Neural encoding of multiple motion speeds in visual cortical area MT

Xin Huang, Bikalpa Ghimire, Anjani Sreeprada Chakrala, Steven Wiesner

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

Correspondence should be addressed to:

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

Keywords: *neural encoding, segmentation, discrimination, speed tuning, velocity, transparent motion, divisive normalization, figure-ground segregation, natural scenes*

Running title: *Neural encoding of multiple motion speeds*

Acknowledgment

We thank Dr. Steven Lisberger for his support in the early phase of this project, Emily Ausloos and Jianbo Xiao for data collection in early human psychophysics experiments, Bryce Arseneau for animal training, Drs. Jennifer Coonen and Kevin Brunner at the Wisconsin National Primate Research Center for excellent veterinary care and surgical assistance, and Drs. Emily Cooper and Greg DeAngelis for valuable comments on the manuscript.

Grant support

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

Conflict of Interest

None.

ABSTRACT

Segmenting objects from each other and their background is critical for vision. The speed at which objects move provides a salient cue for segmentation. However, how the visual system represents and differentiates multiple speeds is largely unknown. Here we investigated the neural encoding of multiple speeds of overlapping stimuli in the primate visual cortex. We first characterized the perceptual capacity of human and monkey subjects to segment spatially overlapping stimuli moving at different speeds. We then determined how neurons in the motion-sensitive, middle-temporal (MT) cortex of macaque monkeys encode multiple speeds. We made a novel finding that the responses of MT neurons to two speeds of overlapping stimuli showed a robust bias toward the faster speed component when both speeds were slow ($\leq 20^{o}$/s). The faster-speed bias occurred even when a neuron had a slow preferred speed and responded more strongly to the slower component than the faster component when presented alone. The faster-speed bias emerged very early in neuronal response and was robust over time and to manipulations of motion direction and attention. As the stimulus speed increased, the faster-speed bias changed to response averaging. Our finding can be explained by a modified divisive normalization model, in which the weights for the speed components are proportional to the responses of a population of neurons elicited by the individual speeds. Our results suggest that the neuron population, referred to as the weighting pool, includes neurons that have a broad range of speed preferences. As a result, the response weights for the speed components are determined by the stimulus speeds and invariant to the speed preferences of individual neurons. Our findings help to define the neural encoding rule of multiple stimuli and provide new insight into the underlying neural mechanisms. The faster-speed bias would benefit behavioral tasks such as figure-ground segregation if figural objects tend to move faster than the background in the natural environment.

INTRODUCTION

Neuroscientists have been investigating how neurons in the brain represent sensory information for decades. Previous studies were often concerned with the neural coding of a single visual stimulus. However, natural environments are abundant with multiple entities that often co-occupy visual neurons' receptive fields (RFs). Segmenting visual objects from each other and their background is a fundamental function of vision (Braddick, 1993). However, how the visual system represents multiple visual stimuli to achieve segmentation is not well understood. As the field progresses to unravel visual processing in natural vision, it becomes increasingly important to understand the principles of neural coding of multiple visual stimuli.

Visual motion provides a salient cue for scene segmentation. Common motion helps to group elements into the same object, whereas entities moving at different velocities can often be segregated from each other. An object moving at a speed different from its background, for example, is easier to segment. Here we investigated how the primate visual system represents multiple motion speeds. The extrastriate middle-temporal cortex (area MT) is important for motion processing and motion-based segmentation (Allman et al., 1985; Britten 2003; Born and Bradley 2005; Pasternak et al., 2020; Born et al., 2000; Huang et al., 2007, 2008). To investigate the neural representation of multiple moving stimuli, it is advantageous to start with overlapping stimuli so the effects of motion cues can be isolated from spatial cues. Segmentation of overlapping stimuli moving at different directions and speeds gives rise to the perception of transparent motion (Braddick, 1997; Braddick et al, 2002; Mestre et al., 2001; Masson et al., 1999). Previous studies have investigated how neurons in area MT represent two motion directions of transparently moving stimuli (Snowden et al., 1991; Qian and Andersen, 1994; McDonald et al., 2014; Xiao et al., 2014; Xiao and Huang, 2015; Wiesner et al., 2020; Stoner and Albright, 1992; Krekelberg and van Wezel, 2013). Although how cortical neurons represent the speed of a single stimulus has been well-studied (Maunsell and van Essen, 1983; Lisberger and Movshon, 1999; Nover et al., 2005; Pack et al., 2005; Krekelberg et al., 2006a; Perrone and Thiele, 2001; Priebe et al., 2003, 2006; Liu and Newsome 2003), how neurons represent multiple speeds is largely unknown.

In characterizing how MT neurons represent multiple directions of transparently moving stimuli, we have previously shown that many neurons do not pool two directions equally, but weigh one direction more than the other (Xiao and Huang, 2015). We have also found that some MT neurons show response nonlinearity in pooling two motion directions in a way that better represents the individual direction components. The heterogeneous response weights and response nonlinearity in representing multiple directions can benefit the neural coding of multiple stimuli (Orhan and Ma, 2015; Xiao and Huang, 2015), and may constitute an optimal population representation of visual motion with multiple directions (Huang et al., 2017). Unlike two motion directions for which the individual component directions appear to be balanced and symmetrical in perceptual quality and salience, visual stimuli moving at two speeds appear to be asymmetrical – one slower and one faster. The goal of this study is to determine the neural coding principle for multiple speeds of overlapping stimuli. It is conceivable that the responses of MT neurons elicited by two motion speeds may follow one of the following rules: 1) averaging the responses elicited by the individual speed components; 2) bias toward the speed component that elicits a stronger response, i.e. "soft-max operation" (Riesenhuber and Poggio, 1999); 3) bias toward the slower speed component, which may better represent the more probable slower speeds in nature scenes (Weiss et al., 2002); 4) bias toward the faster speed component, which may benefit the

segmentation of a faster-moving stimulus from a slower background. We also asked whether the encoding rule was dependent on the stimulus speeds and the speed preference of the individual neurons.

We first characterized the perception of overlapping stimuli that moved simultaneously at two speeds. Our results showed that human and monkey subjects can segment overlapping stimuli based only on speed cues. The performance was better when the separation between two stimulus speeds was larger and the ability of speed segmentation was reduced when stimulus speeds were fast. We next recorded neuronal responses from area MT of male macaque monkeys. We made a novel finding that MT neurons showed a strong faster-speed bias when stimulus speeds were slow and as stimulus speeds increased, the faster-speed bias gradually shifted to response averaging. We also showed that a classifier could differentiate a two-speed stimulus from a single-speed stimulus based on MT responses, in a way generally consistent with perception. We proposed a model in which each speed component was weighted by the responses of a population of neurons with a broad range of speed preferences elicited by that speed component. The model extends the standard divisive normalization and can well explain our results. This study helps to fill a gap in understanding the neural coding principle of multiple visual stimuli and provides new insight into the mechanism underlying the neural representation of multiple stimuli and scene segmentation.

RESULTS

Perception of overlapping stimuli moving at different speeds

Human psychophysics

To establish the perceptual basis for our study, we first characterized how human subjects perceived overlapping stimuli moving at different speeds. We used similar visual stimuli in our psychophysics experiments as in our neurophysiology experiments. We asked how perceptual segmentation was impacted by the separation between two stimulus speeds, and as the mean stimulus speed changed from slow to fast.

The visual stimuli were two overlapping random-dot patches presented within a stationary square aperture 10° wide and centered at 11° eccentricity. The random dots translated within the aperture in the same direction at two different speeds. It has been suggested that the neural representation of speed in the visual cortex is encoded on a logarithmic scale (Maunsell and van Essen, 1983; Lisberger and Movshon, 1999; Nover et al., 2005), so we used a fixed ratio between two speeds, which gave rise to a fixed speed difference in the logarithmic scale. One set of stimuli had a "large speed separation", and the speed of the faster component was four times (x4) that of the slower component. The five speed pairs used were 1.25 and 5°/s, 2.5 and 10°/s, 5 and 20°/s, 10 and 40°/s, and 20 and 80°/s (Fig. 1B1). Another set of stimuli had a "small speed separation", and the speed ratio was two (x2). The five speed pairs were 1.25 and 2.5°/s, 2.5 and 5°/s, 5 and 10°/s, 10 and 20°/s, and 20 and 40°/s (Fig. 1B2). Experimental trials of bi-speed stimuli that had large and small speed separations were randomly interleaved.

Human subjects first performed a standard two-alternative-forced-choice (2AFC) task to discriminate a bi-speed stimulus from the corresponding single-speed stimulus that moved at the

log mean speed of the two component speeds. In each trial, the bi-speed and single-speed stimuli were presented in two consecutive time intervals in a random and balanced order (Fig. 1A). At large (x4) speed separation, all four subjects could perform the task well when the component speeds were less than 20 and 80°/s (Fig. 1C1). At 20 and 80°/s, the discrimination performance was poor (mean *d'* = 0.74, standard error STE = 0.5), indicating that subjects could not segment the speed components. At the small (x2) speed separation, the discriminability was worse than at the x4 separation. When the component speeds were less than 20 and 40°/s, subjects on average could differentiate the bi-speed stimulus from the single-speed stimulus (*d'* > 1.5), but not when speeds were at 20 and 40°/s (mean *d'* = 0.17, STE = 0.1) (Fig. 1C2).

In the standard 2AFC task, it is possible that subjects could not segment the bi-speed stimulus into two separate speeds, but were still able to differentiate the bi-speed from single-speed stimuli based on their appearances (e.g., the distribution of the random dots of the bi-speed stimulus may appear less uniform). Because our goal was to measure discriminability based on perceptual segmentation, we designed a novel 3AFC task to address this concern. In the modified task, subjects still discriminated the bi-speed stimulus from the corresponding single-speed stimulus but had the option to make a third choice on trials when they thought neither stimulus interval appeared to contain two speeds ("no two-speeds" choice) (Fig. 1A). Panels D1 and D2 of Figure 1 show the percentage of trials in which subjects made the no two-speeds choice (NTC). At x4 speed separation, the percentage of NTC was low at most of the speed pairs, except at the highest speeds of 20 and 80°/s, subjects reported they could not see two speeds in most of the trials (Fig. 1D1). At x2 speed separation, the percentage of NTC showed a U-shape as a function of the stimulus speed, and was near 100% at 20 and 40°/s (Fig. 1D2). These results confirmed that human subjects had difficulty segmenting two speeds when stimulus speeds were high. In addition, at low stimulus speeds with a small (x2) speed separation, subjects tended to perceive only one speed (Fig. 1D2). We incorporated the NTC into the *d'* calculation by evenly splitting the NTC trials into "hit" trials and "false alarm" trials (see Methods). In this way, the NTC trials were accounted for by *d'*, in the sense that they did not contribute to successful discrimination.

The *d'* from the 3AFC task were similar to those of the 2AFC task, with a slight reduction of *d'* across conditions as the NTC trials reduced discrimination performance (Fig. 1E1 vs. 1C1, 1E2 vs. 1C2). The small performance difference between the 2AFC and 3AFC tasks suggests that human subjects generally relied on speed segmentation to perform the 2AFC task. Based on the results from the 3AFC task, we performed a two-way ANOVA, in which the two factors were the mean speed of the stimulus components and the speed separation (x4 or x2). We found that both factors had significant effects. *d'* changed significantly with the mean stimulus speed ($F(4,30) = 26.8$, $p = 1.60\times10^{-9}$) and the *d'* at x4 separation differed significantly from that at x2 separation ($F(1,30) = 84.1$, $p = 3.29\times10^{-10}$). *d'* was higher at x4 (Fig. 1E1) than at x2 speed separation except at the fastest speeds of 20 and 80°/s vs. 20 and 40°/s (Fig. 1E2). Our results also showed that segmentation was significantly worse at fast speeds – *d'* dropped significantly as the stimulus speeds increased from 10 and 40°/s to 20 and 80°/s for x4 separation (one-way ANOVA, $F(1,6) = 38.6$, $p = 8.1\times10^{-4}$) (Fig. 1E1), and from 10 and 20°/s to 20 and 40°/s for x2 separation (one-way ANOVA, $F(1,6) = 32.7$, $p = 1.24\times10^{-3}$) (Fig. 1E2).

