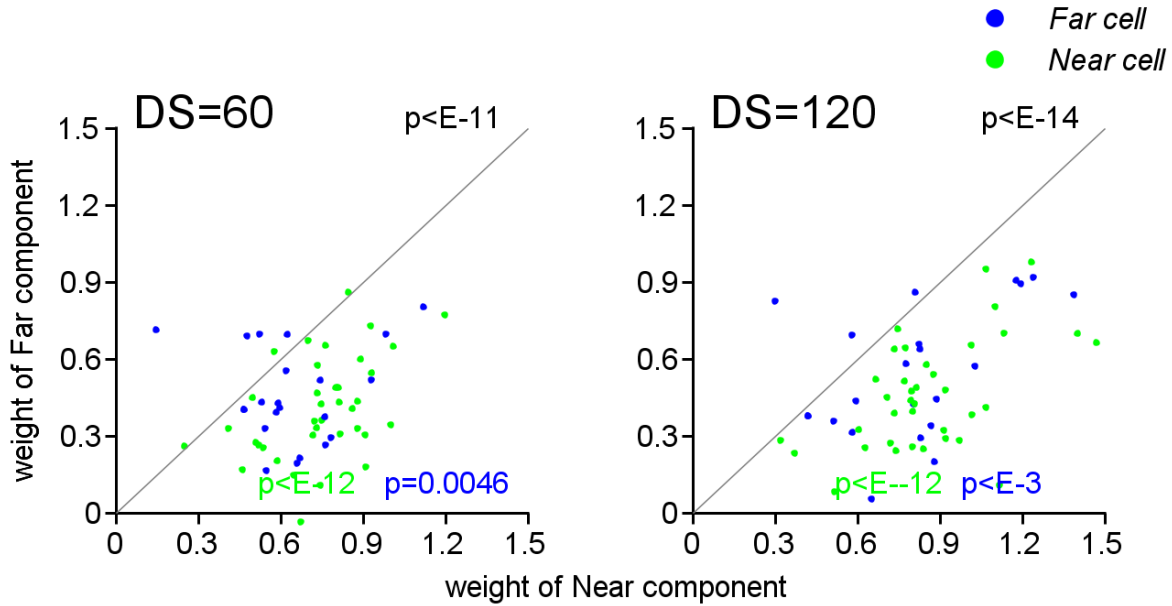


Figure 4.5. Population distribution of component response weights. Weights were calculated from the SNL model fit, weights for near component was plotted as a function of weights for the far component. Each dot represents one cell, both near cells (shown in green) and far cells (shown in blue) have significantly higher weights for near component across two different angle separations (left panel is when $DS=60^\circ$, right panel is when $DS=120^\circ$).

To further assess the near-dominance across the whole population of 59 neurons, we fitted the bi-directional responses with the SNL model (see Equation 4.1) based on the component responses for each individual neuron and compared the weights for near component and those for the far component. We found that the weight of near component on average was significantly greater than the weight of the far component across different angle separation conditions and cell types as shown in **Figure 4.5** (paired t-test, $p < 10^{-11}$ for 60° separation, $p < 10^{-14}$ for 120° separation; $p < 10^{-12}$ for near cells under the conditions of both 60° separation and 120° separation; $p < 10^{-2}$ for far cells under the conditions of both 60° separation and 120° separation). The value of the component weight shown in **Figure 4.5** was the averaged value of the same component weights across attentional states and configurations (see **Appendix 2** for a detailed description of each condition).

Significant attention-switch effect

Another goal of this study was to evaluate the effects of attention on neuronal responses when the animal switched attention from one surface to another. **Figure 4.6** shows the responses profile of one sample neuron which showed significant attention effect across the two angle separations. Comparing the bi-directional responses between the two attention states, we found that the sample neuron's bi-directional response increased when the attended component moved near the PD of the cell recorded and decreased when the ignored component moved near the PD. We call this firing rate modulation when attention was switched between two planes as an attention-switch effect. **Figure 4.4** shows the population-averaged response profile, which shows this attention-switch effect is constrained within a small range around the PD.

To further evaluate the attention-switch effect at the population level, we computed the difference of the area under the response curve (AURC) of the bi-directional response tuning when the animal attended to near component with the area when the animal attended to the far component. As shown in **Table 4.2**, we found the increase of the attention-switch is marginally significant only when the near component moved in a direction less than 30° away from the PD of the recorded cell ($N=59$, paired t-test, $p=0.066$ for DS60 and 0.042 for DS120). However, the attention-switch effect was not significant elsewhere, even when the far component moved near the PD ($N=59$, paired t-test, $p=0.356$ for DS60 and 0.668 for DS120).

