Neural coding 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, population decoding, speed tuning, velocity, transparent motion, divisive normalization, segmentation, figure-ground segregation, natural scenes*

Running title: *Neural coding 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 of early human psychophysics experiments, Bryce Arseneau for animal training, and Drs. Jennifer Coonen and Kevin Brunner at the Wisconsin National Primate Research Center for excellent veterinary care and surgical assistance.

Grant support

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

Conflict of Interest

None.

ABSTRACT

Segmenting visual objects from each other and their background is critical for vision. Motion speed provides a salient cue for scene segmentation – an object moving at a speed different from its background is easier to be perceived. However, how the visual system represents and differentiates multiple speeds to achieve segmentation is largely unknown. We first characterized the perceptual capacity in segmenting overlapping stimuli moving simultaneously at different speeds. We then investigated the rule of how neurons in the motion-sensitive, middle-temporal (MT) cortex of macaque monkeys represent multiple speeds. We found that the responses of neurons to two speeds showed a robust bias toward the faster speed component when both speeds were slow (< 20º/s). Our finding can be explained by a divisive normalization model with a novel implication that the weights for the speed components are proportional to the responses of a population of neurons elicited by the individual components and the neurons in the population have a broad range of speed preferences. We also showed that it was possible to decode two speeds from MT population response in a way consistent with perception when the speed separation was large, but not when it was small. Our results provide strong support for the theoretical framework of coding multiplicity and probability distribution of visual features in neuronal populations and raise new questions for future investigation. The faster-speed bias would benefit 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), but 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 that belong to the same object together, whereas different motion velocities help to segment visual scenes. Here we investigated how the visual system represents and extracts multiple motion velocities. The extrastriate middle-temporal cortex (area MT) in primates 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 multiple 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. The goal of this study is to characterize the neural encoding of multiple speeds of overlapping stimuli in area MT and determine whether it is possible to decode multiple speeds from the population neural response in MT, in a manner consistent with perceptual segmentation.

It is conceivable that the responses of MT neurons elicited by two motion speeds may follow one of the following encoding 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) "slower-speed bias" toward the slower speed component, which may better represent slower speeds which are more probable in nature scenes (Weiss et al., 2002); 4) "faster-speed bias" toward the faster speed component, which may benefit the segmentation of a faster-moving stimulus from a slower background. Also, the encoding rule may be consistent across a full range of stimulus speeds or may change depending on the stimulus speeds. How a given neuron represents two speeds may also depend on the neurons' speed preference and the difference between two stimulus speeds.

Regarding neural decoding, previous studies successfully extracted single stimulus speeds from neuronal populations in area MT using decoders such as vector-averaging and maximum likelihood estimators (Lisberger and Movshon, 1999; Churchland and Lisberger, 2001; Priebe and Lisberger, 2004; Huang and Lisberger, 2009; Yang and Lisberger, 2009; Krekelberg et al., 2006a,

b; Krekelberg and van Wezel, 2013). However, it is unclear whether simultaneously presented multiple speeds can be extracted from population neural responses, which would be difficult to achieve for decoders that only read out a single value. Zemel and colleagues developed a decoding framework that recovers the probabilistic distribution of a stimulus feature (Zemel et al., 1998; Pouget et al., 2003). Decoders of this type remain to be tested with neurophysiology data to determine whether they can extract multiple speeds from population neural responses.

We investigated the rule by which neurons in area MT represent overlapping stimuli that moved simultaneously at multiple speeds. We first characterized the perception of these visual stimuli in human and monkey subjects and then recorded the neuronal responses from macaque MT. We made a novel finding that MT neurons showed a strong faster-speed bias when stimulus speeds were low, and as stimulus speeds increased the faster-speed bias gradually shifted to response averaging. We also found that information about multiple speeds was carried in the population neural response in MT and it was possible to extract either a single speed or multiple speeds from area MT in a way largely consistent with perception, but also with limitations when two stimulus speeds were less separated from each other. Our results provided new insight into the neural coding of multiple visual stimuli and the mechanism underlying scene segmentation. Our work also raised new questions for future investigation.

MATERIALS AND METHODS

We conducted psychophysical experiments using human subjects, and psychophysical and neurophysiological experiments using macaque monkeys. Visual stimuli were identical for all experiments except where noted.

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 consents were 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^o 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. 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 difference" and the speed of the faster component was always twice (x2) that of the slower component (see Fig. 1A, B). For each bi-speed stimuli, 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 stimuli: $Spd_{mean} = e^{[\ln(Spd_1)+\ln(Spd_2)]/2}$, 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 bi-speed stimuli 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 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 bi-speed stimuli by pressing one of two buttons (left or right) within a 1500-ms window. 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 (bi-speed stimuli 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 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-speed" 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 that a subject correctly picked the bi-speed stimuli 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 bi-speed stimuli and the corresponding single-speed stimulus, 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 SD set to 0 and 1, respectively. When the hit or false alarm rate was occasionally greater than 0.99, 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}.

Neurophysiological and psychophysical experiments

Subjects

Five adult male 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 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 mentioned above. 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 translated 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). Figure 1A, B illustrates the speeds of the stimulus components for all bi-speed stimuli used. 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. 1A). At x2 speed separation, the five 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. 1B). 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 8A). We varied the vector average direction of the two component directions across 360° in a step of 15° to characterize the direction-tuning curves of MT neurons in response to the bi-speed and bi-direction stimuli. 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 one stimulus or two overlapping stimuli moving in the same direction (the preferred direction of the recorded neuron) at different speeds. 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. 6). 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. 8). 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).

Analysis of population neural coding of multiple speeds

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 again 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 population neural response to the bi-speed stimuli and the corresponding single-speed stimulus moving at the log mean speed. Trial-by-trial population responses were generated randomly according to a Poisson process with the mean set to the trial-averaged neuronal response. For each speed combination, we generated 200 trials of responses to the bi-speed stimuli and the corresponding single-speed stimulus, respectively. Data were partitioned into training and testing sets using k-fold cross-validation (k = 40). The data 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. 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.

Population Decoding

We define a given probability distribution of stimulus speed as: $\emptyset_m = \{P_{m,j}\}$, in which $P_{m,j}$ is the probability of speed S_j, $j = 1, 2, 3, \ldots, 121$, and j evenly samples speeds from 1.25º/s to 80º/s (referred to as the "full speed range") in a natural logarithm scale and at a "*speed interval*" of 0.0347. Because $\emptyset_m$ is a probability distribution, $\sum_j P_{m,j} = 1$. m is an index for different distributions.

The estimated response (*ES*) of neuron i to the stimulus speeds with a probability distribution $\emptyset_m$ is a linear sum of the responses of neuron i to each single speed S_j within the full speed range, weighed by the probability of each speed in $\emptyset_m$. The probability can also be considered as the weight (signal strength) of the speed.

$$ES_i(\emptyset_m) = \sum_j P_{m,j} f_i(S_j) \ , \qquad (1)$$

where f_i is the spline-fitted speed tuning curve of neuron i in response to single speeds.

The estimated population response (*EP*) of N neurons to $\emptyset_m$ is:

$$EP_m(\ln(PS_i)) = ES_i(\emptyset_m), \qquad (2)$$

where PS_i is the preferred speed of neuron i, $i = 1, 2, 3, \ldots, N$. N = 100 in our neural data.

We then spline-fitted the estimated population response $EP_m(\ln(PS_i))$ using a smoothing parameter of 0.93, interpolating the PS within the full speed range from 1.25º/s to 80º/s in natural logarithm with 121 evenly spaced values. The spline-fitted estimated population response is represented as $spEP_m(\ln(PS_j))$, $j = 1, 2, 3, \ldots, 121$.

Similarly, we spline-fitted the recorded and normalized population neural response $RP_m(\ln(PS_i))$, $i = 1, 2, 3, \ldots, 100$, and interpolated the PS to the same 121-speed values in a logarithm scale within the full speed range as above. The spline-fitted, recorded population neural response is represented as $spRP_m(\ln(PS_j))$, $j = 1, 2, 3, \ldots, 121$.

The decoded probability distribution of the stimulus speed $\emptyset_e$ is the $\emptyset_m$ that maximizes the objective function (OF), which is defined as the negative value of the SSE (sum squared error) between the spline-fitted estimated population response and the recorded neural response:

$$OF(\emptyset_m) = -\sum_j \{[spEP_m(\ln(PS_j)) - spRP_m(\ln(PS_j))]^2\}, \qquad (3)$$

$$\emptyset_e = \arg\max_{\emptyset_m}[OF(\emptyset_m)] \ . \qquad (4)$$

Rather than finding an arbitrary distribution, we constrained $\emptyset_e$ to contain either a single speed with a probability (referred to as the "weight") of 1 or two speeds with the same or different weights that sum to 1.