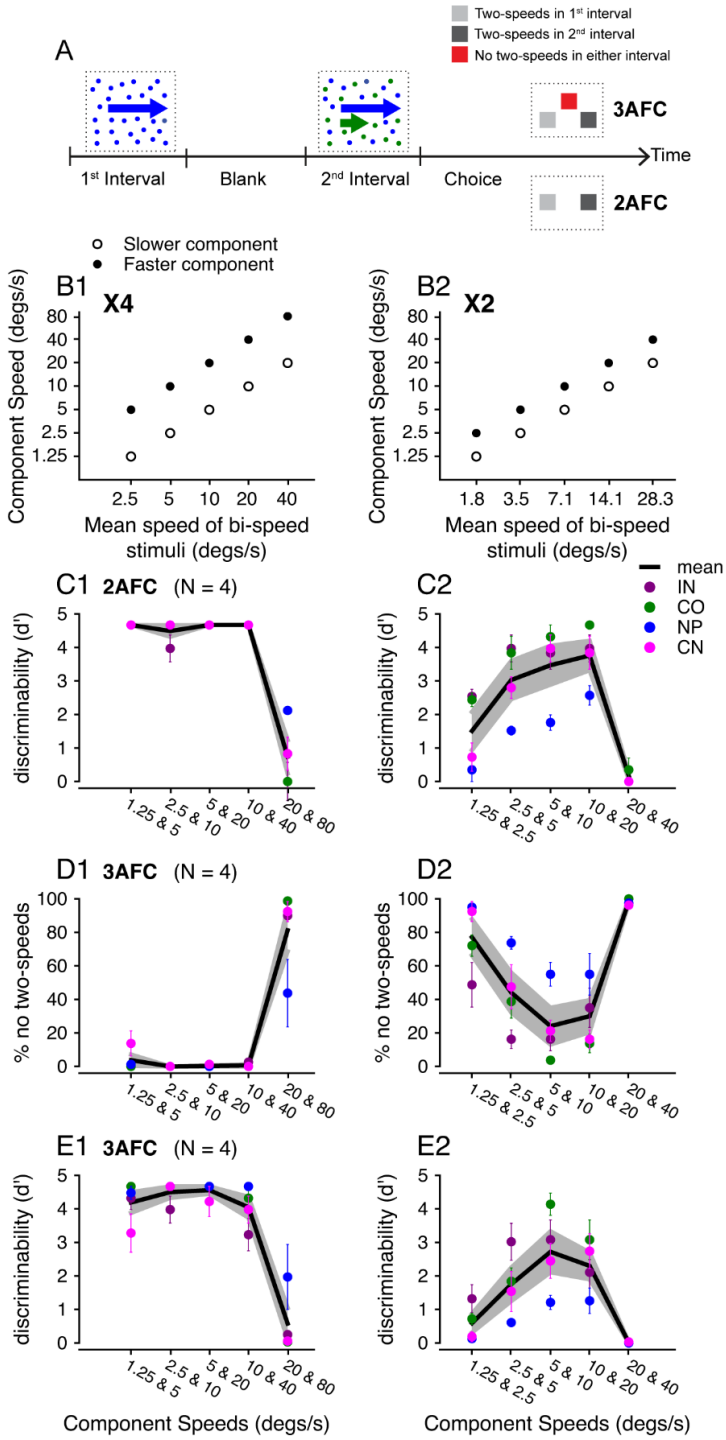


Figure 1. Psychophysical tasks and performance of human subjects. A. Illustration of the 2AFC and 3AFC tasks. **B.** Motion speeds of visual stimuli. The speeds of two stimulus components were plotted versus the log mean speed of each bi-speed stimulus. **C.** Discriminability of four human subjects performing a standard 2AFC task. **D.** In the 3AFC task, the percentage of trials that human subjects reported "no two-speeds". **E.** Discriminability of the same subjects performing the 3AFC task. **B1-E1.** X4 speed separation. **B2-E2.** X2 speed separation. Each color represents data from one subject. The solid line shows the subject-averaged result. Error bars and error bands represent ±STE.

Monkey psychophysics

We next measured the monkey's ability to segment overlapping stimuli moving at two speeds. We trained one monkey to perform a 2AFC task to report whether a stimulus contained one or two speeds (Fig. 2A, see Methods). The monkey's performance at x2 speed separation (Fig. 2B2) was very similar in shape to that of humans (Fig. 1C2 of the 2AFC task). In addition, the monkey's performance was generally better at x4 separation than at x2 separation (Fig. 2B1 vs 2B2).

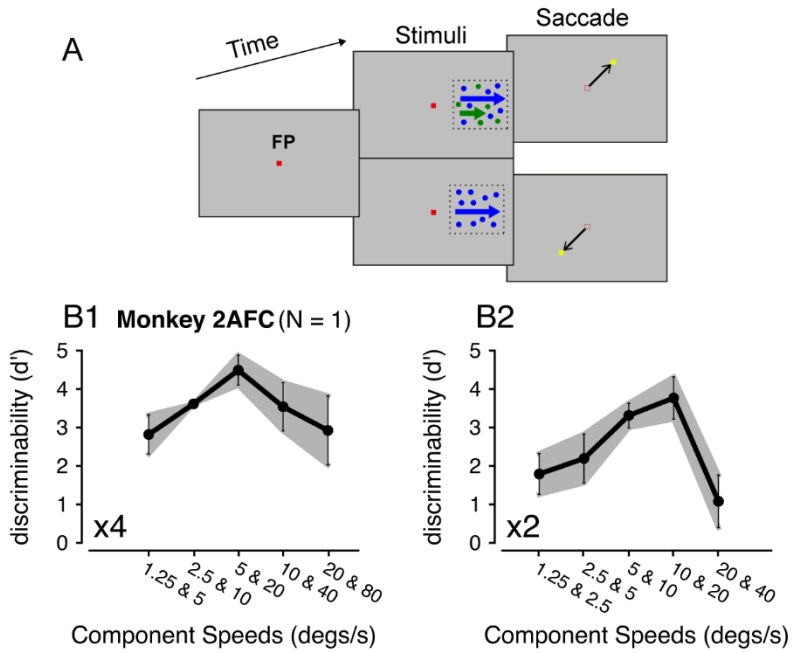


Figure 2. Monkey psychophysics. A. Behavioral task and visual stimuli. **B.** Discriminability of a monkey subject performing a 2AFC task. **B1**. X4 speed separation. **B2**. X2 speed separation. Error bars and error bands represent ±STE.

At x4 separation, the performance improved as the stimulus speeds increased from 1.25 and 5°/s to 5 and 20°/s. As the stimulus speeds increased from 5 and 20°/s to 20 and 80°/s, the performance declined (Fig. 2B1), similar to the human results (Fig. 1C1). However, the monkey was still able to differentiate the bi-speed and single-speed stimuli at the fastest speeds of 20 and 80°/s (Fig. 2B1), whereas the average human performance was poor (Fig. 1C1). Note that one human subject (NP) performed better than other subjects at 20 and 80°/s (mean *d'* = 2.12, STE = 0.12) (Fig. 1C1). The difference between the monkey and human results may be due to species differences or individual variability. It was also possible when it was difficult to segment two speeds at fast stimulus speeds, the monkey subject may resort to using the appearance (e.g. the apparent coherence) of the stimulus rather than speed segmentation to perform the task. Whereas human subjects were more likely to perform the task based on speed segmentation, as indicated by the similar results obtained using the 2AFC and 3AFC tasks.

Another notable difference between the monkey and human results was that, at low stimulus speeds of 1.25 and 5°/s, human subjects could differentiate the bi-speed stimulus from

the corresponding single-speed (2.5°/s) stimulus nearly perfectly. In comparison, the ability of the monkey subject to segment 1.25 and 5 °/s was lower (*d'* = 2.8, STE = 0.51), although still good (Fig 2B1 vs 1C1). This may be explained by how the monkey performed the task. For human subjects, while the motion of the faster component (5°/s) of the bi-speed stimulus appeared to be salient, it required effort to notice the very slow component (1.25°/s) to be moving rather than stationary. In some trials, the monkey may be able to segment the 5°/s component from the bi-speed stimulus but consider the slower component of 1.25°/s as stationary and, therefore, reported that the stimulus contained only one speed. Despite some differences between the human and monkey results, the two general trends – better segmentation performance at larger than smaller speed separation and reduced segmentation ability at very fast speeds were consistent across species.

Neuronal responses in MT elicited by bi-speed stimuli and single-speed components

To characterize how neurons in the visual cortex encode two overlapping stimuli moving at different speeds, we recorded extracellularly from 100 isolated neurons in the extrastriate area MT of two male monkeys (60 neurons from IM and 40 neurons from MO) while the monkeys performed a fixation task. Figure 3 shows the responses from four example neurons. To visualize the relationship between the responses to the bi-speed stimulus (red) and the constituent speed components, the plots of the response tuning curves to the slower (green) and faster (blue) components are shifted horizontally so that the responses elicited by the bi-speed stimulus and its constituent single-speed components are aligned along a vertical line as illustrated in Figure 3A1.

We found that the relationship between the responses elicited by the bi-speed stimulus and the constituent components depended on the stimulus speeds. Figure 3A1-D1 shows the results of four example MT neurons obtained when the separation between the two component speeds was large (×4). The component speeds were the same as the bi-speed stimuli used in the psychophysics experiments (Fig. 1B1, B2). When the two component speeds were slow (1.25 and 5°/s), the response to the bi-speed stimulus nearly followed the response elicited by the faster-speed component (the leftmost data points in Fig. 3A1-D1). Importantly, the response elicited by the bi-speed stimuli did not simply follow the stronger component response. When the preferred speed of a neuron was sufficiently low such that the response elicited by the faster component was weaker than that elicited by the slower component, the response to the bi-speed stimulus still followed the weaker response elicited by the faster component (Fig. 3A1). When the speeds of the two stimulus components were at 2.5 and 10°/s, the response elicited by the bi-speed stimulus was also biased toward the faster component, albeit to a lesser degree. As the mean speed of the two stimulus components increased, the bi-speed response became closer to the average of the two component responses (Fig. 3A1-D1). We found similar results when the speed separation between the two stimulus components was small (×2) (Fig. 3A2-D2).

We found the same trend in the neural responses averaged across 100 neurons (Fig. 4A). At ×4 speed separation, the population-averaged response showed a strong bias toward the faster component when the stimulus speeds were low and shifted toward the average of the component responses as the speeds increased (Fig. 4A1). To determine whether this trend held for neurons with different preferred speeds, we divided the neuron population into three groups with "low" (<2.5°/s), "intermediate" (between 2.5 and 25°/s), and "high" (>25°/s) preferred speeds. For 10

neurons that preferred low speeds, the response to the faster component was weaker than that to the slower component. However, the response to the bi-speed stimuli was strongly biased toward the faster component when the stimulus speeds were low (Fig. 4B1). This finding suggests that the bi-speed response is not biased toward the stimulus component that the neuron prefers when presented alone but biased toward the faster speed component.

For 61 neurons that preferred intermediate speeds (Fig. 4C1) and 29 neurons that preferred high speeds (Fig. 4D1), we also found a strong bias toward the faster speed component when the stimulus speeds were low, and a gradual change toward the average of the component responses as the stimulus speeds increased. At the lowest stimulus speeds of 1.25 and 5°/s, the bi-speed response was nearly identical to that elicited by the faster component, showing "faster-component-take-all". For neurons that preferred high speeds, faster-component-take-all was also found for the stimulus speeds of 2.5 and 10°/s (Fig. 4D1). We found similar results at x2 speed separation (Fig. 4A2-D2), although the effect is not as pronounced as x4 speed separation.

Relationship between the responses to bi-speed stimuli and constituent stimulus components

To quantify the relationship between the response elicited by the bi-speed stimuli and the corresponding component responses, we expressed the response R of a neuron elicited by two component speeds v_s *and* v_f as a weighted sum of the component responses R_s and R_f elicited by the slower (v_s) and faster (v_f) component speed, respectively:

$$R(v_s, v_f) = w_s(v_s, v_f) \bullet R_s + w_f(v_s, v_f) \bullet R_f \quad , \qquad (5)$$

$$R_s = f(v_s), \quad R_f = f(v_f),$$

in which, w_s and w_f are the response weights for the slower and faster component, respectively, and f is the speed-tuning function of the neuron in response to a single speed.

For three data points of R, R_s, and R_f, as long as $R_f \neq R_s$, R can always be expressed as:

$$R = \frac{R_f - R}{R_f - R_s} R_s + \frac{R - R_s}{R_f - R_s} R_f \quad , \qquad (6)$$

The response weights can be expressed as $w_s = \frac{R_f - R}{R_f - R_s}$, $w_f = \frac{R - R_s}{R_f - R_s}$, and w_s and w_f sum to 1. By this definition, if R were closer to one component response, that stimulus component would have a higher weight. Note that Equation 6 is not intended for fitting the response R using R_s and R_f, but rather to use the relationship among R, R_s, and R_f to define the weights for the faster and slower components. We aim to determine whether, and if so how, the response weights change with the stimulus speeds.

Using this approach to estimate the response weights for individual neurons can be inaccurate because, at each speed pair, the weights are determined only by three data points. Also, R_s and R_f can sometimes be similar so the denominator in Equation 6 could be close to zero. We therefore used the neuronal responses across the population to determine the response weights (Fig.

5). For each pair of stimulus speeds, we plotted (R - R_s) in the ordinate versus (R_f - R_s) in the abscissa. Figure 5A1-E1 shows the results obtained at ×4 speed separation. Across the neuronal population, the relationship between (R - R_s) and (R_f - R_s) is remarkably linear (Type II regression, R^2 ranged from 0.94 to 0.76, see Table 1), and can be well described as:

$$R - R_s = k(R_f - R_s) + b \ . \qquad (7)$$

Because all the regression lines in Figure 5 nearly go through the origin (i.e. intercept $b \approx 0$, Table 1), the slope k obtained from the linear regression approximates $\frac{R - R_s}{R_f - R_s}$, which is the response weight w_f for the faster component (Eq. 6). Therefore, for each pair of stimulus speeds, we can estimate the response weight for the faster component using the slope of the linear regression of the responses from the neuronal population. Our results showed that the bi-speed response changed progressively from a scheme of "faster-component-take-all" to “response-averaging” as the speeds of the two stimulus components increased (Fig. 5F1). We found similar results when the speed separation between the stimulus components was small (×2), although the bias toward the faster component at low stimulus speeds was not as strong as x4 speed separation (Fig. 5A2-F2 and Table 1).

So far we have shown that human and monkey subjects can segment two overlapping speeds, but the segmentation becomes harder when the mean stimulus speed is fast. We have also established that the neural representation of two speeds in area MT can be well described by weighted summation of the responses elicited by the individual speed components, and the weights change from faster-take-all to response-averaging as the mean stimulus speed increases. In the following sections, we will show that this neural encoding rule is robust over time and across different motion directions, and is not due to attention modulation. We will further show the result from a decoding analysis using a linear classifier that suggests a connection between the neural representation of multiple speeds in MT and perception. Finally, we will describe a normative model that captures the encoding rule.