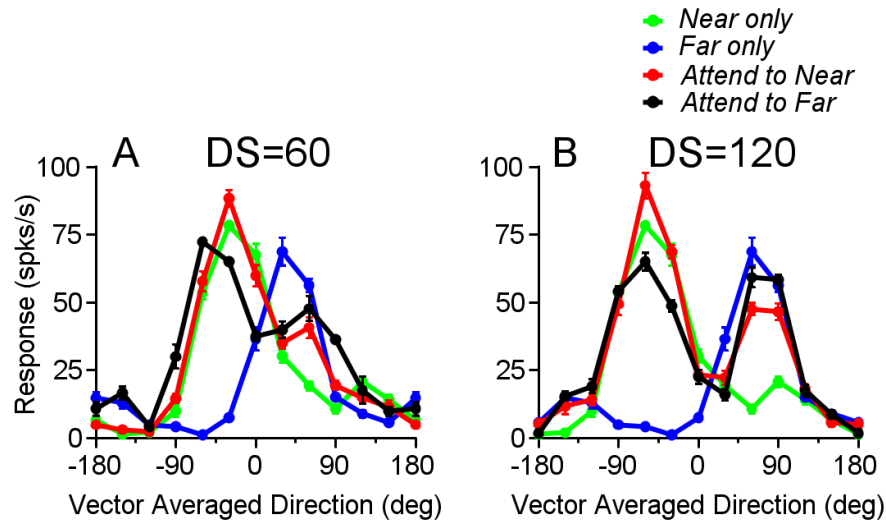


Figure 4.6 An example neuron showing an effect of attention switching. The bi-directional response tuning is biased toward the response tuning of the near component. **A.** the transparent motion stimuli were separated by 60°, and the near component moved at the relative CC-side. Bi-directional response tuning when attending to Near is denoted in red, whereas attending to far is denoted in black. Near and far component response tunings were denoted in green and blue, respectively; **B.** the transparent motion stimuli were separated by 120°, and the near component moved at the relative CC-side.

Table 4.2. Attention-switch effect when behavioral performance is 49%+.

PD $\pm$ 30°, (N=59)	DS60 (Near at C-side)	DS60 (Near at CC-side)	DS120 (Near at C-side)	DS120 (Near at CC-side)
Near component	p = 0.148	p = 0.218	p = 0.145	p = 0.159
moved near PD	p = 0.066		p = 0.042	
Far component	p = 0.815	p = 0.083	p = 0.198	p = 0.327
moved near PD	p = 0.356		p = 0.668	

Attention-switch effect correlates with behavioral performance

To evaluate the correlation between the strength of the attention-switch effect and the behavioral performance, we further analyzed the attention-switch effect for cells recorded when the correct rate of the animal's performance was greater than or equal to 75%, and we found a more robust attention-switch effect between the two attentional states (**Figure 4.7**, apparently p-values of **Table 4.3** are smaller than their counterparts in **Table 4.2**). Further, we compared the attention-switch effect between 49% performance and 75% performance in terms of the difference of the AURC, we found a stronger attention-switch effect associated with the better behavioral performance (sign-rank test, p-value <0.01). This finding suggests that the attention-switch effect positively correlates with the behavioral performance and therefore may serve as the neural correlate of the attentional modulation of the animal behavior.

Table 4.3. Attention-switch effect when behavioral performance is 75%+.

PD $\pm$ 30°, (N=11)	DS60 (Near at C-side)	DS60 (Near at CC-side)	DS120 (Near at C-side)	DS120 (Near at CC-side)
Near component	p = 0.067	p = 0.03	p = 0.015	p = 0.031
moved near PD	p = 0.004		p = 0.0007	
Far component	p = 0.065	p = 0.08	p = 0.004	p = 0.004
moved near PD	p = 0.01		p = 0.003	