Algorithm to search for the probability distribution of stimulus speed

We first searched for the best-fit distribution $\emptyset_{e1}$ that contained a single speed *SP* with non-zero probability (P=1) that gave rise to the maximum OF across the full speed range (OF_{max1}). We next searched for the best-fit distribution $\emptyset_{e2}$ that contained two speeds SP_1 and SP_2 with non-zero probability and gave rise to the maximum OF for two speeds (OF_{max2}). We varied the speed separation, the center position, and the probabilities of the two speeds. For each speed separation and center position, the probabilities of SP_1 and SP_2 were varied from 0 to 1 at a step of 0.01, with the constraint that they summed to 1. We searched the speed separation, $\ln(SP_2)$- $\ln(SP_1)$, from 0.0693 (i.e. 2 *speed intervals*) to 3.3271 (i.e. 96 *speed intervals*), in a step of 0.0693. The search range covered the speed ratio SP_2/SP_1 from x1.07 to x27.86, sufficiently broader than x2 and x4 used in our visual stimuli. For each speed separation, we started the search where the center position of the two speeds $[\ln(SP_1)+\ln(SP_2)]/2$ was in the middle of the 121 possible speed values, referred to as the "speed axis". We then moved the center position toward the left border of ln(1.25) at a step of 0.0347 to find the maximum OF value ($OF_{leftmax}$) along the left half of the speed axis. If the OF value at the center position next to the current position was higher, the search moved to the next position. Otherwise, the current position was considered a local maximum. After we found a local maximum, the search continued in the same direction for up to another 30 *speed intervals* until one of the component speeds hit a border, or 30 intervals were reached, or an OF value greater than the previous local maximum was found. If a larger OF was found, the local maximum was updated and the search jumped to that position, and the procedure repeated until $OF_{leftmax}$ was found. We then returned to the middle of the speed axis and searched speed pair toward the right border ln(80) to find the maximum $OF_{rightmax}$. The larger one of $OF_{leftmax}$ and $OF_{rightmax}$ was the maximum OF for two speeds (OF_{max2}). The $\emptyset_e$ was either $\emptyset_{e1}$ or $\emptyset_{e2}$, whichever gave rise to the larger value of OF_{max1} and OF_{max2}.

Discrimination readout speeds using a classifier

Based on our decoding results we evaluated the discriminability between the readout speed(s) of the bi-speed stimuli and the corresponding single speed stimulus moving at the log-mean speed. To correct for certain artifacts due to our decoding procedure (see Results), we first made the following correction: In a given trial, when the difference between the weights of two readout speeds was greater than 0.7, the trial was considered to have a single readout speed and the separation between the two readout speeds was set to zero. After this correction, we applied a speed separation threshold of x1.3 (i.e. 0.26 in natural log scale) to the distributions of the readout speed separation from trials that contained two speeds and a single speed. If a readout speed separation in a trial was greater (or less) than the threshold, the visual stimulus in this trial was considered to have two (or one) speeds. A hit was achieved when the readout speed separation of a trial was considered to have two speeds and the visual stimulus in the trial indeed had two speeds. A false alarm occurred when the readout speed separation of a trial was considered to have two speeds, whereas the visual stimulus in the trial only had one log-mean speed. We calculated discriminability *d'* using the hit rate and false-alarm rate as described in human psychophysics.

RESULTS

Perception of overlapping stimuli moving at different speeds

To establish the perceptual basis for our study, we first characterized how human subjects perceived overlapping stimuli moving at different speeds. We used the same visual stimuli in our psychophysics experiments as in our neurophysiology experiments and investigated how the mean stimulus speed and the separation between two speeds impacted the perceptual segmentation of speeds.

The visual stimuli were two overlapping random-dot patches presented within a stationary square aperture. 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. 1A). Another set of stimuli had a "small speed separation", and the speed of the faster component was twice (x2) that of the slower component. The five 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. 1B). Experimental trials of bi-speed stimuli that had large and small speed separations were randomly interleaved.

Human subjects first performed a standard 2AFC task to discriminate the bi-speed stimuli from the corresponding single-speed stimulus that moved at the log mean speed of the two component speeds (see Methods). At the large (x4) speed separation, all four subjects could well differentiate the bi-speed stimuli from the single-speed stimulus at speeds less than 20 and 80°/s (Fig. 1C). At the highest speeds of 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 x4 separation. Nevertheless, when the speeds of stimulus components were less than 20 and 40°/s, subjects on average could differentiate the bi-speed stimuli from the single-speed stimulus, but not at the highest speeds of 20 and 40°/s (mean d' = 0.17, STE = 0.1) (Fig. 1D).

In the standard 2AFC task, it is possible that subjects could not segment the bi-speed stimuli 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 stimuli may appear less uniform). Because our goal was to measure discriminability based on perceptual segmentation, we designed a 3AFC task to address this concern. In the modified task, subjects still discriminated the bi-speed stimuli 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-speed" choice). Therefore, the discrimination between the bi-speed and single-speed stimuli was only on trials when subjects thought they could perceive two different speeds of the bi-speed stimuli. The results from the 3AFC task were similar to those of the 2AFC task, with a slight reduction of d' across conditions as the "no two-speed trials" reduced discrimination performance. The small performance difference between the 2AFC (Fig. 1C, D)

and 3AFC (Fig. 1E, F) tasks suggests that human subjects generally relied on speed segmentation to perform the 2AFC task.

Our results showed that human subjects can segment overlapping stimuli based only on speed cues and the performance was better when the separation between the two-speed components was larger. Our results also showed that the ability to segment two motion speeds was reduced at fast speeds, even when the ratio between the speed components was fixed. Based on the results from the 3AFC task, we performed a two-way ANOVA, in which the two factors were the speed separation (x4 or x2) and the speeds of the stimulus components. We found both factors had significant effects on *d'* (for speed separation, $F(1,30) = 94.2$, $p = 9.11 \times 10^{-11}$; for stimulus speeds, $F(4,30) = 46.8$, $p = 1.76 \times 10^{-12}$). The discrimination performance was significantly better at x4 than at x2 speed separation (paired t-test, $p = 0.015$), except at the fastest speeds of 20 and 80°/s vs. 20 and 40°/s (Fig. 1E vs F). The discrimination performance was poor 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, $F(1,6) = 67.6$, $p = 1.75 \times 10^{-4}$) (Fig. 1E), and as the stimulus speeds increased from 10 and 20°/s to 20 and 40°/s for x2 separation (one-way ANOVA, $F(1,6) = 47.1$, $p = 4.71 \times 10^{-4}$) (Fig. 1F).

To measure the monkey's ability to segment overlapping stimuli moving at two speeds, we trained one monkey to perform a 2AFC task and discriminate the bi-speed stimuli from the corresponding single-speed stimulus moving at the log-mean speed (see Methods). In a given trial, either the bi-speed or the single-speed stimulus was presented. The monkey made a saccadic eye movement to one of two targets to indicate whether the stimulus had two speeds or a single speed. The monkey's performance at x2 speed separation (Fig. 1H) was very similar to that of humans (Fig. 1D of the 2AFC task). At x4 separation, the monkey's performance was generally better than at x2 separation (Fig. 1G vs 1H). As the stimulus speeds increased from 5 and 20°/s to 20 and 80°/s, the performance declined (Fig. 1G), similar to the human results (Fig. 1C). 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. 1G), whereas the averaged human performance was poor (also note that one human subject (NP) performed better than others at the fastest speeds) (Fig. 1C). The difference between the monkey and human results may be due to species difference and individual variability. It is also possible that at the fastest speeds, the monkey may use the appearance of the stimulus, rather than speed segmentation to perform the task. Although we could not rule out this possibility, the similar trend between the results of monkey and human subjects across speeds and speed separations suggests that the monkey's ability to segment overlapping motion speeds is similar to that of humans.

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 2 shows the responses from four example neurons. To visualize the relationship between the responses to the bi-speed stimuli (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 stimuli and its constituent single-speed components are aligned along a vertical line as illustrated in Figure 2A1.

We found that the relationship between the responses elicited by the bi-speed stimuli and the constituent components depended on the stimulus speeds. Figure 2A1-D1 shows the results obtained when the separation between the two component speeds was large (×4). The component speeds are shown in Figure 1A. When the two component speeds were slow (1.25 and 5°/s), the response to the bi-speed stimuli nearly followed the response elicited by the faster-speed component (the leftmost data points in Fig. 2A1-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 stimuli still followed the weaker response elicited by the faster component (Fig. 2A1). When the speeds of the two stimulus components were at 2.5 and 10°/s, the response elicited by the bi-speed stimuli 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. We found similar results when the speed separation between the two stimulus components was small (×2) (Fig. 2A2-D2).

We found the same trend in the neural responses averaged across 100 neurons (Fig. 3). 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. 3A1). 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. 3B1). 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. 3C1) and 29 neurons that preferred high speeds (Fig. 3D1), 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. 3D1). We found similar results at x2 speed separation (Fig. 3A2-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, at a given pair of stimulus speeds, we expressed the response R of a neuron elicited by the bi-speed stimuli as a weighted sum of the component responses R_s and R_f elicited by the slower and faster stimulus component, respectively:

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

in which, w_s and w_f are the response weights for the slower and faster component, respectively. 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.