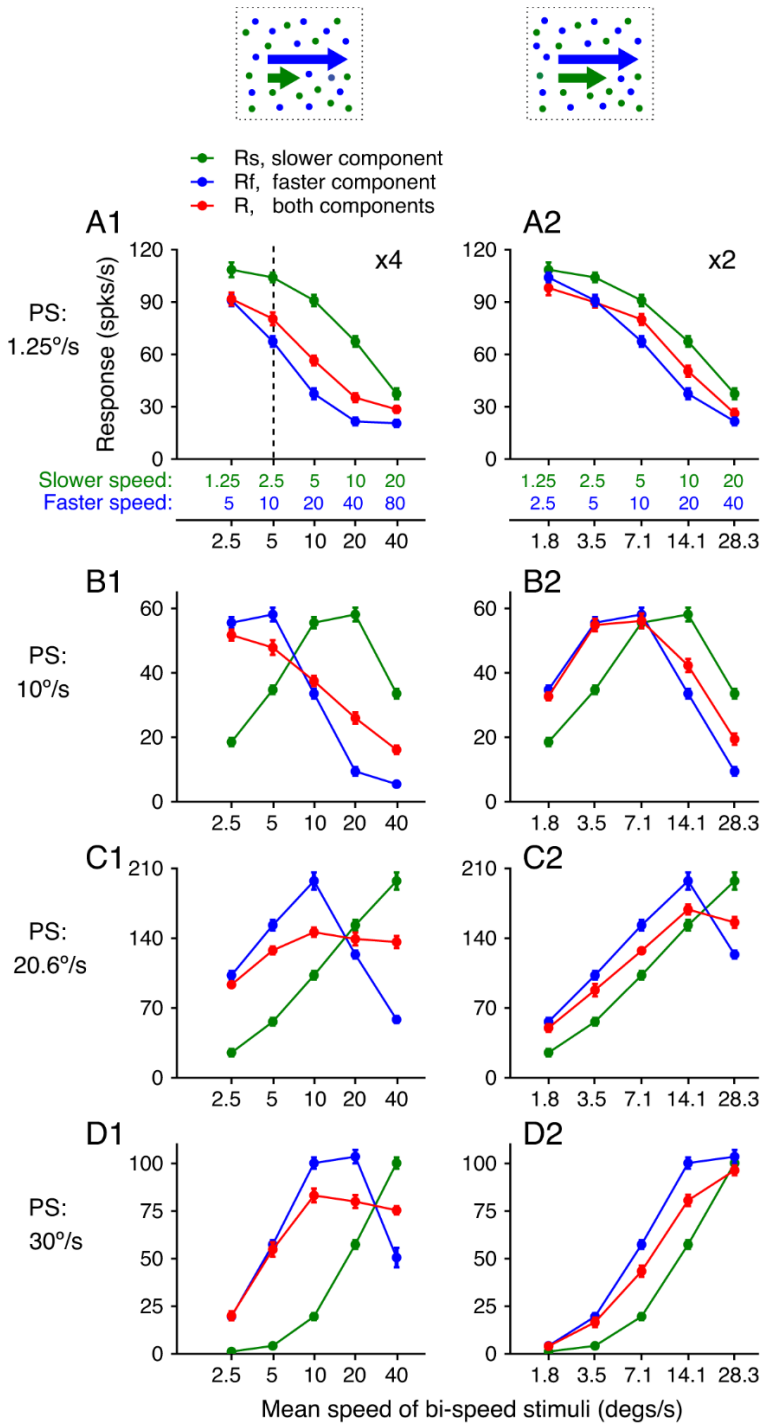


Figure 3. Speed tuning curves of four example neurons to bi-speed stimuli and constituent single-speed components. A. Illustration of the visual stimuli and the response tuning curves of an example neuron. Green and blue dots in the diagram indicate two overlapping achromatic random-dot patterns moving in the same direction at different speeds. Colors are used for illustration purposes only. The abscissas in green and blue show the speeds of the slower and faster components, respectively. The abscissa in black shows the log mean speed of the two speed components. **A-D.** Four example neurons are sorted by their preferred speeds (PS) from slow to fast. Error bars represent ±STE. For some data points, error bars were comparable to the symbol size. **A1-D1.** X4 speed separation. **A2-D2.** X2 speed separation.

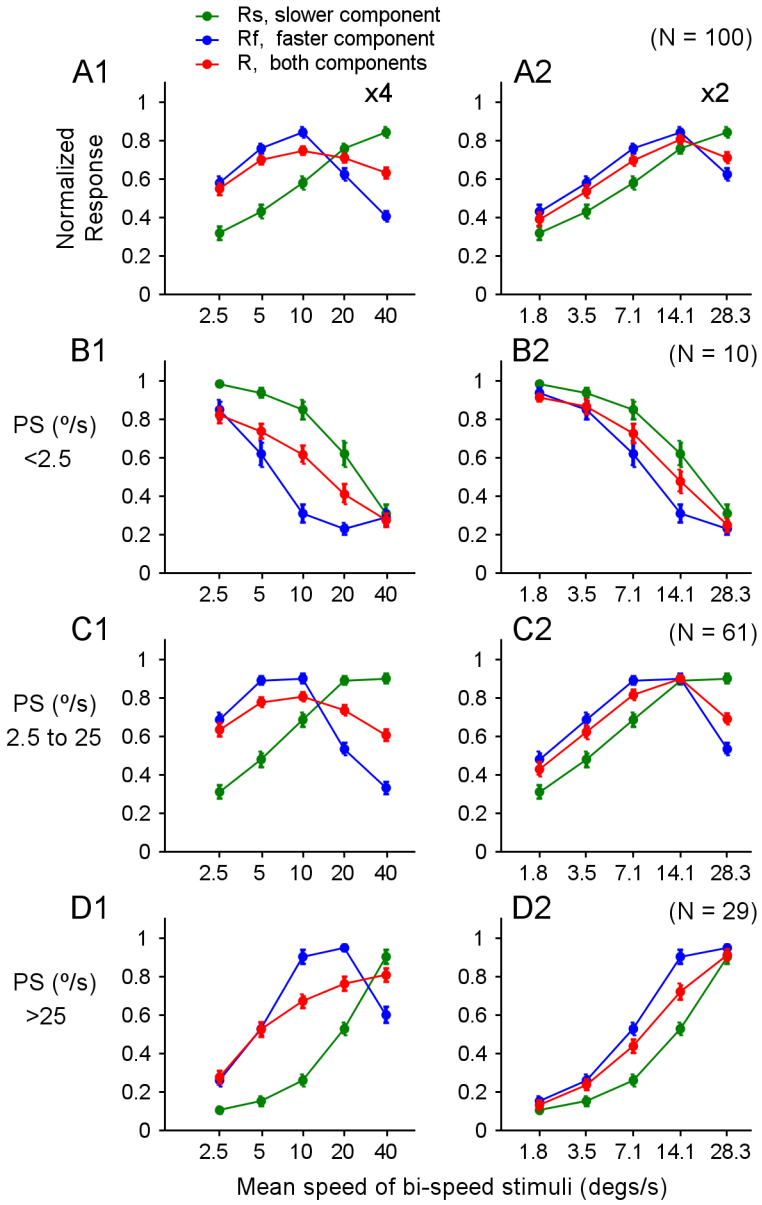


Figure 4. Population-averaged speed tuning curves to bi-speed stimuli and constituent single-speed components. Speed tuning curves averaged across **A.** 100 neurons in our dataset. **B.** 10 neurons that had PS lower than 2.5º/s. **C.** 61 neurons that had PS between 2.5 and 25º/s. **D.** 29 neurons that had PS greater than 25º/s. Error bars represent ±STE. For some data points, error bars were comparable to the symbol size. **A1-D1.** X4 speed separation. **A2-D2.** X2 speed separation.

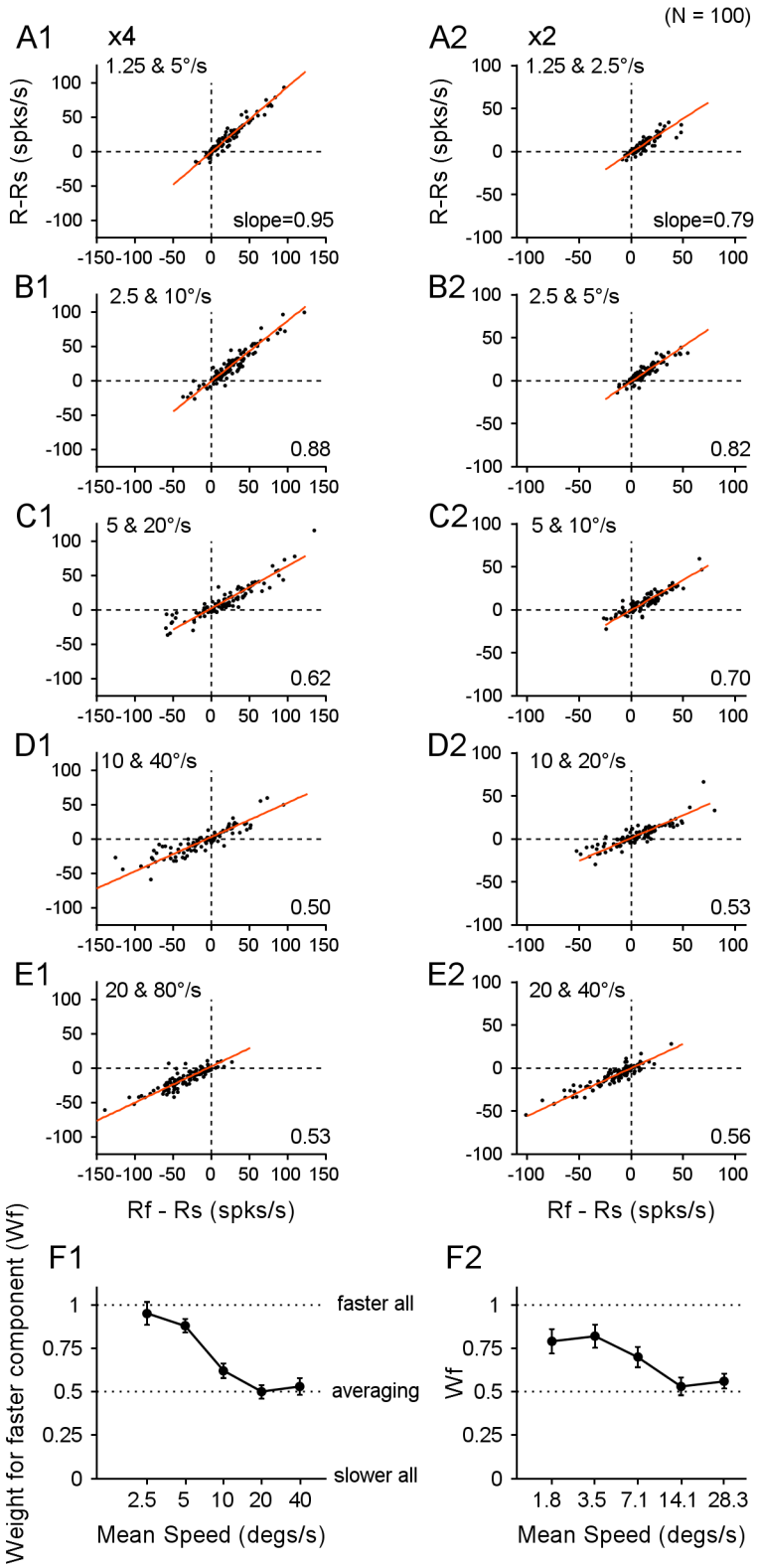

(N = 100)
A1
x4
1.25 & 5°/s
slope=0.95
A2
x2
1.25 & 2.5°/s
slope=0.79
B1
2.5 & 10°/s
0.88
B2
2.5 & 5°/s
0.82
C1
5 & 20°/s
0.62
C2
5 & 10°/s
0.70
D1
10 & 40°/s
0.50
D2
10 & 20°/s
0.53
E1
20 & 80°/s
0.53
E2
20 & 40°/s
0.56
R-Rs (spks/s)
Rf - Rs (spks/s)
F1
F2
faster all
averaging
slower all
Weight for faster component (Wf)
Wf
Mean Speed (degs/s)

Figure 5. Relationship between the responses to the bi-speed stimuli and the constituent stimulus components. A-E. Each panel shows the responses from 100 neurons. Each dot represents the response from one neuron. The ordinate shows the difference between the responses to a bi-speed stimulus and the slower component (R - R_s). The abscissa shows the difference between the responses to the faster and slower components (R_f - R_s). Type II regression line is shown in red. **F.** Response weights for the faster stimulus component obtained from the slope of the linear regression based on the responses of 100 neurons. **A1-F1**. X4 speed separation. **A2-F2.** X2 speed separation.

Table 1. Response weight for faster component based on linear regression (N = 100)

	Large Speed Difference (×4)					**Small Speed Difference (×2)**				
Components Speeds (°/s)	1.25 & 5	2.5 & 10	5 & 20	10 & 40	20 & 80	1.25 & 2.5	2.5 & 5	5 & 10	10 & 20	20 & 40
Slope (w_f) *and 95% Confidence Interval*	0.95 ± 0.097	0.88 ± 0.113	0.62 ± 0.094	0.50 ± 0.089	0.53 ± 0.104	0.79 ± 0.139	0.82 ± 0.134	0.70 ± 0.118	0.53 ± 0.100	0.56 ± 0.084
R^2	0.94	0.90	0.86	0.80	0.76	0.80	0.83	0.82	0.78	0.86
Intercept (b) *and 95% Confidence Interval*	-1.29 ± 2.161	-1.32 ± 2.739	1.48 ± 1.771	2.44 ± 1.115	2.00 ± 3.544	-1.48 ± 1.425	-1.32 ± 1.632	-0.56 ± 1.604	0.89 ± 0.532	-0.25 ± 1.508

Timecourse of MT responses to bi-speed stimuli

We asked whether the bias toward the faster speed component was robust over time. We also asked whether the faster-speed bias occurred early in the neuronal response or developed gradually over time. Figure 6 shows the timecourse of the response averaged across 100 neurons in the population. The bias toward the faster speed component occurred at the very beginning of the neuronal response when the stimulus speeds were less than 20º/s (Fig. 6A-C). The first 20-30 ms of the neuronal response elicited by the bi-speed stimulus was nearly identical to the response elicited by the faster component alone, as if the slower component were not present. The early dominance of the faster component on the bi-speed response cannot be explained by the difference in the response latencies of the faster and slower components. Faster stimuli elicit a shorter response latency (Lisberger and Movshon, 1999), which can be seen in Figure 6A-C. However, the bi-speed response still closely followed the faster component for a period of time after the

response to the slower component started to rise. The effect of the slower component on the bi-speed response was delayed for about 25 ms, as indicated by the arrows in Figure 6A-C. During the rest of the response period, the bias toward the faster component was persistent. As the stimulus speeds increased, the bi-speed response gradually changed to follow the average of the component responses (Fig. 6E). We found similar results when the speed separation between the two stimulus components was x4 (Fig. 6A1-E1) and x2 (Fig. 6A2-E2).