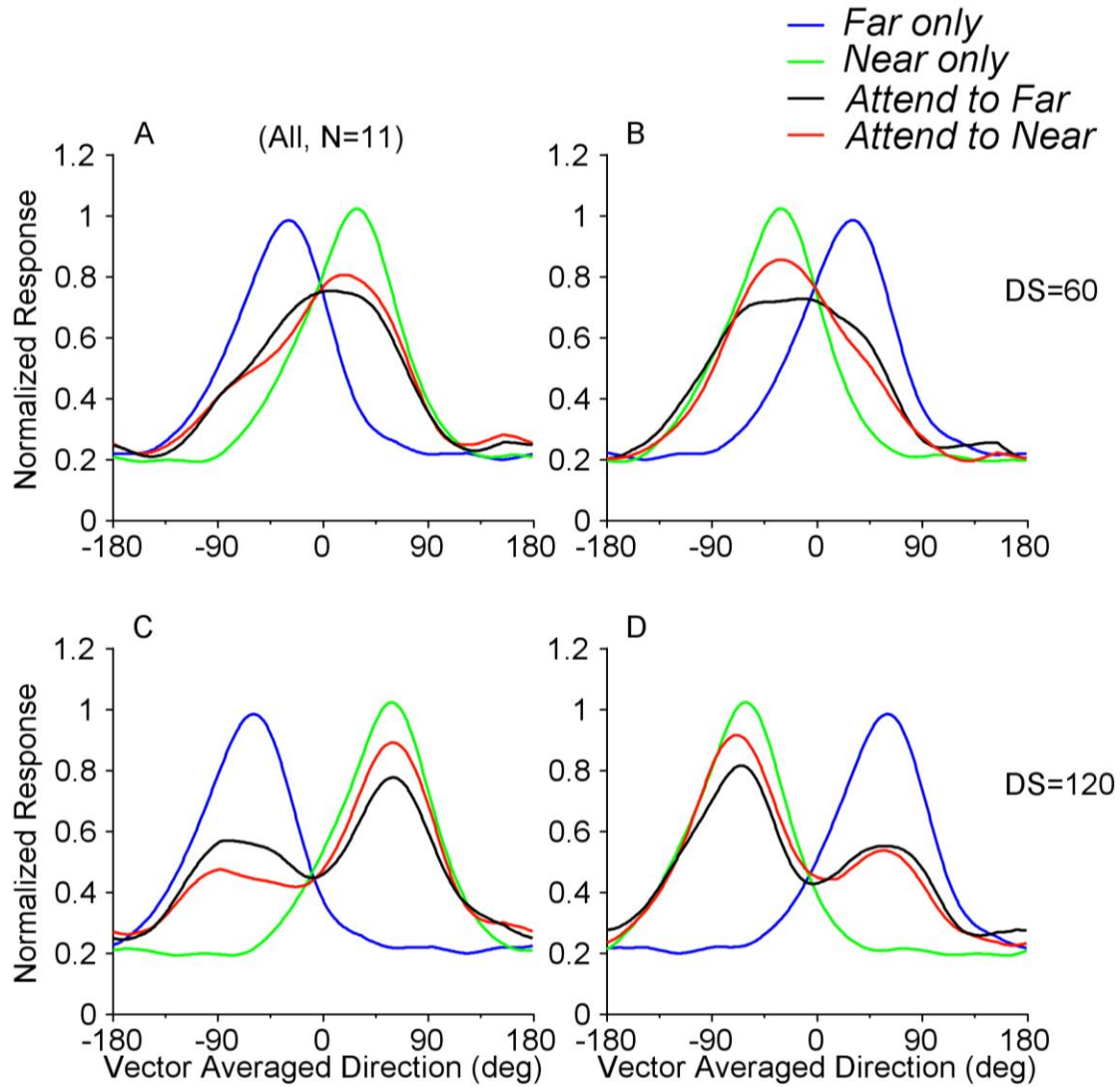


Figure 4.7 Attention-switch effect when the correct rate of the animal's behavioral performance is above 75%. The bi-directional response tuning when the animal attended to Near depth is denoted in red when the animal attended to Far depth is denoted in black, The Near component response tuning is denoted in green and the Far component response tuning is denoted in blue; **A.B.** the transparent motion stimuli were separated by 60°. **C.D.** the transparent motion stimuli were separated by 120°.

Development of the Near-dominance

To further evaluate the temporal development of dominance of the near component in the neuronal representation of the transparent stimuli moving at different depths, we plotted the bi-directional responses

within a 50-ms sliding time window at 10-ms per step for Far cells population and Near cells population separately.

We found that the dominance of near component evolved over time and the near dominance developed slightly faster at larger angular separation. When the angle separation equals to 60° , it took both Near cells and Far cells roughly about 150-ms to bias the bi-directional response tuning toward the Near component; when the angle separation was 120° , it took both Near cells and Far cells roughly about 100ms to develop the bias toward the Near component. Between Near and Far cells, the initial response of Far cells was stronger at Far component side while the initial response of Near cells is more symmetric.

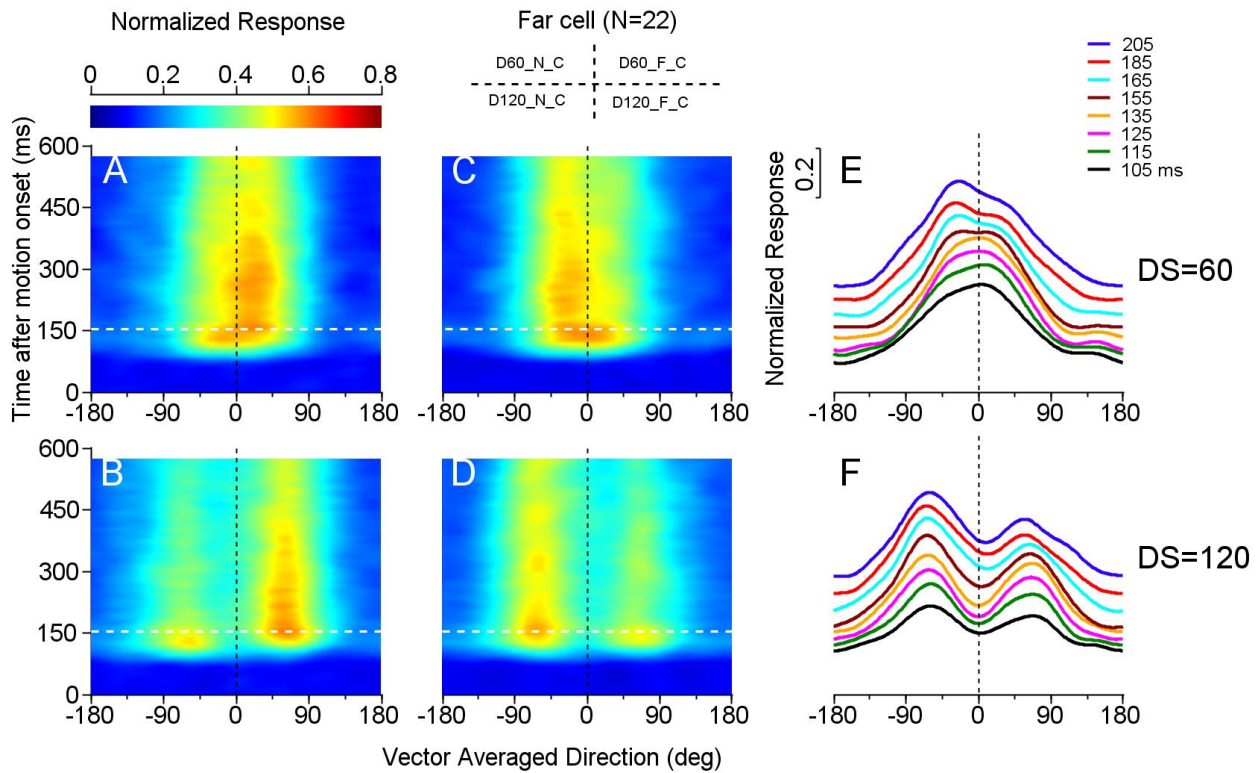


Figure 4.8 Time-course of the bi-directional response tuning of far cells shows the build-up of near-dominance. *A, B.* The animal attended to the near component and the near component located at the C-side. Angle separation was 60° for *A*, and 120° for *B*; *C, D.* The animal attended to the far component and the far component located at the C-side, Angle separation was 60° for *C*, and 120° for *D*; *E, F.* Tuning curves of bi-directional response at different time epochs after the stimulus onset when the animal attended to the far component and the far component located at the C-side. Angle separation was 60° for *E*, and 120° for *F*. Please note the scale bar for the normalized neural response in panel *E*.