Using this approach to estimate the response weights for individual neurons can be inaccurate because 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. 4). For each pair of stimulus speeds, we plotted (R - R_s) in the ordinate versus (R_f - R_s) in the abscissa. Figure 4A1-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:

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

Because all the regression lines in Figure 4 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. 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. 5A). 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. 4A2-E2, Fig. 5B, and Table 1).

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

Normalization model fit of MT responses to bi-speed stimuli

We found that a divisive normalization model (Carandini and Heeger, 2012) can well describe the MT responses elicited by the bi-speed stimuli. We used the following equation (Eq. 8) to fit each neuron's responses to the bi-speed stimuli across 5 speed combinations with either x4 or x2 speed separation:

$$R_{bi} = \frac{S_s^n}{S_s^n + \beta S_f^n + \sigma} R_s(V_s) + \frac{\beta S_f^n}{S_s^n + \beta S_f^n + \sigma} R_s\left(V_f\right) + c \quad . \qquad (8)$$

R_{bi} is the model-fitted response to the bi-speed stimuli; R_s is the response tuning to a single-speed stimulus; V_s and V_f are the slower and faster component speed, respectively; S_s and S_f are the population neural responses in MT to the slower and faster component speed, respectively, and were estimated based on the population-averaged speed tuning curve to single speeds of our recorded MT neurons (Fig. 6A); n, β, σ, and c were model parameters and followed the following constraints: $0 \leqslant n \leq 100$, $0.01 \leq \beta \leq 100$, $0 \leq \sigma \leq 500$, $0 \leq c \leq 100$. In the model, the component responses of a given neuron elicited by the slower and faster components are weighted by S_s and S_f, which are the population neural responses to the corresponding speed components. The parameter β appears in the denominator that controls the contribution of S_f to the response of the normalization pool. β also appears in the numerator for the component response elicited by the faster stimulus $R_s\left(V_f\right)$, which provides a control for weighting the component responses in addition to S_s and S_f. Therefore, we refer to this model as the "weighted normalization" model. In comparison, we also fitted the bi-speed response using a "tuned normalization" model (Rust et al 2006; Ni et al., 2012; Carandini et al., 1997), in which the parameter α ($0.01 \leq \alpha \leq 100$) only appears in the denominator (Eq. 9):

$$R_{bi} = \frac{S_s^n}{S_s^n + \alpha S_f^n + \sigma} R_s(V_s) + \frac{S_f^n}{S_s^n + \alpha S_f^n + \sigma} R_s\left(V_f\right) + c \quad . \qquad (9)$$

Intuitively, when the stimulus components move at low or intermediate speeds less than 20°/s, population-averaged MT response to the higher-speed component S_f is stronger than that to the lower-speed component S_s at the x4 (Fig. 6B) and x2 speed separation (Fig. 6C). This difference between S_f and S_s would weigh the faster stimulus component more than the slower component (Eqs. 8, 9). When both stimulus components move at speeds greater than 20°/s, S_f is smaller than S_s (Fig. 6B, C) and would weigh the faster component less than the slower component.

At x4 speed separation, the weighted normalization model accounted for on average 84.7% of the response variance across 100 neurons, which is significantly better than the tuned normalization model (mean = 78.3%) (one-tailed paired t-test, p = 0.0018) (Fig. 6D). At x2 speed separation, the weighted normalization model accounted for 97.7% of the response variance, which is also significantly better than the tuned normalization model (mean = 95.9%) (one-tailed paired t-test, p = 0.00029) (Fig. 6E).

Based on the model-fitted bi-speed responses across the population, 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 neural responses (Eqs. 6, 7). The weights obtained using the weighted normalization model matched the weights calculated based on the data well (R^2 = 0.925 for x4 and R^2 = 0.93 for x2 speed separation) (Fig. 6F, G). The weights obtained using the tuned normalization model fit also matched the data well but were slightly worse (R^2 = 0.90 for x4 and R^2 = 0.87 for x2 speed separation) (Fig. 6F, G).

For both x4 and x2 speed separations, the median β value of the weighted normalization model fit is 1.2, which is significantly different from one (Wilcoxon signed-rank test, p = 0.00011 and 0.00018 for x4 and x2 speed separation, respectively). This result indicates a stronger weighting for the faster speed component than the slower component across all stimulus speeds, in addition to the weighting prescribed by S_f and S_s dependent on the stimulus speeds. Taking the weighting contributed by β and S_f and S_s together, the overall response weight for the faster component is greater than the slower component at the low and intermediate speed range, but the two weights are similar at high stimulus speeds, which matches the neural data well (Fig. 6F, G). In contrast, the relative weights for the faster and slower component fitted with the tuned normalization model are only determined by S_f and S_s (Eq. 9), which predicts a stronger weight for the slower than the faster component at fast stimulus speeds, slightly deviating from the neural data at this speed range.

Time course of MT responses to bi-speed stimuli

We asked whether the bias toward the faster speed component occurred early in the neuronal response or developed gradually over time. Figure 7 shows the time course 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. 7A-C). The first 20-30 ms of the neuronal response elicited by the bi-speed stimuli 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 7A-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 7A-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. 7E). We found similar results when the speed separation between the two stimulus components was x4 (Fig. 7A1-E1) and x2 (Fig. 7A2-E2).

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

We showed that at low and intermediate speeds, MT response to bi-speed stimuli was biased toward the faster stimulus component when two overlapping stimuli 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 8A 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. 8A, 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).

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:

$$R_{bi}(\theta_1, \theta_2) = w_s R_s(\theta_1) + w_f R_f(\theta_2) + c\,, \qquad (10)$$

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 on average 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. 8B). 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 also 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). 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 and green curves in Fig. 5C, D), similar to the results from the fixation paradigm (Fig. 5A, B, and the blue curves in Fig. 5C, D). 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.

Population neural responses elicited by bi-speed and single-speed stimuli

We asked whether information of multiple speeds of overlapping stimuli was carried in the responses of neuronal populations in MT and whether multiple speeds can be decoded from the population neural response. Because we used identical visual stimuli when recording from neurons in different experimental sessions, this allowed us to pool the responses from different neurons to form a pseudo-population. We first examined the difference between the population neural responses elicited by the bi-speed stimuli and the corresponding single-speed stimulus that moved at the log-mean speed (see Methods of human psychophysics). Figure 9 shows the population neural response plotted as a function of neurons' PS, constructed from 100 neurons that we recorded using the fixation paradigm (see Methods). To capture the population neural response evenly across a full range of PS, we spline-fitted the recorded response elicited by the bi-speed stimuli (the red curves) and by the single, log-mean speed (the black curves) (Fig. 9). At x4 and x2 speed separations, the population neural 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. 9A-E). 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 inferred 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 very similar to that elicited by the single log-mean speed, with the two-speed population response being just slightly broader (Fig. 9F-J).

To evaluate the discriminability between MT population responses elicited by the bi-speed stimuli 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 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 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 x4 speed separation, the discriminability was generally very high and slightly decreased as the stimulus speed increased (Fig. 9K), which was generally consistent with our psychophysics results (Fig. 1C, E, G). One difference is that at 20 and 80°/s, the classifier's performance did not drop to a very low level as human performance (compare Fig. 9K with Fig. 1C, E), but was comparable to that of the monkey subject (Fig. 1G). At the highest stimulus speeds, it was possible that the monkey subject picked up other appearance differences between the bi-speed and single-speed stimuli to perform the task. The classifier can also pick up the difference in population neural responses to the bi-speed and single-speed stimuli (Fig. 9E). However, it is unclear whether this neural response difference can be mapped to the perception of two speeds versus one speed, or as a noticeable appearance difference between the two stimuli. At x2 speed separation, the classifier's performance (Fig. 9L) was similar to that of the human (Fig. 1D, F) and monkey (Fig. 1H) subjects, and was poorer than that at x4 speed separation (Fig. 9K).

Decoding either a single speed or two speeds from trial-averaged population neural response

We next asked whether it is possible to extract either a single speed or multiple speeds from the population neural response in MT. Our decoder found readout speeds and their weights that minimized the difference (in terms of sum squared error) between the estimated population response elicited by the readout speeds and the recorded population neural response (see Methods). Rather than searching for a probability distribution of speed with a general shape, we constrained the search to either one speed or two speeds. The weights of the speed components sum to one, akin to a probability distribution. Figure 10 demonstrates the decoding procedure and shows the results of extracting speeds from the recorded trial-averaged responses of 100 neurons to the bi-speed stimuli. To capture the population neural response evenly across the PS of the neurons, we spline-fitted the recorded population response and the estimated population response. The best-fit estimated population responses (Fig. 10 blue curves) were highly correlated with the recorded neural responses (Fig. 10 red curves) (for five speed pairs, $R^2 > 0.96$ at x4 speed separation; $R^2 > 0.99$ at x2 speed separation). At x4 speed separation, the decoder extracted two separate speeds for all speed combinations (Fig. 10 A-E). The readout speeds were generally close to the veridical stimulus speeds. At low stimulus speeds of 1.25 and 5º/s (Fig. 10A) and 2.5 and 10º/s (Fig. 10B), the decoded faster speed component had a higher weight than the slower component. At the highest speeds of 20 and 80º/s, the decoder extracted two speeds (Fig. 10E), whereas human subjects could not perceive two speeds (Fig. 1E) (see below). At x2 speed separation, the decoder extracted two

separated speeds only at low stimulus speeds of 1.25 and 2.5º/s (Fig. 10F). At higher stimulus speeds, the decoder extracted a dominant single speed that was between the two stimulus speeds, with or without a second nearby speed that had a very low weight (Fig. 10G-J). In contrast, human subjects could perceive two speeds when stimulus speeds were below 20 and 40º/s (Fig. 1F) (see below and Discussion).