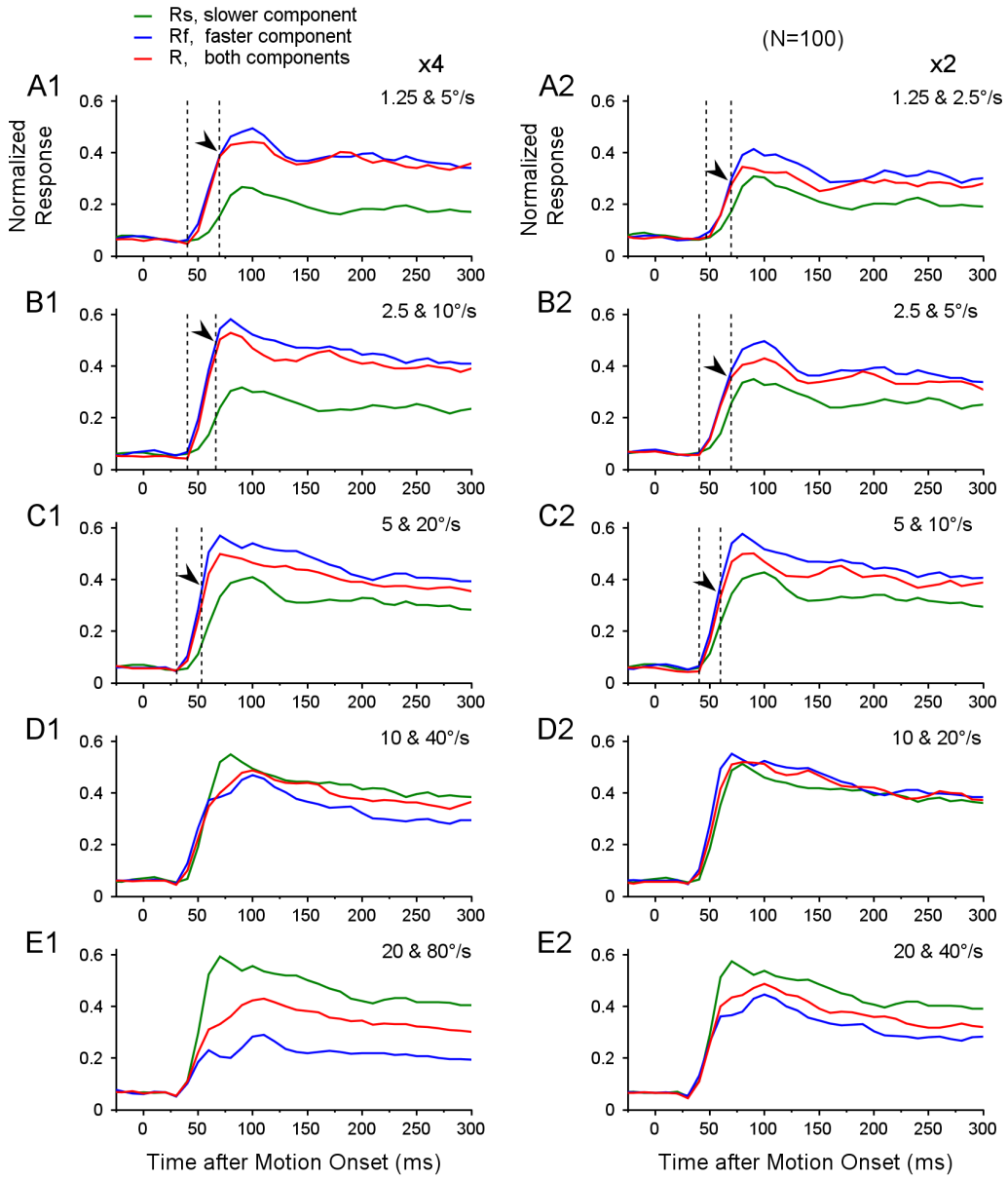


Figure 6. Timecourse of MT responses averaged across neurons to bi-speed stimuli. Peristimulus time histograms (PSTHs) were averaged across 100 neurons. The bin width of PSTH was 10 ms. **A1-E1.** X4 speed separation. **A2-E2.** X2 speed separation. In A-C, the left dash line indicates the latency of the response to a bi-speed stimulus, and the right dash line and the arrow indicate when the response to a bi-speed stimulus started to diverge from the response to the faster component.

Faster speed bias also occurs when stimulus components move in different directions

We showed that at low and intermediate speeds, MT response to the bi-speed stimulus was biased toward the faster stimulus component when two overlapping components moved in the same direction (at the preferred direction of the neuron). We asked whether this faster-speed bias also occurred when visual stimuli moved in different directions. We presented overlapping random-dot stimuli moving in two directions separated by 90° in the RF. The two stimulus components moved at different speeds. The speed of the stimulus component moving on the clockwise side of the two directions was 10°/s, whereas the speed of the other component was 2.5°/s. We varied the vector-average (VA) direction of the two component directions across 360° to characterize the direction tuning curves. Figure 7A shows the direction tuning curves averaged across 21 neurons (13 neurons from monkey RG, 8 neurons from monkey GE). The direction tuning curve of each neuron was first fitted with a spline and rotated such that the VA direction 0° was aligned with the neuron's preferred direction before averaging across neurons. The peak response to the faster component (Fig. 7A, blue curve) was stronger than that to the slower component (green curve). MT responses elicited by the bi-directional stimuli (red curve) showed a strong bias toward the faster component, more than expected by the average of the two component responses (black curve).

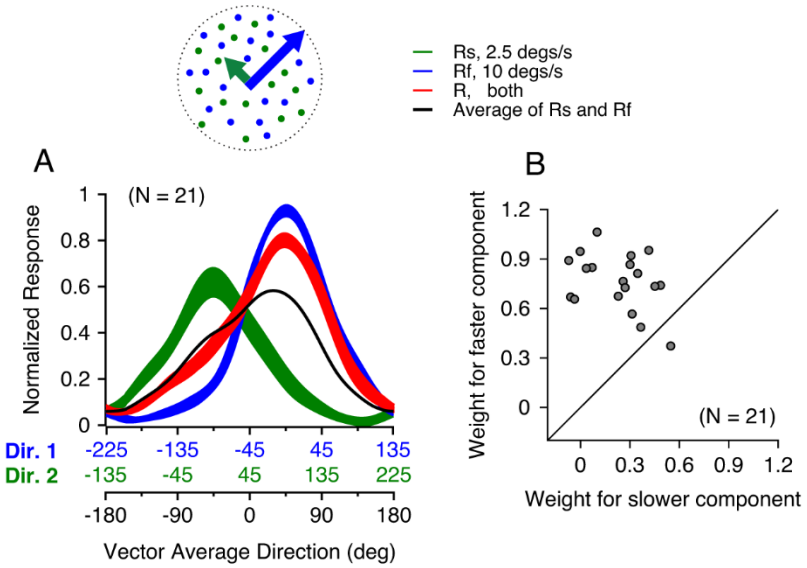


Figure 7. MT responses to bi-speed stimuli moving in different directions. A. Population-averaged direction tuning curves of 21 neurons in response to stimuli moving at two speeds and in two directions separated by 90º (red). The component direction Dir. 1 (blue) moved at 10º/s and the component direction Dir. 2 (green) moved at 2.5º/s. Dir. 1 was on the clockwise side of Dir. 2. The abscissas in blue and green show the directions of stimulus components Dir. 1 and Dir. 2, respectively. The blue and green axes are shifted by 90º relative to each other. The abscissa in black shows the corresponding VA direction of the two direction components. **B.** Response weights for the stimulus components obtained using a linear weighted summation fit. Each dot represents the response from one neuron.

We fitted the MT raw direction tuning curve to the bi-directional stimuli as a weighted sum of the direction tuning curves to the individual stimulus components moving at different speeds:

$$R_{bi}(\theta_1,\ \theta_2) = w_s \bullet R_s(\theta_1) + w_f \bullet R_f(\theta_2) + c\ , \qquad (8)$$

in which, R_s and R_f are the direction tuning curves to the slower and faster stimulus components, respectively; θ_1 and θ_2 are the motion directions of the two components; w_s and w_f are fitted response weights for the slower and faster components, respectively and they are not required to sum to 1. An implicit assumption of the model is that, at a given pair of stimulus speeds, the response weights for the slower and faster components are fixed across motion directions. The model fitted MT responses very well, accounting for an average of 90.3% of the response variance (std = 8.4%, N = 21). The median response weights for the slower and faster components were 0.26 and 0.74, respectively, and were significantly different (Wilcoxon signed-rank test, $p = 8.0 \times 10^{-5}$). For most neurons (20 out of 21), the response weight for the faster component was larger than that for the slower component (Fig. 7B). This result suggests that at low to intermediate speeds the faster-speed bias is a general phenomenon that applies to overlapping stimuli moving either in the same direction or different directions.

Faster-speed bias is not due to attention

We asked whether the faster-speed bias was due to bottom-up attention being drawn toward the faster stimulus component. To test this hypothesis, we recorded neural responses from one monkey (RG) as the animal directed attention away from the single- and bi-speed stimuli presented in the RFs. We trained the monkey to perform a demanding fine direction-discrimination task in the visual field opposite to that of the RFs. The perifoveal/peripheral viewing of the attended stimulus and the use of a fine direction-discrimination task made the task attention-demanding (see Methods). The monkey performed the task reliably with an average correct rate of 86.7 ± 7.3% (mean ± std) across 23 sessions and a total of 5184 trials. The correct rates for 10°, 15°, and 20° direction offsets of the fine direction-discrimination task were 78.8 ± 9.7%, 87.5 ± 8.3%, and 93.9 ± 5.8%, respectively (see Methods).

We recorded the responses from 48 MT neurons in 23 experimental sessions while the monkey performed the task. Among the 48 neurons, 32 neurons were recorded using both the attention-away paradigm and a fixation paradigm. We found a similar faster-speed bias at low and intermediate speeds. The results obtained using the attention-away paradigm and the fixation paradigm were similar (Supplementary Fig. 1). The faster-speed bias was more evident at x4 speed separation than at x2 speed separation. Based on the neuronal responses across the population, we calculated the weight for the faster stimulus component at each of the five speed pairs using linear regression (Eqs. 6, 7) as we did for Figure 5. When attention was directed away from the RF, the response weight for the faster component decreased from a strong faster-speed bias to response averaging as the stimulus speeds increased (red curves in Fig. 8), similar to the results from the fixation paradigm (blue and black curves in Fig. 8). Together, these results suggest that the faster-speed bias at low to intermediate speeds was not due to attention being drawn to the faster-speed component.

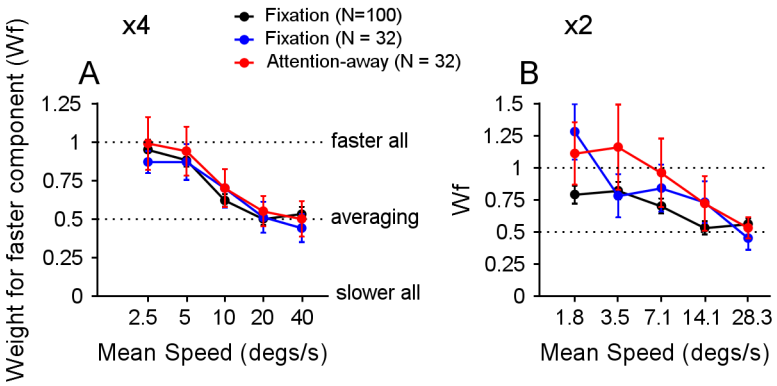


Figure 8. Comparison of response weights between attention-away and fixation paradigms. The red and blue curves indicate the response weights for the faster speed component in an attention-away paradigm and a fixation paradigm, respectively, obtained from the same population of 32 neurons. The black curves are the replot of the data in Figure 5F, obtained from 100 neurons in a fixation paradigm. **A.** X4 speed separation. **B.** X2 speed separation.

Discriminate bi-speed and single-speed stimuli based on neuronal responses in area MT

We asked whether the responses of MT neurons contain information about bi-speed and single-speed stimuli suitable to support the perceptual discrimination of these stimuli. To address this question, we first examined the responses elicited by the bi-speed and single-speed stimuli from a population of MT neurons that have different preferred speeds (PS). We next used a classifier to discriminate the bi-speed stimuli from the single, log-mean speed stimuli based on MT responses.