4.4 Discussion

This study demonstrated that neurons in area MT preferentially represent the near component in response to multiple stimuli moving at different depths. In addition, a tuned feature-attention effect was observed as visual attention was switched between near and far surface. Significant attention effect was observed only when the stimulus component at the near surface moved in a direction close to the PD of the recorded neuron. The time-course analysis showed that the near-dominance representation took roughly about 100~150-ms to develop (this delay is a rough estimate and will need some quantification in the future).

The Near-dominance

We found that in representing multiple stimuli moving transparently at different depths, MT neurons preferentially represented the near motion component regardless of whether the animal was attending to the near surface or the far surface. In addition, we found that the monkey had a better behavioral 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 the neuronal and behavioral level?

Early studies have demonstrated that Near disparity require less stimulus duration than far disparity for stereoscopic depths perception (Cormack et al 1997, Patterson et al 1995, Stelmach & Tam 1998). Later, the study by Hibbard and Bradshaw (1999) revealed 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. However, in our experiment, the animal performed equally well at detecting the motion direction of a component stimulus when they were presented alone at either the near or far depth. This may be due to the fact that the motion coherence used in my study (100%) is much higher than the perceptual threshold. The animal's asymmetric performance in detecting two motion components emerged only when the transparent motion stimuli were presented. This might be

because the threshold for detecting each individual component increased in the presence of an opponent stimulus and the new threshold for the far component is higher. If so, how to explain the higher threshold for the far component? Is it related to the creation and sensitivity theory of the attention modulation proposed by (Luo & Maunsell 2015)? Future study is required to examine this hypothesis.

For the neuronal response to stimulus moving at various depths, Smolyanskaya et al. (2013) have shown that the majority of MT neurons represent motion direction and binocular disparity in a separable manner. In the present study, we found that the direction tuning of the near component and the far component had the same selectivity but usually different amplitude which depends on the neuron's disparity selectivity -- the closer a stimulus component's depth is to a neuron's preferred disparity, the stronger is the amplitude of the direction tuning. This is consistent with the findings by Smolyanskaya et al. (2013). Then, how come both near-disparity preferred cells and far-disparity preferred cells preferentially represented Near component when both stimuli were presented together if the disparity selectivity is consistent across motion direction? What makes the far disparity preferred neuron bias toward the near stimulus component?

Previous studies (Prince et al 2002a, Tanabe et al 2005) have revealed that there are more neurons in area MT, V1 and V4 showing a preference for near disparities (crossed disparities) relative to the neurons preferring far disparities (uncrossed disparities). Considering the success of the divisive normalization model in accounting for the neuronal representation of transparent motion, this prevalence of Near disparity-preferred neurons might provide an explanation for the Near-dominance at both behavioral level and neuronal activity level. From the perspective of divisive normalization, the population size of a specific group of neurons can determine the weight of the information conveyed by that group of the neuron. In addition, from the perspective of top-down processing, a larger population size may translate to a higher prior rate associated with attention manipulation (Summerfield & Egnér 2009). However, the prevalence does not guarantee a higher weight in the normalization model, even a higher weight does not guarantee a

higher contribution to the final output of the whole population either, future study is required to test these hypotheses.

Along the same line, it is critical to understand where the near-dominance is first developed. Since V1 and V4 neurons are also found to be selective for the binocular disparity and show a stronger preference for near (crossed) disparities like MT neuron does (Tanabe et al. 2005), it would be interesting to further look at whether this Near-dominance in representing transparent motion is also present in the visual area V1 and V4. Also, for future study, we would like to investigate if the “side-bias” feature we discovered in the second project (chapter 3) is preserved in the representation of transparent motion stimuli moving at different depths. If so, how is it related to the attentional state and the Near-dominance we found in the present study?

The study by Bradley et al., (1995) has shown that adding disparity cues can reduce the inhibition by motion in the null direction and thus facilitate surface segregation. Further analysis can be performed to compare the neuronal responses evoked by two transparent stimuli moving at the same plane versus moving at different planes.

Tuned feature attention effect

In the present study, we found that the feature attention effect was tuned to both motion direction and binocular disparity – the strongest attention effect was observed when the near component moved near the PD of the recorded neuron. This is consistent with the results from a previous study by Ni and her colleagues (Ni et al 2012), in which they found attention to the preferred stimulus in the presence of the null stimulus significantly increased the bi-directional responses while attending to null stimulus in the presence of preferred stimulus did not lower the response much. They used tuned normalization model to explain this symmetric attention modulation. Based on this, we can understand why a significant attention

effect was only observed when the animal attended to a stimulus component moving near the PD of the neuron being recorded, but how can we account for the asymmetric attention modulation between near and far component? Even for the far-preferred cells and attention was directed to the far component, we also observed stronger attention modulation when the near component moved near the PD (**Table 4.2 and Table 4.3**). This may indicate that the computation underlying the near dominance has been completed pre-attentively. The binocular disparity perception may further define an internal ‘saliency map’ to guide the deployment of visual attention and possibly eye movement. It would be interesting to measure the dynamics of the binocular eye convergence over the whole duration of each trial to figure out which plane the animal actually fixated at and the strategy the animal used in performing the task. Further, it is also interesting to compare the spatial attention effect on near plane versus on far plane when they are presented alone to understand whether it is the context of “near-far” structure that causes the near dominance or the near component inherently is more salient and attracts more attention.

Is the facilitation of segmentation by adding disparity related to attention?