Decoding speeds from trial-by-trial population neural responses

To determine the distribution of the readout speed across trials, we randomly generated 200 trials based on the responses of 100 MT neurons in our data sample. In each simulated trial, a given neuron's response was determined by a Poisson process with the mean set to the spike count averaged across the experimentally recorded trials. The trial-by-trial response of each neuron was normalized to construct the population response and then spline-fitted for decoding. Figure 11 shows the speeds extracted from the recorded neural responses to single stimulus speeds (Fig. 11A-G), and from the inferred responses to the log-mean speed of x2 speed separation (Fig. 11H-L). The decoder speeds matched the single stimulus speed very well (Fig. 11M).

Figure 12 shows the speeds extracted from the neural response to the bi-speed stimuli. The decoder often read out two speeds across trials. In some trials, the readout of one speed component had a very small weight. We considered a speed component invisible to an observer if it had a weight less than 0.15 (i.e. the weight of the other readout speed was greater than 0.85). Therefore, if the weight difference between two readout speeds was greater than 0.7, we considered the trial had a "single" readout speed. This usually happened when the readout speed that had a very small weight was either at one of the boundaries of the speed range (i.e. 1.25º/s or 80º/s) or separated from the other readout speed by the largest speed separation searched by the decoder (i.e. x27.86 or 3.33 in log scale) (see Methods). These small weights were likely artifacts due to boundary conditions in our search algorithm.

At x4 speed separation, the decoder was able to extract the speeds of the stimulus components (Fig. 12A-D), except at the fastest speeds of 20 and 80º/s. At low stimulus speeds of 1.25 and 5º/s, and 2.5 and 10º/s, the readout speed around the faster stimulus component had a higher weight than that around the slower stimulus component (Fig. 12A, B). At stimulus speeds of 1.25 and 5º/s, in trials that had two readout speeds (Fig. 12A, on white background), the faster readout speeds were close to the faster stimulus speed of 5º/s. The slower readout speeds were closely aligned with the slower stimulus speed of 1.25º/s (i.e. 0.22 on the logarithm scale), which was also the lower boundary of the speed range. The weights of the faster readout speeds (mean = 0.67) were significantly greater than those of the slower speeds (mean = 0.33) (paired t-test, $p = 2.33 \times 10^{-29}$). We cannot rule out the possibility that the slower readout speeds of 1.25º/s in some trials were due to boundary artifacts. If discounting the potential artifact, the weight for the readout speed near the faster stimulus speed of 5º/s would be even higher. In trials considered to have a

single readout speed, the readout was very close to the faster stimulus speed of 5º/s (Fig. 12A, on grey background). For some of these trials (at the top of Fig. 12A), the faster readout speed was near the upper-speed boundary of 80º/s (i.e. 4.4 on log scale) and had a very small weight (< 0.15). Those faster readout speeds were boundary artifacts. To address the artifact, rather than comparing the weights between the faster and slower readout speeds, we compared the weights of the readout speeds within a window around the faster and slower stimulus speeds. For x4 speed separation, we multiplied or divided the stimulus speed by 1.5 to set the window (i.e. log(stimulus speed) ± log(1.5) in logarithm scale). For the stimulus speed of 5º/s, the window was from 3.3º/s to 7.5º/s. For the left boundary speed of 1.25º/s, we chose the window to be from 1.25º/s to 2.8º/s so the window had the same width in logarithm scale as that around 5º/s. Taking all the single- and two-readout speed trials together, the trial-averaged mean weight for the readout around the faster stimulus speed (5º/s) was 0.72, significantly greater than the mean weight of 0.20 for the readout around the slower stimulus speed (1.25º/s) (paired t-test, $p = 4.04 \times 10^{-44}$).

At stimulus speeds of 2.5 and 10º/s, the decoder extracted two speeds that had a separation close to the veridical separation (Figs. 12B, 14B). In trials considered to have a single-speed readout, the readout speed was close to the faster stimulus speed of 10º/s. In some single- and two-readout speed trials, the slower readout speeds were also aligned with the 1.25º/s boundary and had a small weight, suggesting they were boundary artifacts. We looked at the weights of the speed readouts of all trials that were within a window (± log(1.5)) around the faster and slower stimulus speeds. The mean weight for the readout around the faster stimulus speed (10º/s) was 0.59, significantly greater than the mean weight of 0.04 for the readout around the slower stimulus speed (2.5º/s) (paired t-test, $p = 4.72 \times 10^{-56}$).

At stimulus speeds of 5 and 20º/s, nearly all trials had two readout speeds with a separation well aligned with the veridical speed separation (Figs. 12C, 14C). The faster readout speeds no longer showed a stronger weight. The mean weight for the faster readout speed was 0.43, significantly smaller than the mean weight of 0.57 for the slower readout speed (paired t-test, $p = 6.0 \times 10^{-10}$). At stimulus speeds of 10 and 40º/s, the decoder was able to extract two speeds for most of the trials (Fig. 12D). A small percentage of the trials (about 10%) were considered to have a single readout speed, which was close to the log mean speed of the two stimulus speeds (20º/s) (Fig. 12D on the grey background).

At the fastest stimulus speeds of 20 and 80º/s, about 40% of the total trials were considered to have only a single readout speed, which was near the log mean speed of the stimulus components (40º/s) (Fig. 12E). In other trials, the decoder extracted two speeds – the slower readout speeds were generally higher than the slower stimulus speed (20º/s) and the faster readout speeds were aligned with the faster stimulus speed (80º/s), which was also the upper boundary speed. However, an examination of the objective function as the decoder searched for the best-fit population response across speed separations revealed that the trial-averaged objective function was flat

within a big range of speed separations (Fig. 13A). Across all trials, the mean objective function value peaked at speed separation of x3.25 (mean OF = -0.17, STD=0.14). However, the peak value is not significantly different from the mean objective function value at the largest speed separation (x27.86) searched (mean OF = -0.19, STD=0.14) (paired t-test, p=0.31). The flat objective function suggests high uncertainty of the extracted speed separation at this speed pair.

We divided the trials into two subgroups considered having either one or two readout speeds and calculated the objective function for each subgroup. For trials considered to have one readout speed, the mean objective function showed a peak at the speed separation of x27.86 (3.3 on the log scale), which was the largest speed separation searched (Fig. 13E). As the searched speed separation increased, the dominate faster readout speed approached the log mean speed and the mean weight increased to 0.94, whereas the slower readout speed approached the lower boundary speed with the weight diminishing to negligible 0.06 as a likely artifact (Fig. 13F). For trials considered to have two readout speeds, objective function peaked at the speed separation of x3.25 (1.18 in log scale) (Fig. 13B), corresponding to two readout speeds of 17.8 and 58.0º/s (Fig. 13C). Furthermore, we compared the population neural responses averaged across the one-readout-speed trials and the two-readout-speed trials. The spline-fitted population responses of the two subgroups were highly correlated (R^2=0.99) and statistically indistinguishable (paired t-test, p=0.30) (Fig. 13D). This indicates that a tiny change in the population response (e.g., a slightly higher peak near log mean speed of 3.7) would lead the decoder to exact one speed rather than two speeds (Fig. 13D). In other words, the decoder was uncertain about how many speeds were in the visual stimuli and therefore had difficulty segmenting the visual stimuli at these fast stimulus speeds of 20 and 80º/s.

At x2 speed separation, the decoder was not able to extract two speeds of the stimulus components, except at the slowest speeds of 1.25 and 2.5º/s (Fig. 12F-J). At stimulus speeds of 1.25 and 2.5º/s (Fig. 12F), in 38% of total trials that were considered to have a single readout speed, the readout speed was close to the faster stimulus speed of 2.5º/s (i.e. 0.92 in log scale) (mean = 1.97º/s, STD = 1.08). In trials that had two readout speeds, the slower readout speeds roughly followed the slower stimulus speed (1.25º/s), which was also the lower boundary of the speed range (Fig. 12F). At stimulus speeds higher than 1.25 and 2.5º/s, most trials were considered to have a single readout speed (Fig. 12G-J). The mean speeds of the single readout-speed trials were 3.9º/s (STD = 1.07), 7.3º/s (STD = 1.99), 13.5º/s (STD = 1.06), and 31º/s (STD = 1.07), respectively, for stimulus speeds of 2.5 and 5º/s, 5 and 10º/s, 10 and 20º/s, and 20 and 40º/s. These mean readout speeds were close to the log mean speeds of the two stimulus speeds (3.54º/s, 7.07º/s, 14.14º/s, and 28.28º/s, respectively).