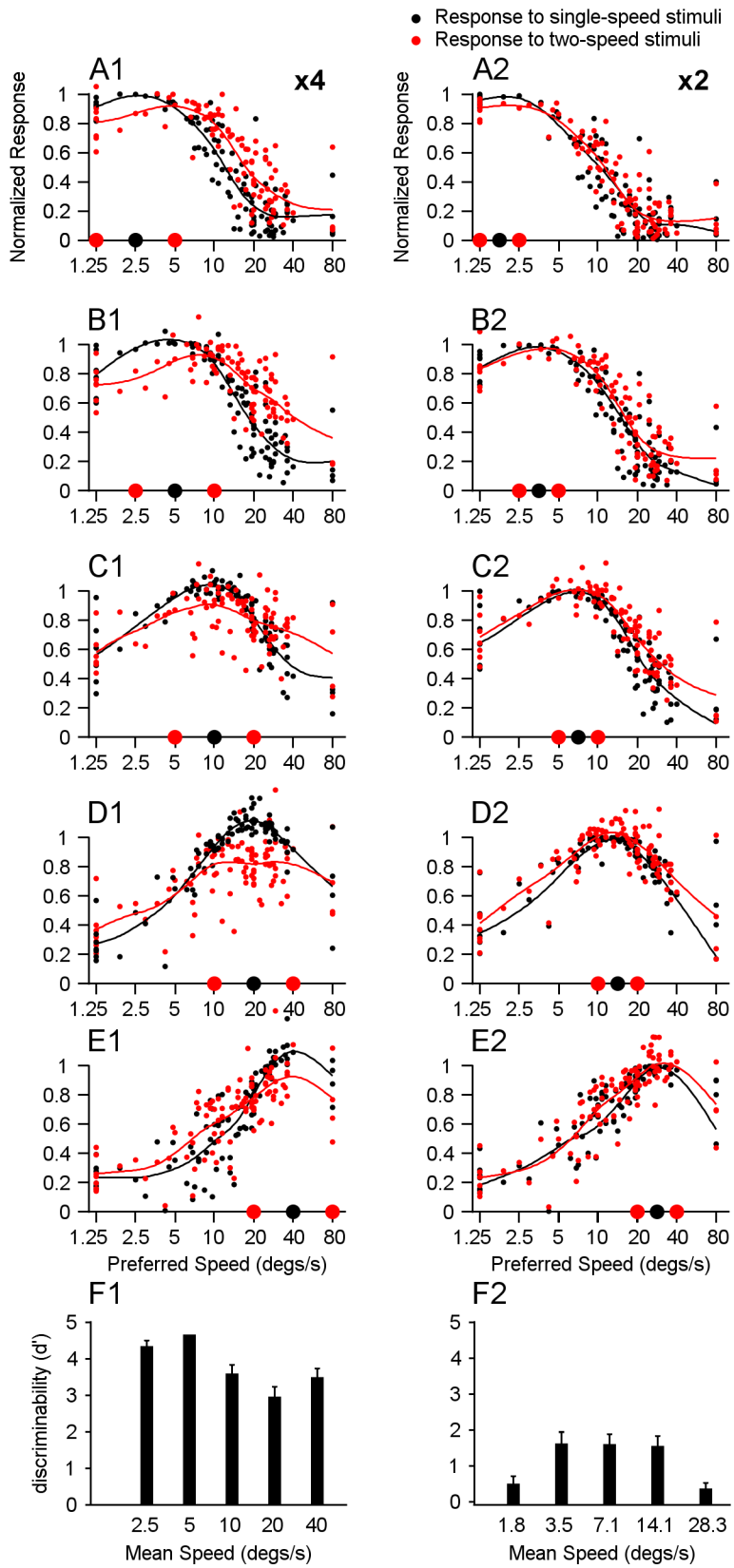
In different experimental sessions, we centered visual stimuli on neurons' RFs. Except for the spatial location, the visual stimuli were identical across experimental sessions. This allowed us to pool the trial-averaged responses recorded from different neurons to form a pseudo-population (see Methods). One can interpret the responses as from a population of neurons with spatially aligned center locations of the RFs elicited by the same visual stimulus. Figure 9 shows the pseudo-population neural response (referred to as the population response) plotted as a function of neurons' preferred speed (PS), constructed from 100 neurons that we recorded using a fixation paradigm (see Methods). To capture the population response evenly across a full range of PS, we spline-fitted the recorded response elicited by the bi-speed stimulus (the red curves) and by the single, log-mean speed (the black curves) (Fig. 9A-E). At x4 and x2 speed separations, the population responses elicited by two speeds did not show two separate peaks but rather had a main hump that shifted from low PS to high PS as the stimulus speeds increased. At x4 speed separation and across all five speed pairs, the population response elicited by two speeds was broader and flatter than that elicited by the single log-mean speed (Fig. 9A1-E1).

In our experiments, we directly measured the neuronal responses elicited by the log-mean speed of x4, but not x2 speed separation. Because we had characterized each neuron's tuning curve to single speeds, we were able to infer the responses elicited by the log-mean speed of x2 separation by interpolating the speed tuning curve using a spline fit. At x2 speed separation, the population

response elicited by two speeds was similar to that elicited by the single log-mean speed, with the two-speed population response being slightly broader (Fig. 9A2-E2).

To evaluate the discriminability between MT population responses elicited by the bi-speed stimulus and the corresponding log-mean speed, we used a linear classifier to perform a discrimination task. Trial-by-trial population responses were generated randomly according to a Poisson process and with the mean response of each neuron set to the trial-averaged neuronal response. The classifier was trained and tested using k-fold cross-validation. The classifier determined whether a population response of the 100 neurons was elicited by two speeds or a single speed (see Methods). Discriminability of the classifier was measured in *d'* as in our psychophysics study. At x2 speed separation, the classifier's performance (Fig. 9F2) had a similar shape as that of the human (Fig. 1C2, E2) and monkey (Fig. 2B2) subjects, but the classifier's performance was worse than perceptual performance. Consistent with perceptual discrimination, the classifier's performance at x4 speed separation (Fig. 9F1) was better than that at x2 speed separation (Fig. 9F2). At x4 speed separation, the discriminability was high and slightly decreased as the stimulus speed increased (Fig. 9F1), which was generally consistent with the psychophysics results (Fig. 1C1, E1). One difference was that at 20 and 80°/s, the classifier's performance did not drop to a low level as human performance (compare Fig. 9F1 with Fig. 1C1, E1), but was more comparable to that of the monkey subject (Fig. 2B1). At the highest stimulus speeds, although the difference between the population responses to the bi-speed and single-speed stimuli is sufficiently large for the classifier to pick up (Fig. 9E1), it is difficult to decode two speeds from the population response elicited by 20 and 80°/s (data not shown, but see Figs. 12-14 of Huang et al., 2023). Therefore, if the subject performed the discrimination task based on whether a stimulus contains one or two speeds (as for human subjects), the performance should be low at these high speeds. Whereas, if the subject performed the task based on other cues such as the apparent coherence of the stimulus at these fast speeds, the performance could be reasonably good. Overall, the discrimination performance of the classifier based on the population responses in MT is largely consistent with the perceptual discrimination of human and monkey subjects, with an exception when stimulus speeds were 20 and 80°/s.

Figure 9. Population neural responses elicited by the bi-speed and single-speed stimuli and the performance of a linear classifier. A population of 100 neurons was constructed by pooling across recordings in different experimental sessions. Each neuron's response was averaged across experimental trials and normalized by the maximum response of the spline-fitted speed tuning curve to single speeds. Each dot represents the response from one neuron plotted as the neuron's PS in the natural logarithm scale. The curves represent the spline-fitted population neural responses. Red: response to the bi-speed stimulus; Black: the response to the corresponding single, log-mean speed. **A1-F1**. X4 speed separation. The speeds of the bi-speed stimuli are 1.25 and 5º/s (A1), 2.5 and 10º/s (B1), 5 and 20º/s (C1), 10 and 40º/s (D1), 20 and 80º/s (E1). **A2-F2**. X2 speed separation. The speeds of the bi-speed stimuli are 1.25 and 2.5º/s (A2), 2.5 and 5º/s (B2), 5 and 10º/s (C2), 10 and 20º/s (D2), 20 and 40º/s (E2). Two red dots on the X-axis indicate two component speeds; the black dot indicates the log-mean speed. **F1, F2.** Performance of a linear classifier to discriminate the population neural responses to the bi-speed stimulus and the corresponding single log-mean speed. Error bars represent STE.


Response to single-speed stimuli
Response to two-speed stimuli
A1
x4
A2
x2
B1
B2
C1
C2
D1
D2
E1
E2
F1
F2
Normalized Response
Preferred Speed (degs/s)
discriminability (d')
Mean Speed (degs/s)
1.25 2.5 5 10 20 40 80
2.5 5 10 20 40
1.8 3.5 7.1 14.1 28.3

A model that accounts for the neuronal responses to bi-speed stimuli

We showed that the neuronal response in MT to a bi-speed stimulus can be described by a weighted sum of the neuron's responses to the individual speed components. We also showed that the weights were dependent on the stimulus speeds, rather than the preferred speed of the neuron (Fig. 5A-E). We will next show that a modified normalization model could explain these results.

The divisive normalization model (Carandini and Heeger, 2012) has been used to explain a wide array of phenomena, including neuronal responses elicited by multiple visual stimuli (e.g. Britten and Heuer 1999; Heuer and Britten 2002; Busse et al. 2009; Xiao et al. 2014; Xiao and Huang 2015; Bao and Tsao 2018). In the normalization model, while the division by the activity of a population of neurons in the denominator (the normalization pool) is well accepted, what constitutes the numerator is unclear. The signal strength such as luminance contrast or motion coherence of each stimulus component was typically used in the numerator. However, it is not always clear how to define the signal strength of a sensory stimulus. In this study, multiple speed components had the same contrast and coherence. Stimulus speed itself is not a good measure of signal strength. We have previously proposed that the weight of a stimulus component is proportional to the activity of a population of neurons elicited by the stimulus component (Xiao et al., 2014; Wiesner et al., 2020), which reflects the signal strength in the "eye" of the neuronal population. We name this neuronal population the "weighting pool". The nature and scope of the weighting pool are currently unclear. Here we assumed that, in response to multiple overlapping speeds, the weighting pool is composed of neurons with a broad range of speed preferences. We used the following equation (Eq. 9) to fit each neuron's responses to the bi-speed stimuli across 5 pairs of speeds (i.e. the speed tuning curve to the bi-speed stimuli), at either x4 or x2 speed separation:

$$R_{bi} = \frac{S_s^n}{S_s^n + \alpha S_f^n + \sigma} \bullet f(v_s) + \frac{S_f^n}{S_s^n + \alpha S_f^n + \sigma} \bullet f(v_f) + c \quad . \qquad (9)$$

R_{bi} is the model-fitted response of a neuron to a bi-speed stimulus. f is the response tuning of the neuron to single speeds. v_s and v_f are the slower and faster component speeds, respectively. S_s and S_f are the population neural responses in MT to the slower and faster component speeds, respectively, and were estimated based on the population-averaged speed-tuning curve to single speeds of our recorded MT neurons (Fig. 10A). n, σ, α, and c are model parameters and have the following constraints: $0 \leq n \leq 100$, $0 \leq \sigma \leq 500$, $0.01 \leq \alpha \leq 100$, $0 \leq c \leq 100$. α is a parameter that controls for the tuned normalization (Ni et al., 2012; Rust et al 2006; Carandini et al., 1997).

When the stimulus components move at low or intermediate speeds ≤ 20°/s, population-averaged MT response to the faster-speed component S_f is stronger than that to the slower-speed component S_s at the x4 (Fig. 10B1) and x2 speed separation (Fig. 10B2). This difference between S_f and S_s would weigh the faster stimulus component more strongly than the slower component (Eq. 9). When one or more stimulus components move at speeds greater than 20°/s, S_f is smaller than S_s (Fig. 10B1, B2) and would weigh the faster component less than the slower component. The model accounted for on average 78.3% of the response variance across 100 neurons at x4 speed separation, and 95.9% of the variance at x2 speed separation. Based on the model-fitted bi-speed responses across the population of 100 neurons, we calculated the weight for the faster stimulus component at each of the five speed pairs using linear regression, as we did for the recorded neuronal responses (Eqs. 6, 7). The weights obtained using the model-fitted responses

R_{bi} matched the weights derived from the data well (R^2 = 0.90 for x4 and 0.87 for x2 speed separation) (Fig. 10C1, C2). Although the model is reasonably successful, it has two small deficiencies: 1) At the fastest speeds of 20 and 80°/s and 20 and 40°/s, the model predicted the weight for the faster component to be slightly less than 0.5, whereas the data showed response averaging (Fig. 10C1, C2); 2) At the slowest speeds of 1.25 and 5°/s, the faster-speed bias predicted by the model was not as strong as the data (Fig. 10C1).

In Equation 9, the parameter α only controls the relative contribution of S_s and S_f to the response of the normalization pool. Although the mechanism underlying the tuned normalization is unknown, we assumed that a similar mechanism may also apply to the response of the weighting pool, and may reflect how faster and slower components are mixed in the feedforward input. In Equation 10, α appears in both the numerator and denominator in determining the weight for the faster speed component.

$$R_{bi} = \frac{S_s^n}{S_s^n + \alpha S_f^n + \sigma} \bullet f(v_s) + \frac{\alpha S_f^n}{S_s^n + \alpha S_f^n + \sigma} \bullet f(v_f) + c \quad . \qquad (10)$$

The model based on Equation 10 improved the fit of the speed tuning to the bi-speed stimuli and accounted for on average 84.7% and 97.7% of the response variance at x4 and x2 separation, respectively, significantly better than Equation 9 (one-tailed paired t-test, $p < 0.002$). The weights obtained using Equation 10 also matched the weight of the faster component derived from data better (R^2 = 0.925 for x4 and 0.93 for x2 separation) (Fig. 10D1, D2). For both x4 and x2 separations, the median α of Equation 10 is 1.2, which is significantly different from 1 (Wilcoxon signed-rank test, $p < 0.0002$). This result indicates that, in addition to the stimulus-dependent weighting prescribed by S_f and S_s, the parameter α provides a stronger weighting for the faster component across speeds. Together, the overall response weight for the faster component is greater than the slower component at the low and intermediate speed range, and the two weights are similar at fast stimulus speeds, consistent with the neural data (Fig. 10D1, D2).