The facilitation effect by adding the depth order in the representation of transparent motion has been illustrated previously at both behavioral level and neuronal activity level (Bradley et al. 1995). However, in this study, we found that the attention-switch effect is much weaker relative to the near-dominance effect in terms of firing rate modulation. From the perspective of decoding, the near-dominance seems to play a more critical role in facilitating motion segmentation in this context. We also demonstrated that the near-dominance is independent of the attentional state and develops in less than 150ms after the motion onset, indicating the near-dominance is likely developed pre-attentively, but further study is required to confirm this.

One interesting finding is that our animal is not better at detecting motion directions when angle separation is larger which is supposed to be easier to be separated, suggesting the motion segmentation is different from motion detection. In particular, between 60° separation and 120° separation, the animal showed similar correct rate in correctly detecting the cued motion direction and similar error rate in reporting the ignored motion direction, indicating that the animal was equally good at separating near plane from far plane to learn the cue in both angle separation conditions. The animal's correct rate in detecting the cued motion direction is controlled by his accuracy in detecting and reporting the motion direction of a single stimulus. Our task was designed to control animal's attention and measure the attention effect on motion discrimination. But to understand the causal role of attention on segmenting objects, a new experiment is needed to be able to turn on and off attention.

Segmentation is a dynamic process including the initial exploratory search and the consolidation during the later phase. Attention may exert a different impact on the behavioral performance and neuronal activity during different stages. So for future studies, in order to better understand the specific role of attention in assisting visual segmentation, it is necessary to design specific behavioral tasks that mimic the various stages or components of visual segmentation process to evaluate the various attention effect. Still taking the transparent motion perception as an example, we could split it into two phases – early phase of separating two overlapping motion vectors and late phase of detecting the velocity of each motion vector. As for the first phase/component, emphasis will be put on studying how attention helps separate two overlapping motion vectors, so it makes sense to just use two reporting targets. However, for the second phase of detecting the velocity of each vector, emphasis should be put on understanding how attention facilitates the accurate coding and decoding of the external motion vector, so the task should be a detection task in the presence of multiple choices (reporting targets). Other than this, the certainty and reaction time of the behavioral performance should also be treated separately in segmentation task. Additional questions

to address in the future would be: How does attention influence the animal's certainty of their perceptual choice and the reaction time? And how is that related to final behavioral performance?

Interaction between top-down and bottom-up pathways

The present study shows that MT neurons, in response to transparent motion stimuli moving at different depths, preferentially represent the near component regardless of the attentional state or the neuron's disparity selectivity. This near-dominance representation and its quick temporal development indicate the processing of the binocular disparity of the transparent motion stimuli is very likely to be a bottom-up process and independent of the top-down influence, such as attentional modulation. In addition to the near-dominance, we observed a rate modulation effect by feature attention which is tuned to both motion direction and binocular disparity. A significant attention effect was only observed when the attended stimulus moved on the near plane and in the PD of the recorded neuron.

Regarding the interaction between the top-down attention effect and the likely bottom-up near-dominance representation, our finding that attention effect is tuned to near disparity suggests that top-down attention modulation is influenced by the bottom-up processing or at least controlled by a common neural mechanism as the near-dominance representation does. It has been demonstrated that tuned normalization can well account for the various attention effect observed in area MT (Ni et al. 2012). Our finding that the feature attention effect on MT activity is tuned to motion direction is consistent with their finding that MT neurons showed stronger attention modulation when attention was directed to preferred stimulus relative to when attention was directed to non-preferred stimulus. Their study also shows MT neurons with strong divisive normalization behavior tend to show strong attention effect. Interestingly, the near-dominance we observed in our study is likely to be explained by the divisive normalization given the fact that there are more near disparity preferred neurons than far disparity preferred in area MT which might serve as a

biological basis for divisive normalization. Therefore, the attention effect observed in our study is very likely rooted from the divisive normalization conducted in the early visual area. However, there is also more near disparity preferred neurons in area V1. To understand whether the near-dominance is achieved locally in area MT or inherited from area V1, future study is required to study the neuronal representation of transparent motion stimuli moving at different depths in area V1.

Regarding the impact of top-down attention modulating on the bottom-up processing, although we have shown that development of the near-dominance is independent of attentional state and the effect is completely evolved in less than 150-ms, suggesting that the near-dominance is a product of bottom-up processing, it's hard to rule out potential impact of attention on the temporal development of the near-dominance. First, it has been shown that the feature-based attention-related facilitation of neural activity begins as early as 90-120 ms after stimulus onset in area MT and V4 of human subjects (Schoenfeld et al 2007), while the modulations associated with spatial attention typically begin even earlier (Di Russo et al 2003, Martinez et al 1999, Martinez et al 2001). Secondly, neural representation bias has also been demonstrated in other areas, like V4 and IT (Hinkle & Connor 2002, Tanabe et al 2005, Uka et al 2000), where cells also show binocular disparity selectivity and even more cells prefer Near disparity than V1 and MT cells do. This bias representation of disparity in V4 could be related to the uncontrolled attentional effects (Hinkle & Connor 2002). Our finding that the presence of the near-dominance is independent of the attentional state -- regardless of whether the near component is attended or ignored provides a strong argument that near-dominance representation is developed pre-attentively. However, besides the controlled attention, the animal can develop its internal top-down anticipation for external stimuli based on the experience since both "cue-near" and "cue-far" trials are randomly interleaved in our paradigm. A near plane dominated internal anticipation may be generated from the successful experience in detecting near component, and therefore it may further cause the near-dominance in neuronal activity. To further evaluate if the near-dominance is developed pre-attentively or not, it is necessary to design a new task in which the

attention should be directed to the opposite hemisphere, such that the transparent motion stimuli are totally unrelated to the attention task.