Discrimination between single- and bi-speed stimuli based on decoded speeds

We used the decoding results to perform a discrimination task similar to that used in our psychophysical experiment. Figure 14 shows the distributions of the speed separation between two readout speeds extracted from the population responses to the bi-speed stimuli and the correspondingly single log-mean speed. As above, when the difference between the weights of two readout speeds in a trial was greater than 0.7, the trial was considered to have a single readout speed and the speed separation was set to zero. At x4 speed separation, the separations between the readout speeds extracted from the response to the bi-speed stimuli were generally matching the veridical speed separation and were larger than those extracted from the response to single log-mean speed (Fig. 14A-E). Based on the distributions of the decoded speeds, we used a speed separation threshold of x1.3 (i.e. log scale of 0.26) to distinguish single- and bi-peed stimuli and to evaluate the hit rate and false alarm rate. The exact choice of the threshold within a range from x1.1 to x1.7 did not change our results qualitatively. We calculated *d'* (see Methods) to measure the ability to discriminate the bi-speed stimuli from the corresponding single-speed stimulus. At x4 speed separation, the *d'* (Fig. 14K) was similar to the psychophysical performance of the monkey subject (Fig. 1G), reaching its peak at 5 and 20º/s. Although *d's* at stimulus speeds of 1.25 and 5º/s and 2.5 and 10º/s were smaller than those of human subjects (Fig. 1C, E), the fact that in many trials the readout speeds matched the faster stimulus speeds (Fig. 12A, B) indicated that the decoder was able to segment the visual stimuli when stimulus speeds were low. If the discrimination were not based on the speed separation as in Figure 14, but rather on whether the readout speed matched the speed of one of the stimulus components, the discriminability would be even higher and better match the human performance at low stimulus speeds.

At x2 speed separation, except at 1.25 and 2.5º/s, the distribution of the speed separation extracted from the response to the bi-speed stimuli was similar to that extracted from the inferred response to single log-mean speed (see Methods) (Fig. 14F-J). At some stimulus speeds, the decoded speed separation extracted from the response to the log mean speed even had slightly more trials showing two readout speeds (Fig. 14G-H), likely due to more noise in the inferred response to the log-mean speed than in the recorded response (see Fig. 11 H-L). The *d'* calculated based on the decoded speed separation (Fig. 14L) was much smaller than the psychophysical performance of human and monkey subjects (Fig. 1D, F, H), suggesting that the decoder was not able to segment the visual stimuli at x2 speed separation, except at the lowest speeds of 1.5 and 2.5º/s.

DISCUSSION

We investigated the neural code in area MT that represents not only a single speed, but also the multiplicity of speeds. Using the same visual stimuli as in our neurophysiology experiments, we showed that human subjects could segment overlapping stimuli based only on speed cues, and the performance reduced as the separation between two speeds was smaller. We also showed that, with a fixed speed separation, perceptual segmentation became harder at fast speeds. These results

are consistent with previous psychophysical studies (Masson et al., 1999; Rocchi et al., 2018). We employed a novel 3AFC task that combined the classification of whether a stimulus had one or two speeds and the discrimination between two-speed and single-speed stimuli (Fig. 1E, F). This approach allowed us to characterize discriminability based on the perception of multiple speeds, rather than other perceptual appearances of the stimuli. This 3AFC task not only has the advantage of a discrimination task that reduces individual biases, but also a subjective report task that pinpoints a perceptual trait of the stimulus. Our behavioral experiment also suggests that macaque monkeys have a similar capacity in discriminating two-speed and single-speed stimuli as humans, consistent with previous studies demonstrating that visual perceptions of macaque monkeys are very similar to humans (e.g. Huang et al., 2002; Krekelberg et al., 2006b). By recording the electrical activity of single neurons from macaque monkeys, 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 (less than 20º/s). We also showed that it was possible to extract two speeds of MT population responses in a way consistent with human visual perception when the speed separation was large, but not when the speed separation was small. Our results revealed the rule of encoding multiple speeds in area MT and demonstrate a population neural code enabling the decoding of multiple motion speeds and perceptual segmentation.

Encoding of multiple speeds of overlapping stimuli and the underlying mechanisms

Our finding of the faster-speed bias at low stimulus speeds cannot be explained by attention 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. Instead, our encoding results can be largely explained by the framework of divisive normalization (Carandini and Heeger, 2012). Our findings also provide several new insights into the mechanisms of representing multiple speeds.

We showed that neuronal response elicited by the bi-speed stimuli can be described by a weighted sum of the responses to the individual speed components. Importantly, the weight for a speed component is proportional to the averaged (or summed) response of a population of neurons elicited by that speed component (S_s and S_f in the numerator of Eqs. 8 and 9). We name this neuron population the "weighting pool". The weighting pool is not necessarily the same as the "normalization pool", which is the neuron population that contributes to the response in the denominator of the normalization model. Our finding of the faster-speed bias at low stimulus speeds suggests that the weighting pool is composed of neurons with a broad range of speed preferences. In this way, the averaged/summed population response depends only on the stimulus speed and is invariant to the individual neuron's speed preference. In our data set, MT population-averaged speed tuning curve peaks around 20º/s (Fig. 6A), 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 so well, regardless of the speed preferences of the individual neurons (Fig. 4). Neurons with similar speed preference 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. 3B).

Previous studies that characterized the neural representation of multiple stimuli have used stimulus strength (e.g. luminance contrast and motion coherence) to weigh the component responses in the normalization model (Busse et al., 2009; Ni et al, 2012; Xiao et al., 2014; Heuer and Britten, 2002; Morgan et al., 2008). Our current study extends these previous findings by showing that weighting by the population neural response elicited by the stimulus component provides a more parsimonious explanation even when it is not obvious how to define the stimulus strength. Bao and Tsao (2018) suggest that responses of neurons in category-selective regions of the inferotemporal cortex to multiple objects are weighted by the responses from neighboring neurons that have the same category selectivity. Our finding is consistent with their study and further constrains the extent of the weighting pool to include neurons in MT with a broad range of preferred speeds in response to multiple speeds of overlapping stimuli.

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. 7A-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. 7A-C), suggesting that divisive normalization requires additional processing time and may involve neural circuits within MT. 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.

Population neural coding of multiple speeds and comparison with perception

Our decoding approach is to apply an encoding rule and find the visual stimuli that generate a population response best matching the recorded neural response, akin to the forward encoding model often used in brain imaging studies for decoding (e.g. Vintch and Gardner 2014). We

assumed an encoding rule of linear combination in the decoder – a neuron's response to multiple speeds is the linear sum of the neuron's responses to individual speed components presented alone based on the neuron's speed tuning curve, weighted by the strength (or probability) of each speed component. This encoding model can be considered a special case of convolution, with the impulse response function (kernel) being a single value in response to a given speed.

This "encoding model for decoding" differs from how MT neurons actually encode multiple speeds as characterized in this study. Our interpretation is that the brain may take linear combination as the default encoding rule and may not be aware of the actual encoding rule. As the result, for two stimulus speeds that have the same strength and are encoded by a rule emphasizing the faster speed, the brain would extract the faster speed with a stronger weight than the slower speed, as found in our decoding results at low speeds (Fig. 12). Alternatively, if the encoding model for decoding captures the bias toward the faster speed, the brain would then likely extract two speeds with equal weight. The assumption that the brain makes about the neural encoding rule is a likely source for visual perception and illusion. Our finding that the speed readouts had a higher weight for the faster speed component suggests that the faster component may be more salient perceptually at a low-speed range, a prediction that can be tested in future psychophysics experiments.

Theoretical studies have proposed neural coding of probability distribution and multiplicity of a visual attribute (Pouget et al., 2000). The key idea of this framework is that neurons are not coding a single stimulus value, but rather coding the full distribution of the stimulus in a sense of either multiplicity or probability distribution (Zemel et al., 1998; Pouget et al., 2003). However, neurophysiological evidence supporting this framework on coding multiplicity is limited. Previous studies have not demonstrated the ability to extract multiple speeds from population neural responses. Our results provide strong experimental support for this framework of coding multiplicity. Our decoding analysis reveals that population neural response in MT carries information about multiple speeds of overlapping stimuli and it is possible to extract multiple speeds and their weights even when the population neural response has a unimodal distribution.

At large (x4) speed separation, our decoding results captured several key features of human and monkey's perception of multiple speeds – the decoded speeds support perceptual segmentation at low to intermediate speeds and show the difficulty of segmentation at high speeds (Figs. 12, 14). At low stimulus speeds (1.25 and 5º/s, 2.5 and 10º/s), the decoded speed is mainly aligned with the faster stimulus speed (Fig. 12A-B), indicating strong segmentation of the faster speed component. This also explains the reduced performance of perceiving two speeds as found in the decoding results (Fig. 14K) and monkey psychophysics experiment (Fig. 1G), because only one (the faster) speed is perceived. At intermediate speeds (from 5 and 20º/s to 10 and 40º/s), the decoded speed separation matches the veridical separation (Figs. 12C-D, 14K). Also consistent with perception, our decoding results show that the ability to segment two speeds is poorer at smaller speed separation (x2) (Fig. 14L).