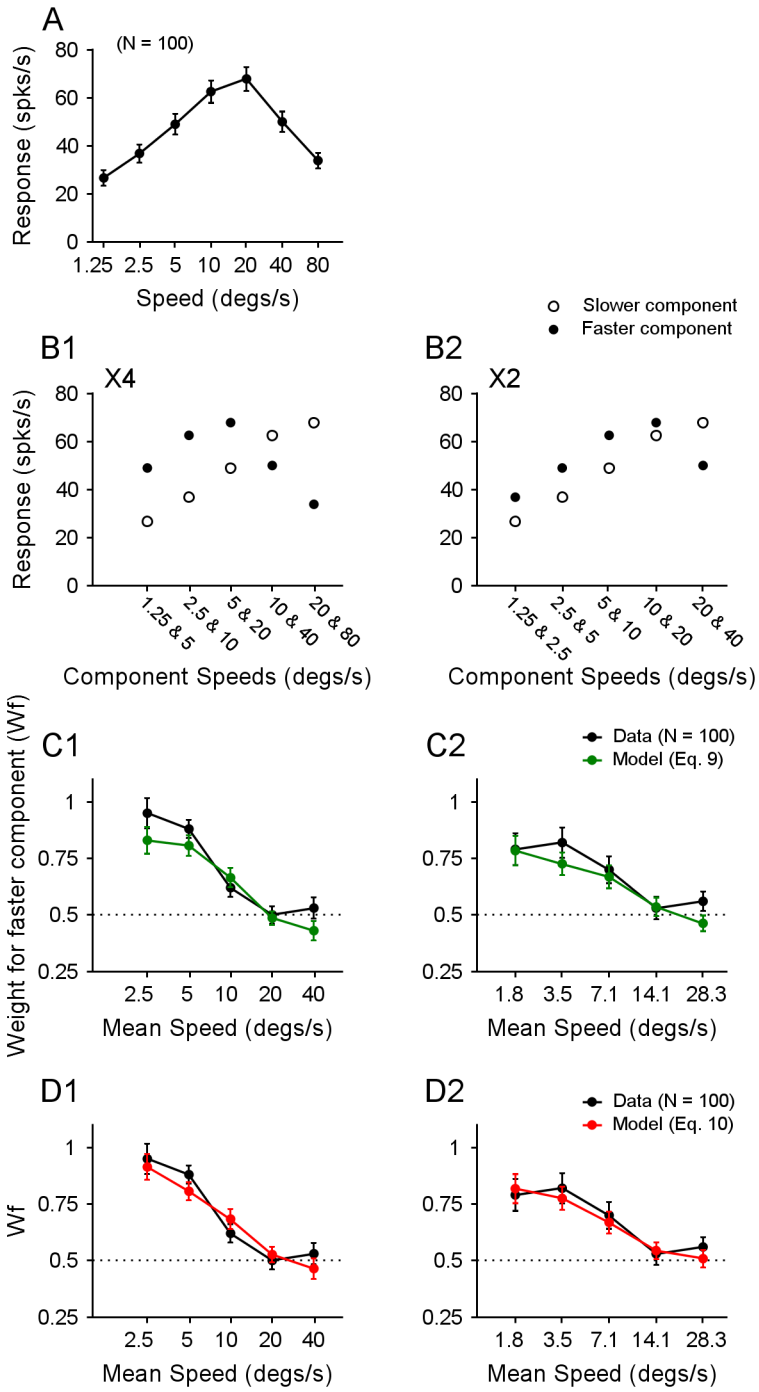
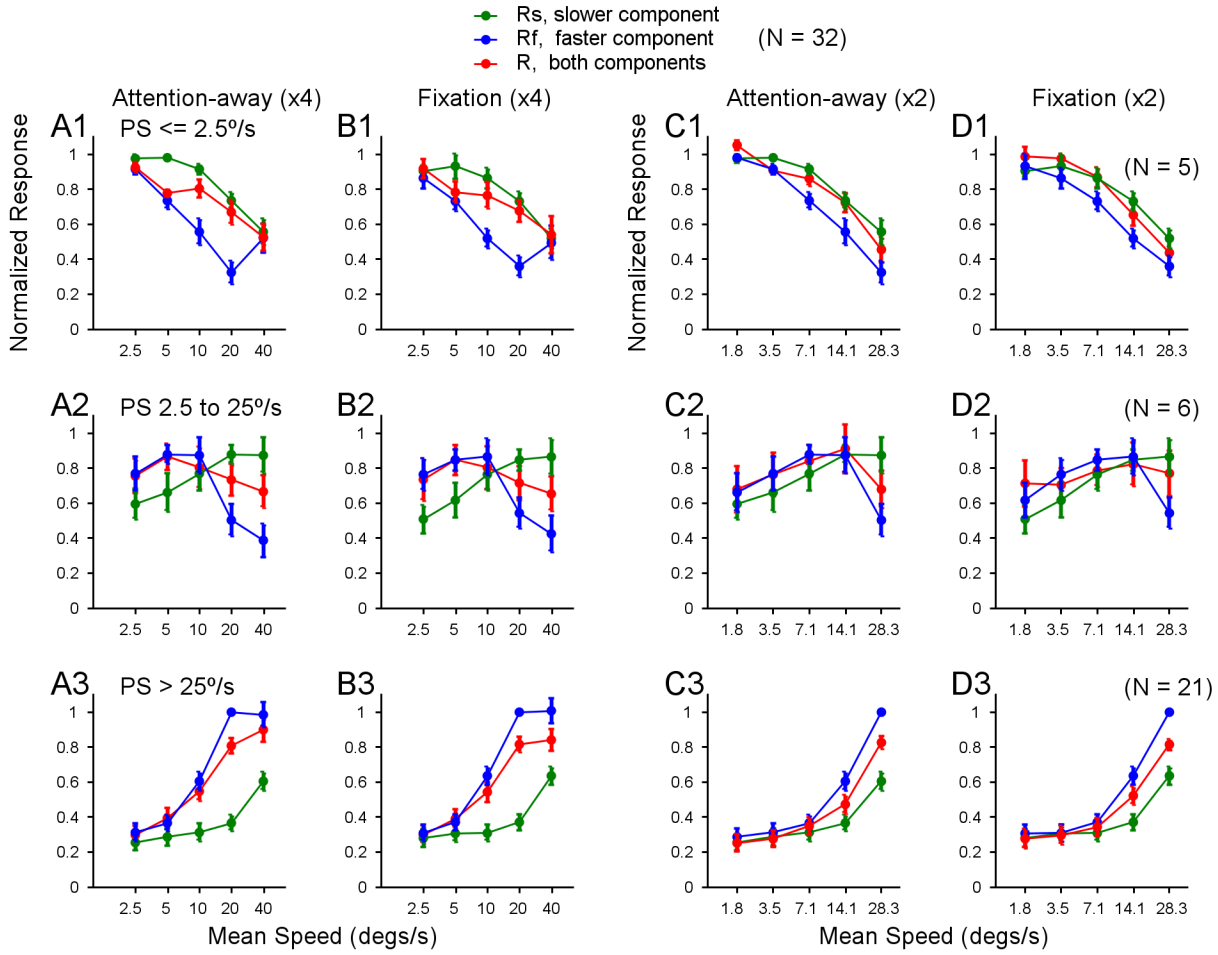


Figure 10. Model fit of MT responses to bi-speed stimuli. A. Speed tuning curve to single-speed stimulus averaged across 100 recorded MT neurons in our data set. **B**. Population-averaged responses to slower (open circle) and faster (solid circle) speed components. The convention for the speed components is the same as in Figure 1B1, B2. **C, D.** The response weights for the faster component calculated based on the data (black) and the models of Equation 9 (green in **C**) and Equation 10 (red in **D**). **B1-D1.** X4 speed separation. **B2-D2.** X2 speed separation.

Neural encoding of multiple motion speeds in visual cortical area MT

Supplementary Material



Supplementary Figure 1. Population-averaged speed tuning curves to bi-speed stimuli and constituent single-speed components recorded in an attention-away and a fixation paradigm. Speed tuning curves from one monkey (RG) averaged across **A1-D1.** 5 neurons that had PS ≤ 2.5º/s, **A2-D2.** 6 neurons that had PS between 2.5 and 25º/s, **A3-D3.** 21 neurons that had PS > 25º/s. Error bars represent ±STE. **A1-A3** and **B1-B3.** X4 speed separation; **C1-C3** and **D1-D3.** X2 speed separation. **A1-A3** and **C1-C3.** Attention directed away from the RF; **B1-B3** and **D1-D3.** Fixation paradigm.

DISCUSSION

Perception of multiple motion speeds and possible neural basis

Our human psychophysical study employed a novel 3AFC task. The task combined an identification task (to report whether a stimulus had one or two speeds) with a discrimination task (to compare a two-speed stimulus with a single-speed stimulus) (Fig. 1A, E1, E2). This approach allowed us to characterize discriminability based on perceptual segmentation, rather than other perceptual appearances of the stimuli. We made two findings. First and intuitively, the performance of speed segmentation was better when the separation between two stimulus speeds was larger. Second, at a fixed speed separation, speed segmentation became harder at fast speeds. Our results are consistent with previous studies. Masson et al. (1999) showed that the speed segmentation threshold increased sharply when the mean stimulus speed increased from 8º/s to 16º/s. By varying the width of a speed notch, Rocchi et al. (2018) showed that transparent motion perception was stronger when the notch width was wider, and that transparent motion was well perceived at slow speeds (mean speed = 4.6º/s), but not at faster speeds (mean speed = 20.6º/s) at a range of notch width from 1 to 6º/s. Our study showed that the segmentation performance dropped sharply at speeds of 20 and 80º/s (x4), and 20 and 40º/s (x2), faster than those shown in the previous studies. This discrepancy is likely due to the larger speed separations used in our study and the difference in stimuli. The visual stimuli used in our study had either one or two speeds, whereas those used by Rocchi et al. (2018) were sampled from a distribution of motion speeds and had multiple elements.

Our MT data from macaque monkeys provide explanations for the perceptual performance of speed segmentation. By comparing the constructed population MT responses elicited by two speeds and a single, log-mean speed, we found that the response difference was larger at x4 speed separation than at x2 separation (Fig. 9). This provides a neural correlate with better perceptual speed segmentation at larger speed separation. The explanation for the poor speed segmentation at fast stimulus speeds is more complicated. One factor may be related to the rule of neural encoding. At slow stimulus speeds, we found that neuronal response in MT showed a faster-speed bias. This would make detecting and extracting the faster speed component easier and therefore benefit speed segmentation. At fast stimulus speeds, we found that the encoding of multiple speeds changed to response averaging. As we have shown previously using visual stimuli moving transparently in different directions, a classifier's performance of discriminating a bi-directional stimulus from a single-direction stimulus is worse when the encoding rule is response-averaging than biased toward one of the stimulus components (Fig. 12, Xiao and Huang, 2015). The same may apply to multiple motion speeds. Indeed, when the stimulus speeds were 20 and 40º/s (x2 separation), the constructed population responses elicited by the bi-speed stimulus and the log-mean speed were very similar (Fig. 9E2), which explained the poor performance of the classifier (Fig. 9F2) and the monkey subject (Fig. 2B2) at these speeds. However, when the stimulus speeds were 20 and 80º/s (x4 separation), the population responses elicited by the bi-speed stimulus and the log-mean speed were noticeably different, likely due to the large speed separation (Fig. 9E1). As a result, the classifier performed well at these fast speeds (Fig. 9F1), which was inconsistent with the performance of the monkey and human subjects. These observations suggest that, even when the population neural response elicited by a bi-speed stimulus is different from that elicited by a single-speed stimulus, there is no guarantee that information of two motion speeds is carried in the population neural response elicited by the bi-speed stimulus. The difference in population neural

responses may contribute to perceptual differences in quality other than motion speeds. We therefore conducted a decoding study to evaluate what speed(s) can be extracted from MT population neural responses. We will report the main findings of the decoding study in a different paper as the current paper focuses on neural encoding. Briefly, we found that at x4 speed separation, it was possible to extract the speeds of motion components from the MT population neural response when the component speeds were slower than 20 and 80º/s. At 20 and 80º/s, the decoder was uncertain about how many speeds were in the visual stimuli and therefore had difficulty segmenting the visual stimuli at these fast speeds. The discrimination performance based on the decoded speeds was comparable to the perceptual performance of the monkey subject (Figs. 12-14 of Huang et al., 2023).

Neural mechanisms underlying the encoding of multiple speeds of overlapping stimuli

We made a novel finding that the responses of MT neurons to overlapping stimuli were biased toward the faster speed component when the stimulus speeds were slow. The faster-speed bias was not due to an attentional modulation because we found similar results when attention was directed away from the RF. The faster-speed bias cannot be explained by the apparent contrast of the stimulus component either – the random dots of the faster speed component had shorter dwell time on the video display and appeared to be dimmer than the slower component. We showed that the neural encoding results can be explained by a modified normalization model.