Relationship between “near-bias” and “side-bias”

In chapter 3, we found that sub-populations of MT neurons preferentially represent one of two similar motion components by showing “side-bias”. How is that related to the “near-bias” observed in the present study? Although the transparent motion stimuli used in chapter 3 were located at the same depth, given the previous finding that transparent motion stimuli are usually perceived as one surface sliding on top of another, would it be possible that the component perceived as moving at the near plane is preferentially represented? To examine this hypothesis, we can train the animal to report the motion direction of the component perceived as moving at the near plane and examine whether the neuronal activity also shows bias toward the near-perceived motion component.

In summary, the present study provided new insights on the neural representation of multiple stimuli moving transparently at different depths and established a neural basis for the perceptual dominance of the near-surface in motion transparency. The dominant representation of the near component was consistent across cells and attentional states, suggesting that a common computational strategy is implemented pre-attentively across neuronal populations. Furthermore, we found that attention directed to the surface was tuned for both motion direction and binocular disparity, suggesting that attention modulation is dependent on the bottom-up stimulus-driven visual processing.

References

- Bradley DC, Qian N, Andersen RA.** 1995. Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature* 373: 609-11
- Carrasco M.** 2011. Visual attention: the past 25 years. *Vision Res* 51: 1484-525
- Carrasco M, Loula F, Ho YX.** 2006. How attention enhances spatial resolution: evidence from selective adaptation to spatial frequency. *Percept Psychophys* 68: 1004-12
- Cormack LK, Stevenson SB, Landers DD.** 1997. Interactions of spatial frequency and unequal monocular contrasts in stereopsis. *Perception* 26: 1121-36
- Di Russo F, Martinez A, Hillyard SA.** 2003. Source analysis of event-related cortical activity during visuo-spatial attention. *Cerebral Cortex* 13: 486-99
- Greenwood JA, Edwards M.** 2006. Pushing the limits of transparent-motion detection with binocular disparity. *Vision Res* 46: 2615-24
- Hibbard PB, Bradshaw MF.** 1999. Does binocular disparity facilitate the detection of transparent motion? *Perception* 28: 183-91
- Hinkle DA, Connor CE.** 2002. Three-dimensional orientation tuning in macaque area V4. *Nature Neuroscience* 5: 665-70
- Luo TZ, Maunsell JHR.** 2015. Neuronal Modulations in Visual Cortex Are Associated with Only One of Multiple Components of Attention. *Neuron* 86: 1182-88
- Martinez A, Anillo-Vento L, Sereno MI, Frank LR, Buxton RB, et al.** 1999. Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience* 2: 364-69
- Martinez A, DiRusso F, Anillo-Vento L, Sereno MI, Buxton RB, Hillyard SA.** 2001. Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision Research* 41: 1437-57
- McMains S, Kastner S.** 2011. Interactions of Top-Down and Bottom-Up Mechanisms in Human Visual Cortex. *Journal of Neuroscience* 31: 587-97
- Ni AM, Ray S, Maunsell JH.** 2012. Tuned normalization explains the size of attention modulations. *Neuron* 73: 803-13
- 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: 891-97
- 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, Adelson EH.** 1994. Transparent Motion Perception as Detection of Unbalanced Motion Signals .1. Psychophysics. *Journal of Neuroscience* 14: 7357-66
- Schoenfeld MA, Hopf JM, Martinez A, Mai HM, Sattler C, et al.** 2007. Spatio-temporal analysis of feature-based attention. *Cerebral Cortex* 17: 2468-77

- Sharon E, Galun M, Sharon D, Basri R, Brandt A.** 2006. Hierarchy and adaptivity in segmenting visual scenes. *Nature* 442: 810-3
- Smolyanskaya A, Ruff DA, Born RT.** 2013. Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable. *Journal of Neurophysiology* 110: 2806-16
- Snowden RJ, Rossiter MC.** 1999. Stereoscopic depth cues can segment motion information. *Perception* 28: 193-201
- Stelmach LB, Tam WJ.** 1998. Stereoscopic image coding: Effect of disparate image-quality in left- and right-eye views. *Signal Process-Image* 14: 111-17
- Summerfield C, Egner T.** 2009. Expectation (and attention) in visual cognition. *Trends Cogn Sci* 13: 403-9
- 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, Maunsell JHR.** 1996. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 382: 539-41
- Uka T, Tanaka H, Yoshiyama K, Kato M, Fujita I.** 2000. Disparity selectivity of neurons in monkey inferior temporal cortex. *Journal of Neurophysiology* 84: 120-32
- Wannig A, Rodriguez V, Freiwald WA.** 2007. Attention to surfaces modulates motion processing in extrastriate area MT. *Neuron* 54: 639-51
- Yeshurun Y, Carrasco M.** 1998. Attention improves or impairs visual performance by enhancing spatial resolution. *Nature* 396: 72-5