Our decoding results also have limitations. At small (x2) speed separation, our decoding results showed very little segmentation (Fig, 12G-J, 14L), except at very low speeds. This is different from the perception at stimulus speeds less than 20 and 40º/s (Fig. 1). What are the potential reasons for the decoder's inadequacy to segment small speed separation? The best-fit population response predicted by the encoding rule of the decoder matched the neural responses remarkably well ($R^2 > 0.99$ for all five speed pairs of x2 separation, Fig. 10F-J). Because we found the same results when performing decoding based on neural responses averaged across experimental trials (Fig. 10G-J), this inadequacy was unlikely due to our assumption of the trial-by-trial response variability following a Poisson process, nor due to the lack of consideration of noise correlations (Zohary et al., 1994; Huang and Lisberger, 2009).

We think several reasons may explain this discrepancy. First, it may be due to the limitation of the linear encoder. If the predicted response to multiple speeds involves a nonlinear combination of the responses to individual speed components, the same population neural response may be best fitted with two speeds that have a wider (or narrower) separation than those decoded using a linear encoder. Second, it may be due to the choice of the "objective function". Our decoder minimized the sum squared error between the predicted population response and the recorded neural response. In contrast, Zemel et al. (1998) found motion directions that maximized the posterior probability $P(\boldsymbol{s}|\boldsymbol{r})$ using a maximum a posteriori (MAP) estimate. It remains to be determined whether maximizing the posterior probability (or other measures) can improve the resolution of segmenting multiple speeds. Third, neuronal response to two stimuli may fluctuate from trial to trial between representing one stimulus component from the other (Li et al., 2016; Caruso et al., 2018). If this trial-varying stimulus multiplexing is also used for representing two stimuli with a small speed separation, information about individual speed components would be lost if to-be-decoded neuronal responses were averaged across experimental trials (with added variability based on a Poisson process), as in our decoding procedure. Fourth, we found that MT population response to two speeds at x2 separation differed from the response to the log mean speed just slightly (Fig. 9F-J). Although a classifier was able to pick up this difference (Fig. 9L), it is unclear whether this small response difference in MT was sufficient to support speed segmentation. Other brain areas may be better suited to represent multiple speeds with a small separation. Finally, although we characterized a monkey's discrimination performance in a psychophysics experiment, we did not measure monkey subjects' perception during neurophysiological recordings. We cannot rule out the possibility that when performing a fixation task, the monkeys may not be able to segment two speeds with a small (x2) separation. Future study that combines a discrimination task with neural recording will help to provide a more definite answer. Together, our results demonstrated the limitation of relying on trial-averaged population neural response in MT to segment similar speeds and raised new questions for future investigation.

Conclusions and functional implication of faster-speed bias

In summary, we characterized the encoding rule of representing multiple motion speeds of overlapping stimuli in area MT. We made a novel finding that neuronal responses in MT to two

speeds showed a robust bias toward the faster speed component when the stimulus speeds were slow, but not when they were fast. The responses to the bi-speed stimuli can be explained by a divisive normalization model, with the weights for the stimulus components proportional to the responses of a population of neurons (i.e. the weighting pool) elicited by the individual stimulus components. Our results also help to constrain the extent of the weighting pool to include neurons that have a broad range of speed preferences when representing multiple speeds of overlapping stimuli. The decoding results match the visual perception well and can account for perceptual segmentation for speeds with a large separation, but have limitations for small speed separation. Our neural data and decoding results support the theoretical framework of coding stimulus multiplicity, with the key idea that neuronal response can represent the distribution of the visual stimulus.

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 slower speeds are more likely to occur than faster speeds in natural scenes (Weiss et al., 2002; Stocker and Simoncelli, 2006; Zhang and Stocker, 2022). Our finding of faster-speed bias seems to disagree with the prior probability of single speeds having a preference for lower speeds, in the sense of efficient coding. However, if in the natural environment a figural object tends to move faster than its background, a faster-speed bias would help to identify the figure and therefore benefit figure-ground segregation. To test this functional implication of faster-speed bias, future study is needed to characterize natural scene statistics of speeds for figural objects and their background.

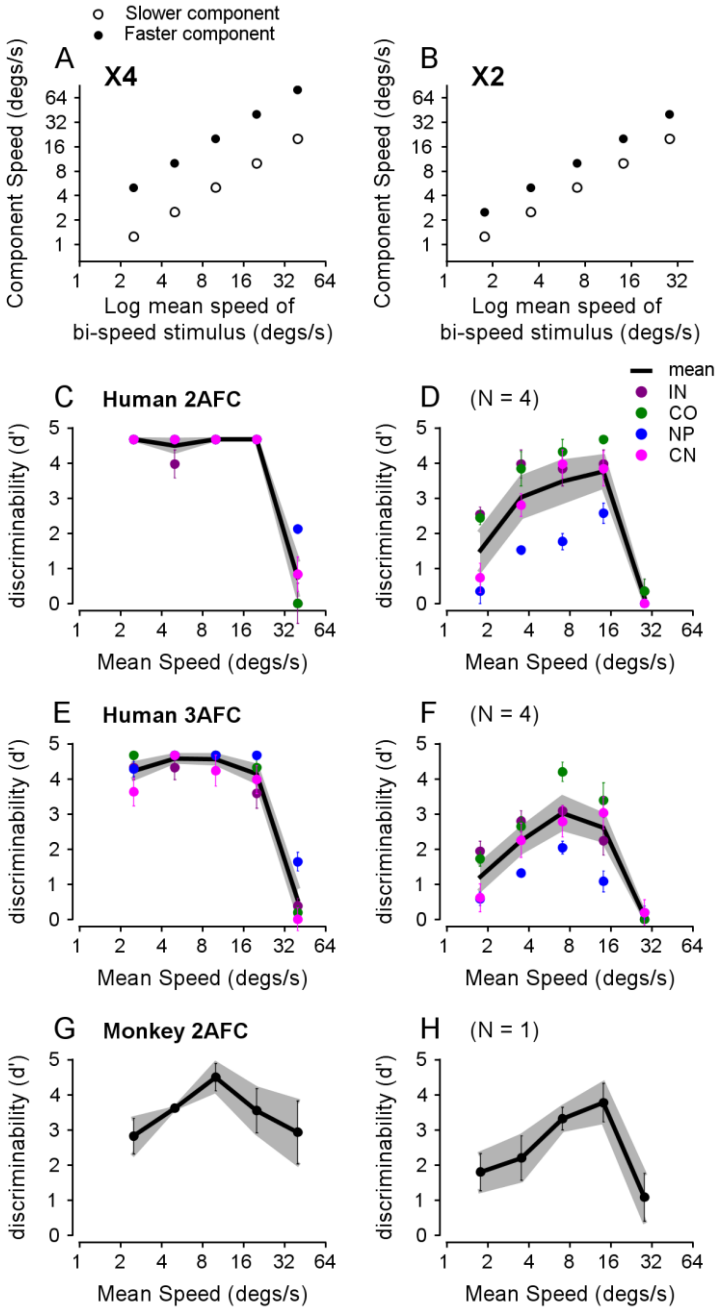


Figure 1. Visual stimuli and psychophysical performance of human and monkey subjects. **A-B.** Motion speeds of visual stimuli. The speeds of two stimulus components were plotted versus the log mean speed of each bi-speed stimulus. **C-D.** Discriminability of four human subjects performing a standard 2AFC task. **E-F.** Discriminability of the same four human subjects performing a 3AFC task. **G-F.** Discriminability of a monkey subject (BJ) performing a 2AFC task. **A, C, E, G**. X4 speed separation. **B, D, F, H**. 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.