Previous studies that characterize the neural representation of multiple stimuli have used stimulus strength to weigh the component responses in the divisive normalization model (Busse et al., 2009; Ni et al, 2012; Xiao et al., 2014; Heuer and Britten, 2002). In comparison to the standard normalization model, our model assumes that the response of a population of neurons (i.e. the weighting pool) defines the numerator of the normalization equation. The weighting pool may or may not be the same as the normalization pool that defines the response in the denominator. We have shown that weighting by the population neural response elicited by the stimulus component provides a parsimonious explanation of the response to multiple speeds even if the stimulus strength is not well defined. In a study investigating how neurons in the inferotemporal cortex represent multiple objects, Bao and Tsao (2018) suggest that the responses of neurons in category-selective regions to multiple objects are weighted by the responses from neighboring neurons that have the same category selectivity. Our study provides new insight into the extent of the weighting pool – in response to multiple speeds of overlapping stimuli, the weighting pool may include neurons with a broad range of preferred speeds. In this way, the summed (or averaged) population response depends only on the stimulus speed and is invariant to the individual neuron's speed preference. In our data, MT population-averaged speed tuning curve peaks around 20º/s (Fig. 10A), which is consistent with previous studies (Maunsell and Van Essen 1983; Lisberger and Movshon, 1999; Nover et al., 2005; Huang and Lisberger, 2009). At speeds less than 20º/s, the population speed tuning has a positive slope, and a faster component would elicit a stronger population response than a slower component. This insight explains the faster-speed bias at low stimulus speeds and why a fixed weight for the faster component fits the responses of individual neurons elicited by a pair of stimulus speeds remarkably well, regardless of the speed preferences of the individual neurons (Fig. 5A-E). Neurons with similar speed preferences are spatially clustered in MT (Liu and Newsome, 2003). Should the weighting pool be composed of locally clustered neurons with similar preferred speeds, the response weight of a neuron would be higher for a speed component that elicits a stronger response. This would be contrary to our finding that neurons

preferring very low speeds also showed a faster-speed bias, rather than a bias toward the slower component that elicited a stronger response (Fig. 4B).

Although in our model we used the responses of a population of MT neurons to estimate the response of the weighting pool, it is also possible that the weighting pool may be composed of neurons that feed signals into MT and have similar population-averaged speed tuning as MT neurons. A recent study from our lab using multiple stimuli competing in more than one feature domain suggests that it is important to consider hierarchical processing in representing multiple stimuli, in particular how multiple stimuli are represented in the feedforward input to a visual area (Wiesner et al., 2020). Our result that the initial MT response to the bi-speed stimuli was nearly identical to the response to the faster component alone (Fig. 6A-C) suggests that the faster-speed bias may be already present in the feedforward input. The suppressive effect due to the presence of the slower component did not appear until 20-30 ms after the MT response onset (arrows in Fig. 6A-C), suggesting possible involvement of neural circuits within MT for additional processing and divisive normalization. MT neurons receive feedforward motion-selective input mainly from V1, and also from V2 and V3 (Ungerleider and Desimone, 1986; Movshson and Newsome, 1996; Anderson et al., 1998; Anderson and Martin 2002; Rockland 2002). Speed-selective complex cells in V1 have preferred speeds in a range similar to that of MT neurons, but the mean preferred speed is slower than MT (Mikami et al. 1986; Orban et al., 1986; Priebe et al., 2006). Normalization in V1 may contribute to the faster-speed bias at low speeds. The roles of neural processing in early visual areas on the faster-speed bias remain to be determined in future studies.

Efficient coding and functional implication of faster-speed bias

We have shown that the faster-speed bias in MT response is a robust phenomenon regardless of whether the stimulus components move in the same or different directions. An efficient way to represent sensory information is to devote limited resources to better represent signals that occur more frequently in the natural environment (Attneave 1954; Barlow 1961; Simoncelli and Olshausen 2001). Previous studies have suggested that slow speeds are more likely to occur than fast speeds in natural scenes (Weiss et al., 2002; Stocker and Simoncelli, 2006; Zhang and Stocker, 2022). If neurons in the primate visual cortex are optimized to efficiently represent speeds that are more likely to occur in natural scenes, one may expect to find neurons showing a slower-speed bias rather than a faster-speed bias. However, besides maximizing information about the environment, neural representation in the sensory cortices may be optimized for other goals such as maximizing the performance of certain behavioral tasks (Simoncelli and Olshausen 2001; Manning et al., 2023). In the natural environment, it is common to encounter multiple motions occurring concurrently in a spatial region, due to object motions in the world coordinate and retinal image slip caused by the self-motion of the observer. If a figural object (e.g. a lion) tends to move faster than its background in natural scenes, a neural representation of multiple motions with a faster-speed bias would help to identify the figure, and therefore benefit the performance of an essential behavioral task – figure/ground segregation. To test this hypothesis, future study is needed to characterize natural scene statistics of speeds for figural objects and their background.

MATERIALS AND METHODS

We conducted psychophysical experiments using human subjects, and psychophysical and neurophysiological experiments using macaque monkeys.

Human psychophysics

Subjects

Four adult human subjects (*CN, CO, IN, NP*), two men and two women, with normal or corrected-to-normal visual acuity participated in the psychophysics experiments. Subject *CN* was naive about the purposes of the experiments. Subjects CO and IN had a general idea about this study but did not know the specific design of the experiments. Informed consent was obtained from the subjects. All aspects of the study were in accordance with the principles of the Declaration of Helsinki and were approved by the Institutional Review Board at the University of Wisconsin-Madison.

Apparatus

Visual stimuli were generated by a Linux workstation using an OpenGL application and displayed on a 19-inch CRT monitor. The monitor had a resolution of 1,024 x 768 pixels and a refresh rate of 100 Hz. The output of the video monitor was measured with a photometer (LS-110, Minolta) and was gamma-corrected. Stimulus presentation was controlled by a real-time data acquisition and stimulus control program "Maestro" (https://sites.google.com/a/srscicomp.com/maestro/) as in the animal behavior and neurophysiology experiments. Subjects viewed the visual stimuli in a dark room with dim background illumination. The viewing distance was 58 cm. A chin rest and forehead support were used to restrict the head movements of the observers. During experimental trials, human subjects maintained fixation on a small spot within a 2 x 2° window. Eye positions were monitored using a video-based eye tracker (EyeLink, SR Research) at a rate of 1kHz.

Visual stimuli

Visual stimuli were two spatially overlapping random-dot patches presented within a square aperture 10° wide. Each square stimulus was centered 11° to the right of the fixation spot, therefore covering 6° to 16° eccentricity. This range roughly matched the RF eccentricity of the recorded MT neurons in our neurophysiological experiments. The random dots were achromatic. Each random dot was 3 pixels and had a luminance of 15.0 cd/m^2. The background luminance was 0.03 cd/m^2. The dot density of each random dot patch was 2 dots/degree^2. The two random-dot patches translated horizontally in the same direction. To reduce adaptation, the motion direction was either leftward or rightward in half of the trials, and stimulus trials were randomly interleaved. In one set of trials, two overlapping random-dot patches had a "large speed separation" and the speed of the faster component was always four times (x4) that of the slower component. In another set of trials, visual stimuli had a "small speed separation" and the speed of the faster component was always twice (x2) that of the slower component (see Fig. 1B1, B2). For each bi-speed stimulus, there was a corresponding single-speed stimulus composed of two overlapping random-dot patches moving in the same direction at the same speed. The single speed was the natural logarithmic (log)

mean speed of the bi-speed stimulus: $Spd_{mean} = e^{[\ln(Spd_1)+\ln(Spd_2)]/2}$, in which Spd_1 and Spd_2 were the two component speeds. The motion coherence of each random-dot patch was always 100%.

Procedure

In a standard two-alternative-forced-choice (2AFC) task, subjects discriminated a bi-speed stimulus from the corresponding single log-mean speed stimulus. The bi-speed and single-speed stimuli were presented in two consecutive time intervals with a 500 ms gap in between, in random, balanced order. In each time interval, the visual stimulus appeared, remained stationary for 250 ms, and then moved for 500 ms. At the end of each trial, subjects reported which time interval contained a bi-speed stimulus by pressing one of two buttons (left or right) within a 1500-ms window. After the button press, the inter-trial interval was 1300 ms. Each block of trials contained 40 trials, i.e. 5 speed pairs × 2 speed separations × 2 temporal orders (the bi-speed stimulus appeared in the first or second time-interval) × 2 motion directions (visual stimuli moved either to the left or right). Each experimental session typically contained 5 blocks, i.e. 200 trials.

Subjects also performed a 3AFC task. As in the 2AFC task, subjects discriminated a bi-speed stimuli from the corresponding single log-mean speed stimulus but had the option to make a third choice by pressing the middle button on trials when they thought neither stimulus interval appeared to contain two speeds ("no two-speeds" choice). When subjects thought one of the two stimulus intervals contained two speeds, subjects then pressed either the left or the right button to indicate which interval had two speeds.

Data analysis

The hit rate was calculated as the percentage of trials in which a subject correctly picked the bi-speed stimulus as having two speeds. The false alarm rate was calculated as the percentage of trials that a subject incorrectly picked the singe-speed stimulus as having two speeds. As a measure of discriminability between the bi-speed and the corresponding single-speed stimuli, we calculated the discriminability index d' = *norminv*(hit rate) – *norminv(*false alarm rate). *norminv* is a MATLAB function that calculates the inverse of the normal cumulative distribution function, with the mean and standard deviation set to 0 and 1, respectively. When the hit or false alarm rate was occasionally close to 1, to avoid infinite d' values, d' was calculated using a modified formula: d' = norminv{[(100 x hit rate)+1]/102} - norminv{[(100 x false alarm rate) +1]/102}. In analyzing the results of the 3AFC task, we incorporated the NTC trials into the *d'* calculation by evenly splitting the NTC trials into "hit" trials and "false alarm" trials. In this way, the NTC trials were still accounted for by the hit rate and false alarm rate, in the sense that they did not contribute to the discrimination. We also examined the percentage of trials in which subjects made the NTC choice at different stimulus speeds.

Neurophysiological and psychophysical experiments

Subjects

Five male adult rhesus monkeys (*Macaca mulatta*) were used in the experiments. Four monkeys were used in the neurophysiological experiments, and one was used in the psychophysical experiment. Experimental protocols were approved by the local Institutional Animal Care and Use Committee and were in strict compliance with U.S. Department of Agriculture regulations and the National Institutes of Health *Guide for the Care and Use of Laboratory Animals*.

Apparatus and electrophysiological recording

Procedures for surgical preparation and electrophysiological recording were routine and similar to those described previously (Huang and Lisberger 2009; Xiao et al., 2014). For subjects IM and MO, horizontal and vertical eye positions were monitored using the search coil method at a sampling rate of 1kHz on each channel. For subjects RG, GE, and BJ, eye positions were monitored using a video-based eye tracker (EyeLink, SR Research) at a rate of 1kHz. For electrophysiological recordings, we lowered single-contact tungsten microelectrodes (Thomas Recording or FHC) either using the MiniMatrix microdrive (Thomas Recording) or the NAN drive (NAN Instruments) into the posterior bank of the superior temporal sulcus. The impedances of the electrodes were 1∼3 MΩ. We identified area MT by its characteristically large proportion of directionally selective neurons, small classical RFs relative to those in the neighboring medial superior temporal area, and location on the posterior bank of the superior temporal sulcus. Electrical signals were filtered, amplified, and digitized conventionally. Single units were identified with a real-time template-matching system (Plexon). Spikes were carefully sorted using Plexon offline sorter.

Stimulus presentation and the behavioral paradigm were controlled by a real-time data acquisition program Maestro as described in the human psychophysics experiment. For neurophysiological recordings from IM and MO, visual stimuli were presented on a 20-inch CRT monitor at a viewing distance of 38 cm. Monitor resolution was 1,280 × 1,024 pixels and the refresh rate was 85 Hz. For RG, GE, and BJ, visual stimuli were presented on a 25-inch CRT monitor at a viewing distance of 63 cm. Monitor resolution was 1,024 × 768 pixels and the refresh rate was 100 Hz. Visual stimuli were generated by a Linux workstation using an OpenGL application that communicated with the main experimental-control computer over a dedicated Ethernet link. The output of the video monitor was gamma-corrected.

Visual stimuli and experimental procedure of the main experiment

All visual stimuli were presented in individual trials while monkeys maintained fixation. Monkeys were required to maintain fixation within a 1.5 × 1.5° window centered around a fixation spot during each trial to receive juice rewards, although actual fixation was typically more accurate. In a trial, visual stimuli were illuminated after the animal had acquired fixation for 200 ms. To assist the isolation of directional-selective neurons in area MT, we used circular translation of a large random-dot patch (30 × 30°) as a search stimulus (Schoppmann and Hoffmann, 1976). After an MT neuron was isolated, we characterized the direction tuning by randomly interleaved trials of 30 × 30° random-dot patches moving at 10°/s in eight different directions from 0 to 315° at 45° steps. Next, we mapped the RF by recording responses to a series of 5 × 5° patches of random dots that moved in the preferred direction of the neuron at 10°/s. The location of the patch

was varied randomly to tile the screen in 5° steps without overlap and to cover an area of either 40 × 30° or 35 × 25°. The raw map of the RF was interpolated using the Matlab function *interp2* 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, testing stimuli were centered on the RF.

Monkeys IM and MO were tested with the main visual stimuli used in our experiments, which were two spatially overlapping random-dot patches presented within a square aperture 10° wide. The random dots were achromatic. The dot density of each random-dot patch was 2 dots/deg^2. Each random dot was 3 pixels at a side and had a luminance of 15.0 cd/m^2. The background luminance was < 0.2 cd/m^2. In each trial, the random dots moved within the aperture. The two random-dot patches translated at two different speeds at 100% motion coherence and in the same direction (the preferred direction of the recorded neuron). The ratio between the two component speeds was fixed either at 4 (i.e. the large speed separation) or 2 (i.e. the small speed separation) (see Methods for human psychophysics above). At x4 speed separation, the five speed pairs used were 1.25 and 5°/s, 2.5 and 10°/s, 5 and 20°/s, 10 and 40°/s, and 20 and 80°/s (Fig. 1B1). At x2 speed separation, the speed pairs used were 1.25 and 2.5°/s, 2.5 and 5°/s, 5 and 10°/s, 10 and 20°/s, and 20 and 40°/s (Fig. 1B2). Experimental trials of bi-speed stimuli that had x4 or x2 speed separations were randomly interleaved. Also randomly interleaved were trials that showed only a single random-dot patch moving at a speed of 1.25, 2.5, 5, 10, 20, 40, or 80°/s, which were the individual stimulus components of the bi-speed stimuli.