Appendixes

Appendix 1

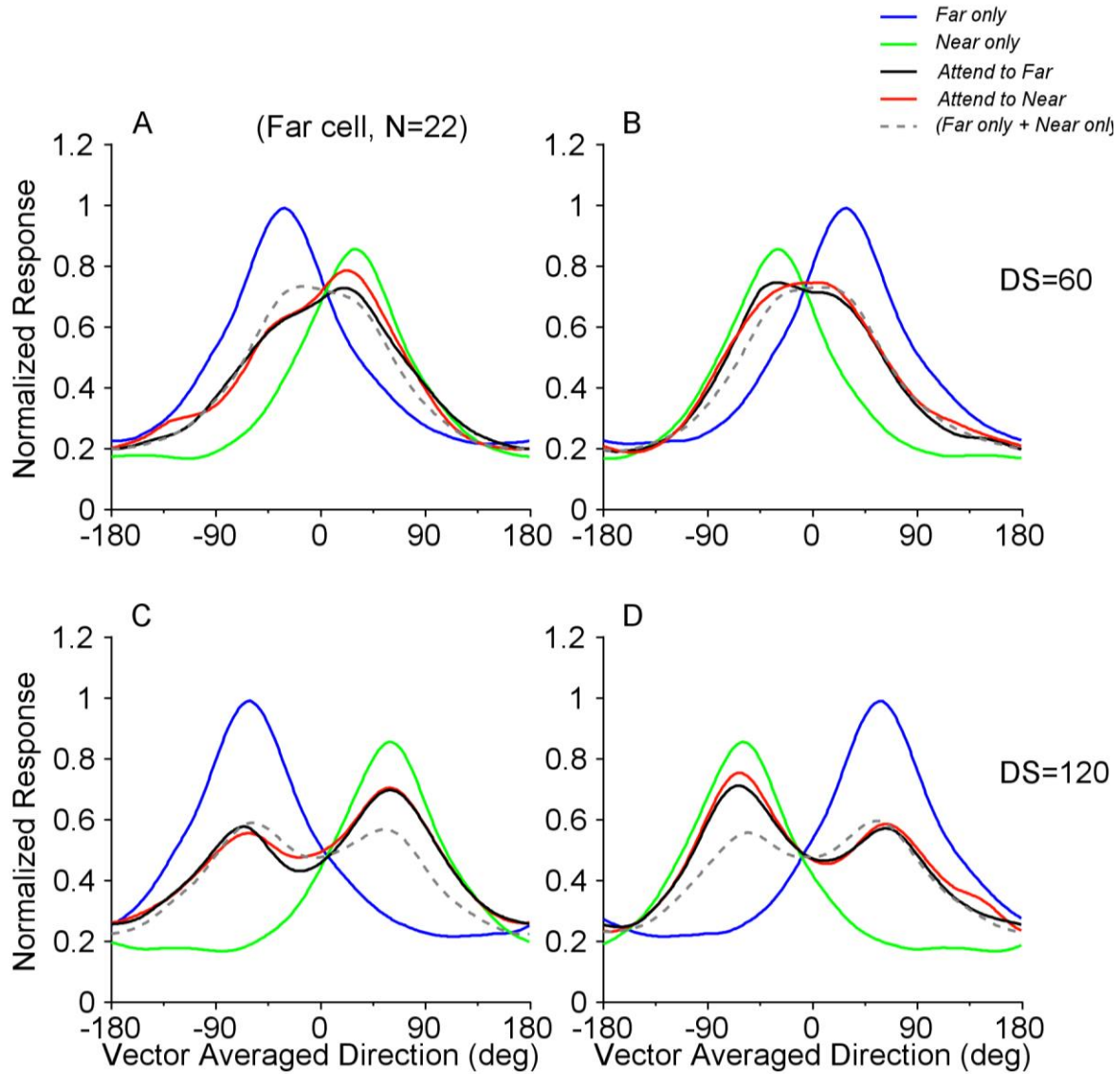


Figure1: Response tuning curves averaged across neurons that preferred the far disparity. The bi-directional response tuning is biased toward the response tuning to the near component. **A.** The transparent motion stimuli were separated by 60° and the near component moved at the relative C-side; **B.** stimuli were separated by 60° and the near component moved at the relative CC-side. **C.** stimuli were separated by 120°, and the near component moved at the relative C-side; **D.** stimuli were separated by 120° and the near component moved at the relative CC-side. The gray curve is the average of the two component response tuning curves.