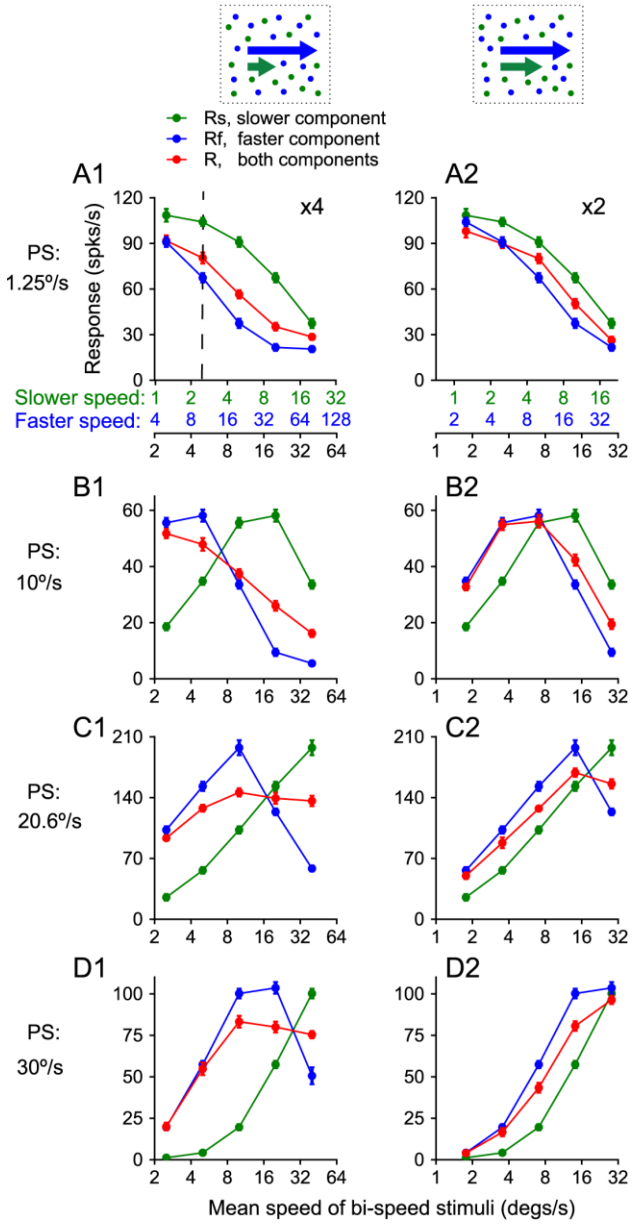


Figure 2. Speed tuning curves of four example neurons to bi-speed stimuli and constituent single-speed components. A. Illustration of the visual stimuli and the response tuning curves of an example neuron. Green and blue dots in the diagram indicate two overlapping achromatic random-dot patterns moving 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.** The four example neurons are sorted by their preferred speeds (PS) from slow to fast. **A1-D1.** X4 speed separation. **A2-D2.** X2 speed separation.

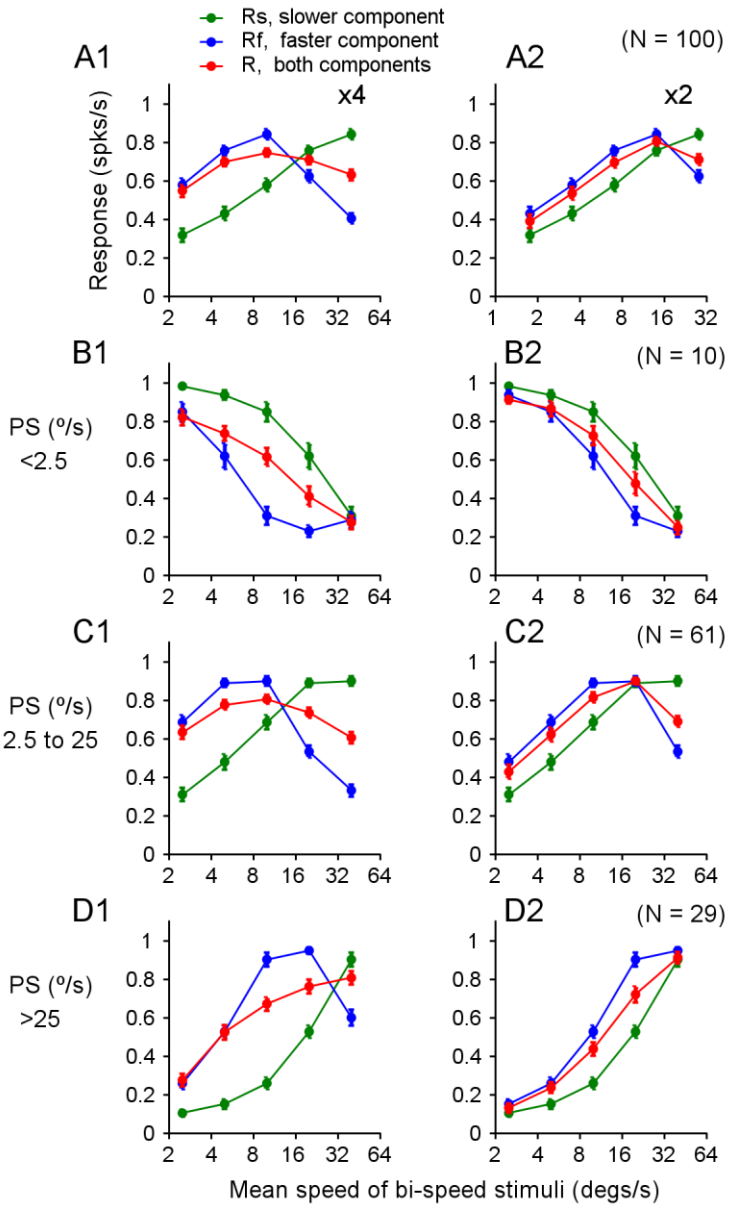


Figure 3. Population-averaged speed tuning curves to bi-speed stimuli and constituent single-speed components. Speed tuning curves averaged across **A.** 100 neurons in our dataset. **B.** 10 neurons that had PS lower than 2.5º/s. **C.** 61 neurons that had PS between 2.5 and 25º/s. **D.** 29 neurons 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.

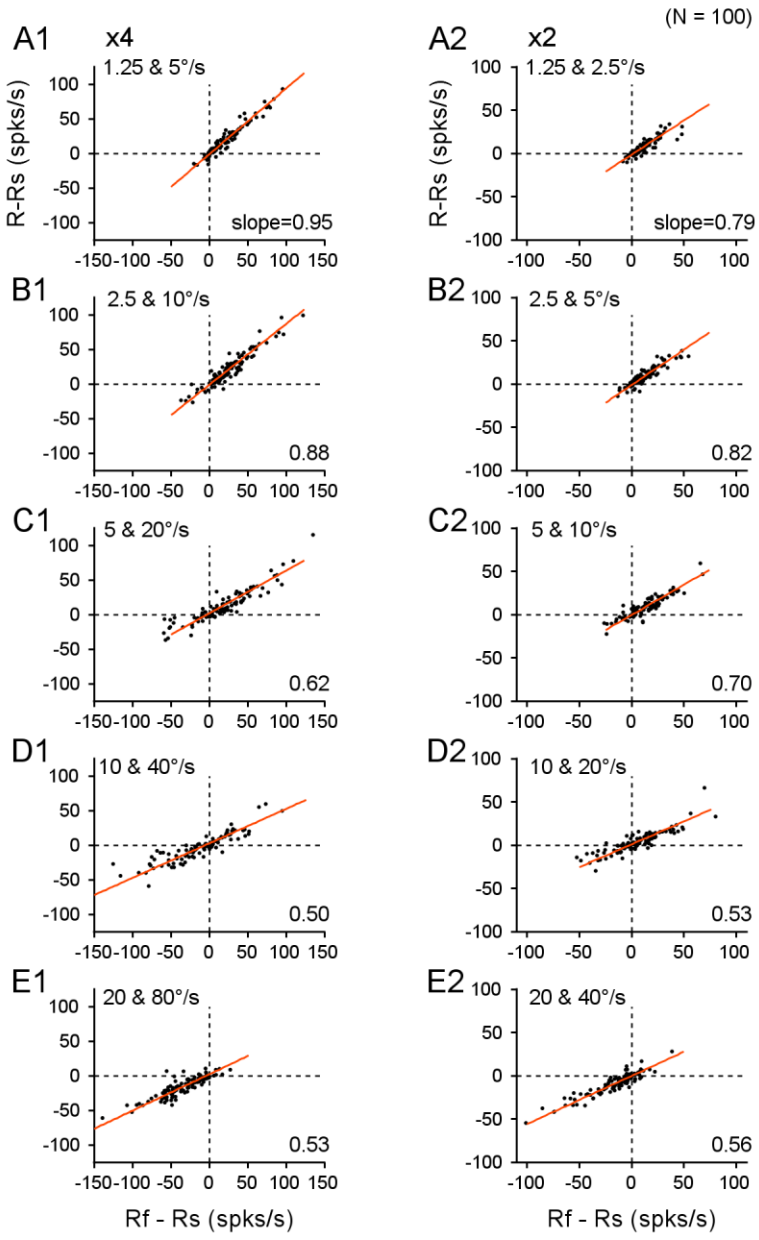


Figure 4. Relationship between the responses to bi-speed stimuli and constituent stimulus components. Each panel shows the responses from 100 neurons. Each dot represents the response from one neuron. The ordinate shows the difference between the responses to a bi-speed stimulus and the slower component (R - R_s). The abscissa shows the difference between the responses to the faster and slower components (R_f - R_s). The type II regression line is shown in red. **A1-E1**. X4 speed separation. **A2-E2.** X2 speed separation.