Monkeys RG and GE were tested with a variation of the main visual stimuli, in which two overlapping random-dot stimulus components moved at two fixed speeds of 2.5 and 10°/s, respectively, and in two different directions separated by 90°. The diameter of the stimulus aperture was 3°. The faster component moved at the clockwise side of the two component directions (illustrated in Figure 7). We varied the vector average direction of the two stimulus components across 360° in a step of 15° to characterize the direction-tuning curves of MT neurons. We also measured the direction-tuning curves to a single stimulus moving at the individual component speeds.

Behavioral paradigm and visual stimuli of attention control

Monkey RG was also tested in a control experiment in which the attention of the animal was directed away from the RFs of MT neurons. The attended stimulus was a random-dot patch moving in a single direction at 100% motion coherence within a stationary circular aperture that had a diameter of 5º. The stimulus patch was centered 10º to the left of the fixation spot, in the visual hemifield contralateral to the hemifield of the recorded MT neurons' RFs. The monkey performed a fine direction-discrimination task to report whether the motion direction of the attended stimulus moved at the clockwise or counter-clockwise side of the vertical direction. While the animal fixated on a point at the center of the monitor, both the attended stimulus and the RF stimulus were turned on and remained stationary for 250 ms before they moved for 500 ms. The attended stimulus translated at a speed of 10°/s and in a direction either clockwise or counter-clockwise from an invisible vertical (upward) direction by an offset of 10°, 15°, or 20°. The RF stimuli were the same as our main visual stimuli, with either a single-speed or bi-speed stimulus moving in the same direction. All trials were randomly interleaved. After the motion period, all the visual stimuli were turned off, and two reporting targets appeared 10° eccentric on the left and

right sides of the fixation point. To receive a juice reward, the animal was required to make a saccadic eye movement within 400 ms after the fixation spot was turned off, either to the left or right target when the motion direction of the attended stimulus was counter-clockwise or clockwise to the vertical direction, respectively.

Monkey psychophysics

Monkey BJ was trained to perform a 2AFC discrimination task. The visual stimuli were the same as our main visual stimuli in the neurophysiological experiments except that the stimulus moving at a single speed was also composed of two overlapping random-dot patches moving in the same direction at the same speed, the same as in the human psychophysics experiments. In this way, the single-speed stimulus and the bi-speed stimuli had the same dot density. Visual stimuli were random-dot patches moving within a square aperture of 10°x10°, centered 10° to the right of the fixation spot. The motion direction of the visual stimuli was always rightward. Experimental trials of bi-speed stimuli that had x4 or x2 speed separations, as well as the single-speed stimulus that moved at the log mean speed of the bi-speed stimuli were randomly interleaved. Visual stimuli were turned on and remained stationary for 250 ms before they moved for 500 ms. Following the stimulus offset, two reporting targets (dots) were presented 5.7° away from the fixation spot, at upper right (4°, 4°) and lower left (-4°, -4°) positions relative to the fixation spot. To receive a juice reward, the animal was required to make a saccadic eye movement to one of the two targets within 300 ms after the fixation spot was turned off. In a majority of the experiment trials, the animal received juice rewards if selecting the upper-right target when visual stimuli moved at two different speeds and selecting the lower-left target when visual stimuli moved at a single speed. Guided by our human psychophysics results, we made an exception to always reward the animal when the bi-speed stimuli moved at 20 and 80°/s or at 20 and 40°/s, regardless of which target was selected to avoid biasing the monkey's choice by veridically rewarding the animal. This was because, at these fast speeds, human subjects could not segment the bi-speed stimuli. During training, the animal was never presented with the bi-speed stimuli of 20 and 80°/s, and 20 and 40°/s. During testing, the trials of 20 and 80°/s, and 20 and 40°/s were randomly interleaved with bi-speed and single-speed trials that were rewarded veridically to anchor the task rule. Among all testing trials, only 10% of the trials were rewarded with a 100% rate. We collected 50 trials of data for x4 speed separation across 5 experimental sessions, and 90 trials for x2 speed separation across 9 sessions during the testing phase. The hit rate, false alarm rate, and the *d'* were calculated in the same way as in the human psychophysics experiments.

Model fit of the tuning curves to bi-speed stimuli

We fitted the response tuning curves to the bi-speed stimuli using a few variants of a divisive normalization model (Fig. 10). We also used a weighted summation model to fit the direction tuning curves to overlapping stimuli moving in different directions and at different speeds (Fig. 7). These model fits were obtained using the constrained minimization tool "fmincon" (MATLAB) to minimize the sum of squared error. To evaluate the goodness of fit of models for the response tuning curves, we calculated the percentage of variance (PV) accounted for by the model as follows: PV = 100 x (1 - $\frac{SSE}{SST}$), where SSE is the sum of squared errors between the model fit and the neuronal data, and SST is the sum of squared differences between the data and the mean of the data (Morgan et al., 2008).

Construction of population neural response

For each recorded MT neuron, we plotted the trial-averaged speed tuning curve in response to the single speed and spline-fitted the tuning curve using the Matlab function *csaps* with the smoothing parameter p set to 0.93. We found $p = 0.93$ best captured the trend of the speed tuning, without obvious overfitting. We then found the preferred speed (PS) of the neuron, which is the speed when the maximum firing rate was reached in the spline-fitted tuning curve. The neuron's responses to all single-speed and bi-speed stimuli were normalized by the maximum firing rate at the PS. To construct the population neural response to a given stimulus, we took the normalized firing rate of each neuron elicited by that stimulus and plotted it against the PS of the neuron. Because the PSs of the neurons in our data sample did not cover the full speed range evenly, we spline-fitted (with a smoothing parameter of 0.93) the population neural response to capture the population neural response evenly across the full range of PS.

Discrimination of population neural responses using a classifier

We trained a linear classifier to discriminate constructed population neural response to a bi-speed stimuli and the corresponding single-speed stimulus moving at the log mean speed. Constructed trial-by-trial population responses were generated randomly according to a Poisson process with the mean set to the recorded neuronal response averaged across experimental trials. For each speed combination, we generated 200 trials of responses to the bi-speed stimuli and the corresponding single-speed stimulus, respectively. Constructed population responses were partitioned into training and testing sets using k-fold cross-validation ($k = 40$). The 200 generated trials were randomly divided into 40 folds. The classifier was trained on 39 data folds and tested on the remaining fold, and the process was repeated 40 times to ensure that each fold was used for testing exactly once. The Matlab *fitclinear* function was used to fit a linear classifier to the training data. The logistic learner and lasso regularization techniques were specified during the model training. The Stochastic Gradient Descent solver was used to optimize the objective function during the training of the classifier. The performance of the classifier was evaluated by d', calculated using the hit rate and false alarm rate as described in human psychophysics.

REFERENCES

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

Anderson, J. C., Binzegger, T., Martin, K. A., & Rockland, K. S. (1998). The connection from cortical area V1 to V5: a light and electron microscopic study. *J Neurosci*, *18*(24), 10525-10540.

Anderson, J. C., & Martin, K. A. (2002). Connection from cortical area V2 to MT in macaque monkey. *J Comp Neurol*, *443*(1), 56-70.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol Rev*, *61*(3), 183-193.

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

Barlow H. B. (1961). Possible principles underlying the transformations of sensory messages. Chapter 13. In: W. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). M.I.T. Press.

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

Born, R. T., Groh, J. M., Zhao, R., & Lukasewycz, S. J. (2000). Segregation of object and background motion in visual area MT: effects of microstimulation on eye movements. *Neuron, 26*(3), 725-734.

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

Braddick, O. (1997). Local and global representations of velocity: transparency, opponency, and global direction perception. *Perception*, *26*(8), 995-1010.

Braddick, O. J., Wishart, K. A., & Curran, W. (2002). Directional performance in motion transparency. *Vision Res*, *42*(10), 1237-1248.

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

Britten, K. H. (2003). The middle temporal area: motion processing and the link to perception. In J. W. LM Chalupa (Ed.), *The visual neurosciences* (pp. 1203-1216). The MIT Press.

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

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51-62.

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

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

Huang, X., Albright, T. D., & Stoner, G. R. (2007). Adaptive surround modulation in cortical area MT. *Neuron*, *53*(5), 761-770.

Huang, X., Albright, T. D., & Stoner, G. R. (2008). Stimulus dependency and mechanisms of surround modulation in cortical area MT. *J Neurosci*, *28*(51), 13889-13906.

Huang, X., Ghimire, B., Chakrala, A. S., & Wiesner, S. (2023). Neural coding of multiple motion speeds in visual cortical area MT. *BioRxiv.* doi: https://doi.org/10.1101/2023.04.08.532456. https://www.biorxiv.org/content/10.1101/2023.04.08.532456v1.full

Huang W., Huang X., & Zhang K. (2017). Information-theoretic interpretation of tuning curves for multiple motion directions. *51st Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, USA, pp. 1-4, doi: 10.1109/CISS.2017.7926142.

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

Krekelberg, B., & van Wezel, R. J. (2013). Neural mechanisms of speed perception: transparent motion. *J Neurophysiol*, *110*(9), 2007-2018.

Krekelberg, B., van Wezel, R. J., & Albright, T. D. (2006a). Adaptation in macaque MT reduces perceived speed and improves speed discrimination. *J Neurophysiol*, *95*(1), 255-270.

Lisberger, S. G., & Movshon, J. A. (1999). Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. *J Neurosci*, *19*(6), 2224-2246.

Liu, J., & Newsome, W. T. (2003). Functional organization of speed tuned neurons in visual area MT. *J Neurophysiol*, *89*(1), 246-256.

Manning, T.S., Alexander, E., Cumming, B. G., DeAngelis, G. C., Huang, X., & Cooper, E. A. (2023). Transformations of sensory information in the brain reflect a changing definition of optimality. *BioRxiv.* doi: https://doi.org/10.1101/2023.03.24.534044. https://www.biorxiv.org/content/10.1101/2023.03.24.534044v1.full

Masson, G. S., Mestre, D. R., & Stone, L. S. (1999). Speed tuning of motion segmentation and discrimination. *Vision Res*, *39*(26), 4297-4308.

Maunsell, J. H., & Van Essen, D. C. (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*(5), 1127-1147.

McDonald, J. S., Clifford, C. W., Solomon, S. S., Chen, S. C., & Solomon, S. G. (2014). Integration and segregation of multiple motion signals by neurons in area MT of primate. *J Neurophysiol*, *111*(2), 369-378.

Mestre, D. R., Masson, G. S., & Stone, L. S. (2001). Spatial scale of motion segmentation from speed cues. *Vision Res*, *41*(21), 2697-2713.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *J Neurophysiol*, *55*(6), 1308-1327.

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

Ni, A. M., Ray, S., & Maunsell, J. H. (2012). Tuned normalization explains the size of attention modulations. *Neuron*, *73*(4), 803-813.

Nover, H., Anderson, C. H., & DeAngelis, G. C. (2005). A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J Neurosci*, *25*(43), 10049-10060.

Orban, G. A., Kennedy, H., & Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. *J Neurophysiol*, *56*(2), 462-480.

Orhan, A. E., & Ma, W. J. (2015). Neural population coding of multiple stimuli. *J Neurosci.* 3*5,* 3825-3841.

Pack, C. C., Hunter, J. N., & Born, R. T. (2005). Contrast dependence of suppressive influences in cortical area MT of alert macaque. *J Neurophysiol*, *93*(3), 1809-1815.

Pasternak, T., & Tadin, D. (2020). Linking Neuronal Direction Selectivity to Perceptual Decisions About Visual Motion. *Annu Rev Vis Sci*, *6*, 335-362.

Perrone, J. A., & Thiele, A. (2001). Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nat Neurosci*, *4*(5), 526-532.

Priebe, N. J., Cassanello, C. R., & Lisberger, S. G. (2003). The neural representation of speed in macaque area MT/V5. *J Neurosci*, *23*(13), 5650-5661.

Priebe, N. J., Lisberger, S. G., & Movshon, J. A. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *J Neurosci*, *26*(11), 2941-2950.

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

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci*, *2*(11), 1019-1025.

Rocchi, F., Ledgeway, T., & Webb, B. (2018). Criterion-free measurement of motion transparency perception at different speeds. *Journal of Vision*, *18*(4).

Rockland, K. S. (2002). Visual cortical organization at the single axon level: a beginning. *Neurosci Res*, *42*(3), 155-166.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*(11), 1421-1431.

Schoppmann, A., & Hoffmann, K. P. (1976). Continuous mapping of direction selectivity in the cat's visual cortex. *Neurosci Lett*, *2*(4), 177-181.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu Rev Neurosci*, *24*, 1193-1216.

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

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci*, *9*(4), 578-585.

Stoner, G. R., & Albright, T. D. (1992). Neural correlates of perceptual motion coherence. *Nature*, *358*(6385), 412-414.

Ungerleider, L. G., & Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *J Comp Neurol*, *248*(2), 190-222.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat Neurosci*, *5*(6), 598-604.

Wiesner, S., Baumgart, I. W., & Huang, X. (2020). Spatial Arrangement Drastically Changes the Neural Representation of Multiple Visual Stimuli That Compete in More Than One Feature Domain. *J Neurosci*, *40*(9), 1834-1848.

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

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

Zhang, L. Q., & Stocker, A. A. (2022). Prior Expectations in Visual Speed Perception Predict Encoding Characteristics of Neurons in Area MT. *J Neurosci*, *42*(14), 2951-2962.