Appendix 2

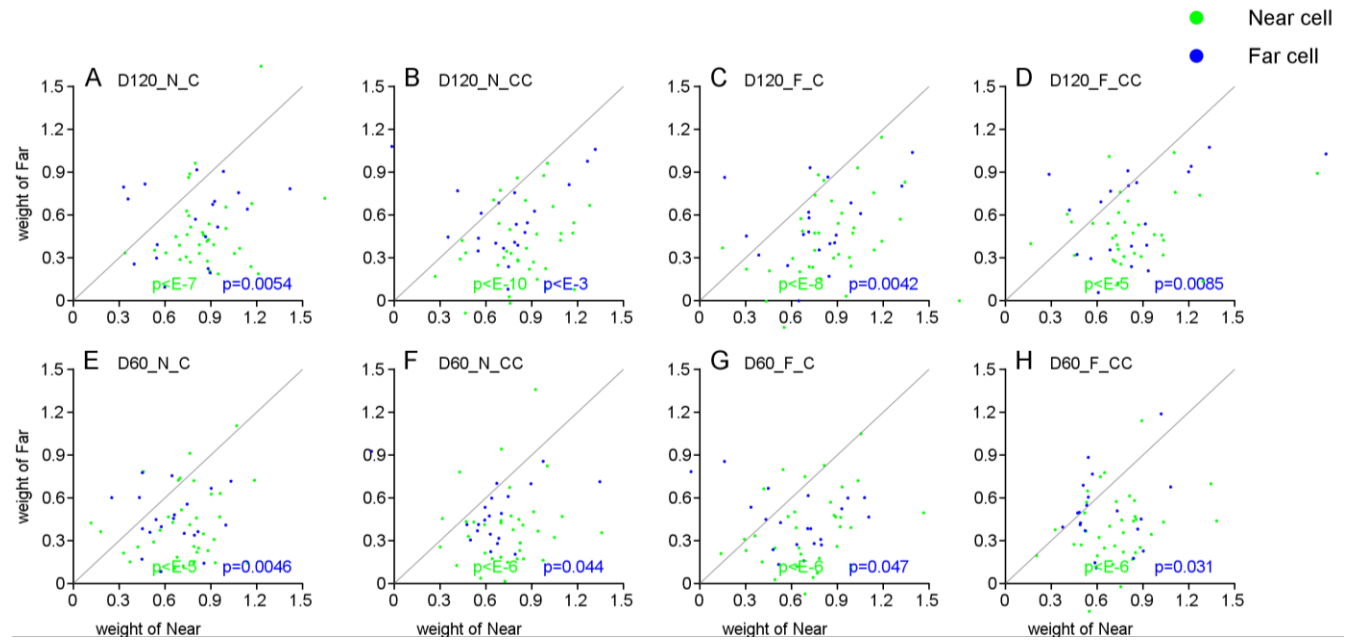


Figure2. Distributions of the response weight for the two stimulus components. Weights were calculated from the SNL model fit. Weights for the near component were plotted vs the weights for the far component. Each dot represents one cell. Both near-preferred cells (shown in green) and far-preferred cells (shown in blue) showed significantly larger weights for the near component than the far component across all stimulus and attention conditions tested. Angle separation between the two components was 120° in the top row, and 60° in the bottom row. **A,E.** Attending to the near component and the near component moved at C-side; **B,F.** Attending near and the near component moved at CC-side; **C,G.** Attending far and the far component moved at C-side; **D,H.** Attending far and the far component moved at CC-side;

Appendix 3

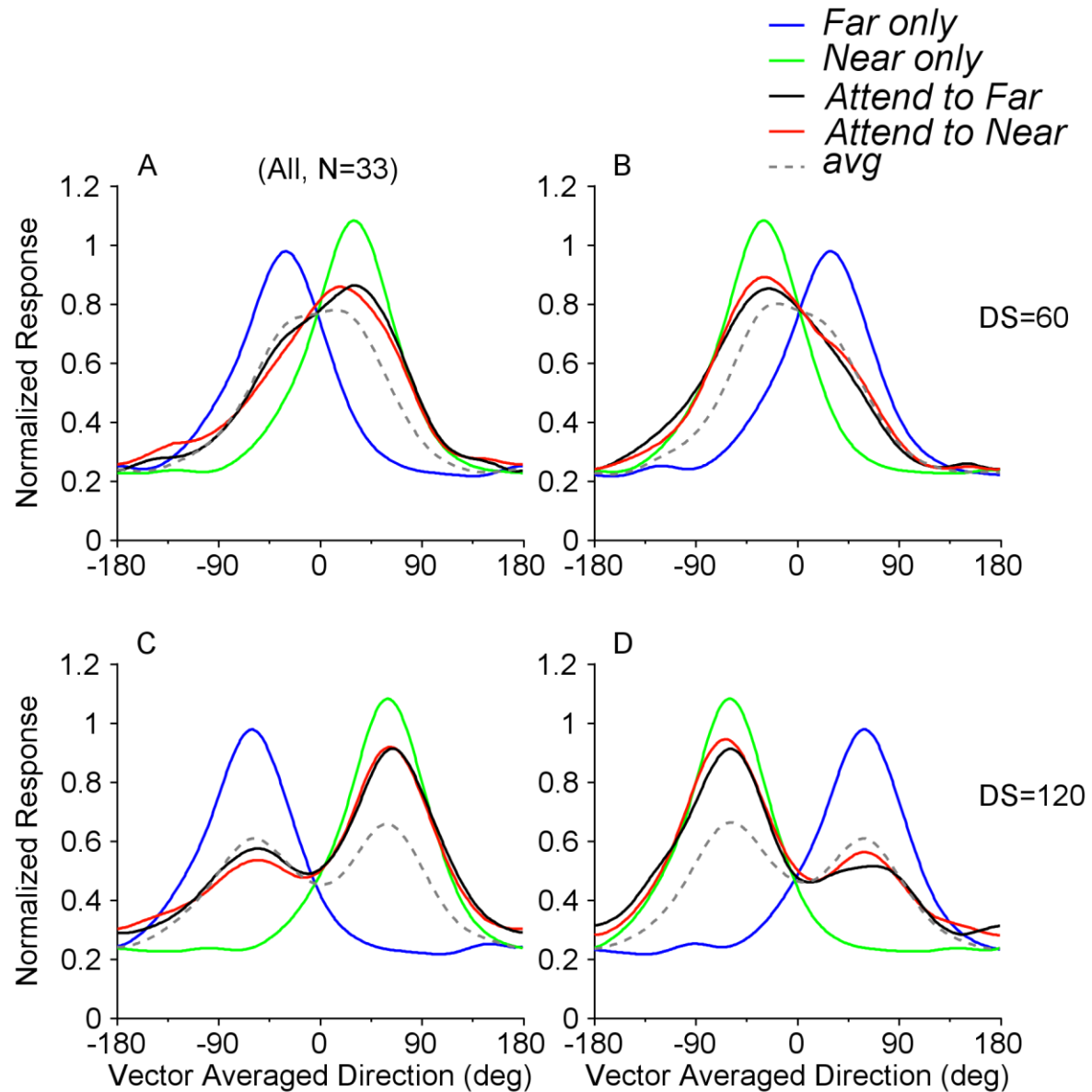


Figure 3: Population-averaged response tuning curves when the average correct rate of the animal was greater than or equal to 70%. The bi-directional response tuning is biased toward the response tuning of the near component. **A.** The transparent motion stimuli were separated by 60° and the near component moved at the relative C-side; **B.** The components were separated by 60° and the near component moved at the relative CC-side. **C.** The components were separated by 120°, and the near component moved at the relative C-side; **D.** The components were separated by 120°, and the near component moved at the relative CC-side. The gray curve is the average of the two component response tuning curves.