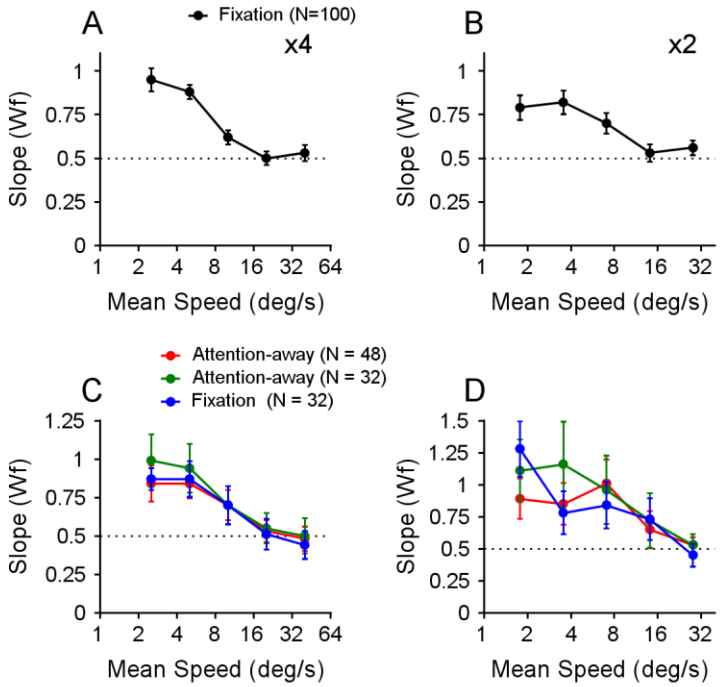


Figure 5. Response weights for the faster stimulus component. **A-B.** Results of linear regression based on the responses from 100 neurons recorded using a fixation paradigm. **C-D.** Results of linear regression based on the responses from 48 neurons recorded using an attention-away paradigm. **A, C.** X4 speed separation. **B, D.** X2 speed separation.

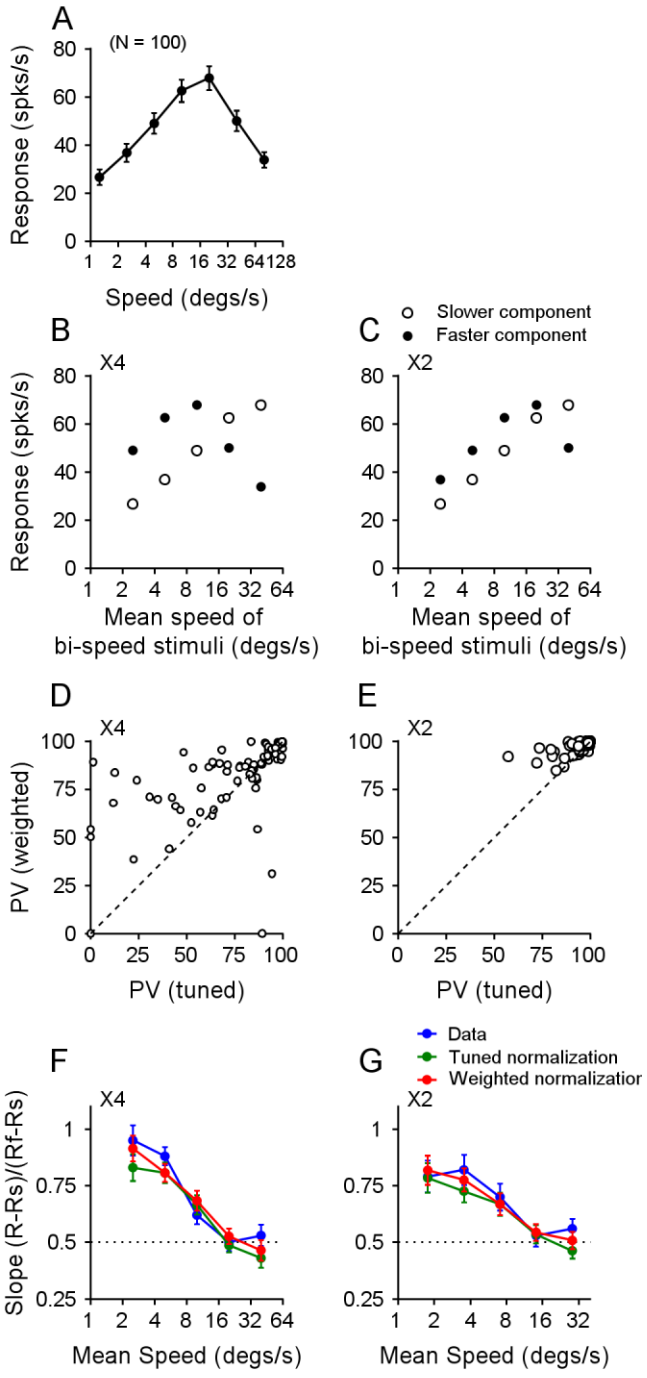


Figure 6. Normalization model fit of MT responses to bi-speed stimuli. A. Speed tuning curve to single-speed stimulus averaged across 100 recorded MT neurons in our data set. **B-C**. Population-averaged responses to slower (open circle) and faster (solid circle) speed components. The convention is the same as in Figure 1A, B. **D-E.** Comparison of percentage of variance accounted for by the weighted and tuned normalization model. **F-G.** The response weights for the faster component were calculated based on the data (blue), weighted (red), and tuned (green) normalization model. **B, D, F.** X4 speed separation. **C, E, G.** X2 speed separation.

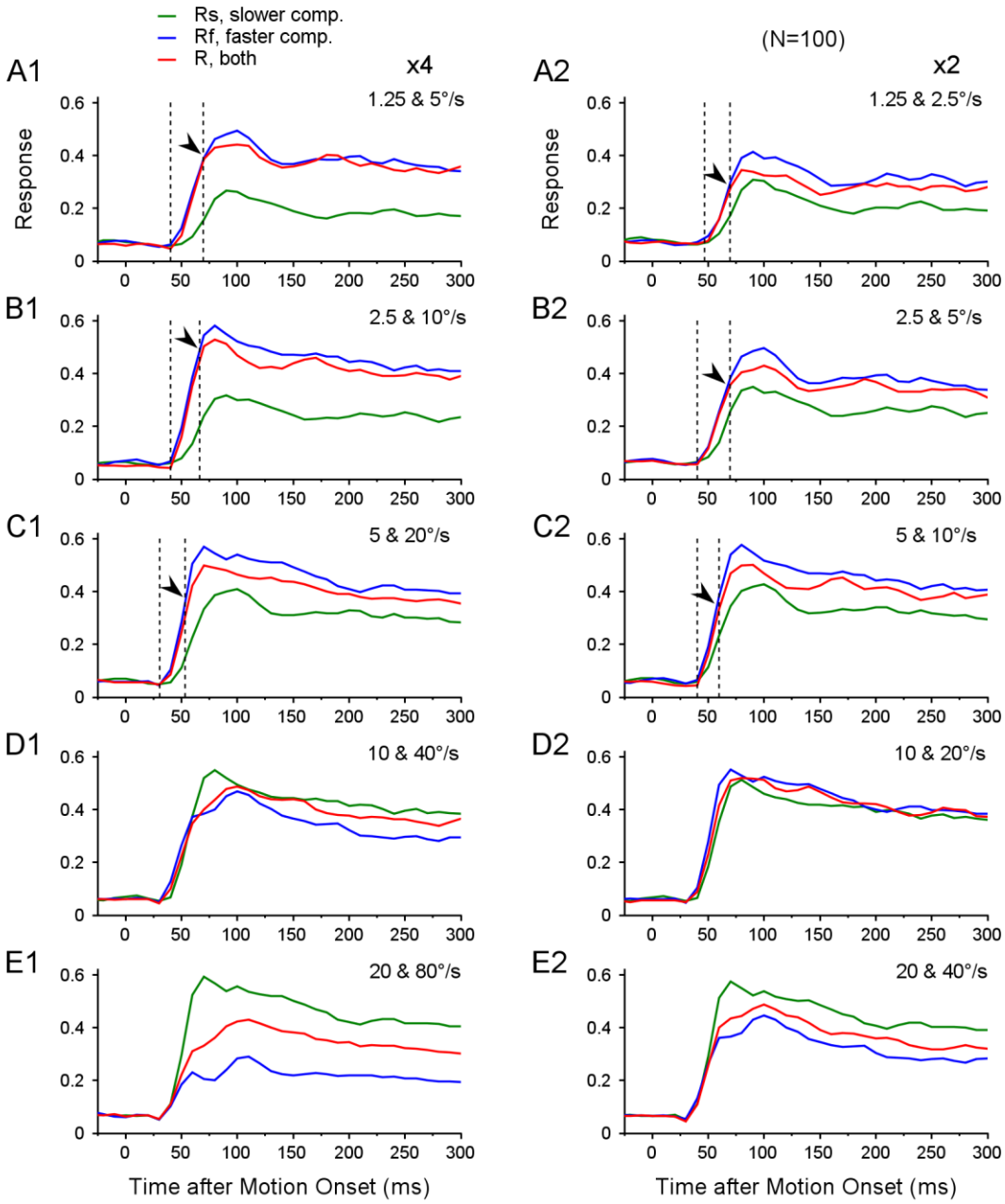


Figure 7. Time course of MT response to bi-speed stimuli averaged across neurons. Peristimulus time histograms (PSTHs) were averaged across 100 neurons. The bin width of PSTH was 10 ms. **A1-E1.** X4 speed separation. **A2-E2.** X2 speed separation. In A-C, the left dash line indicates the latency of 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.

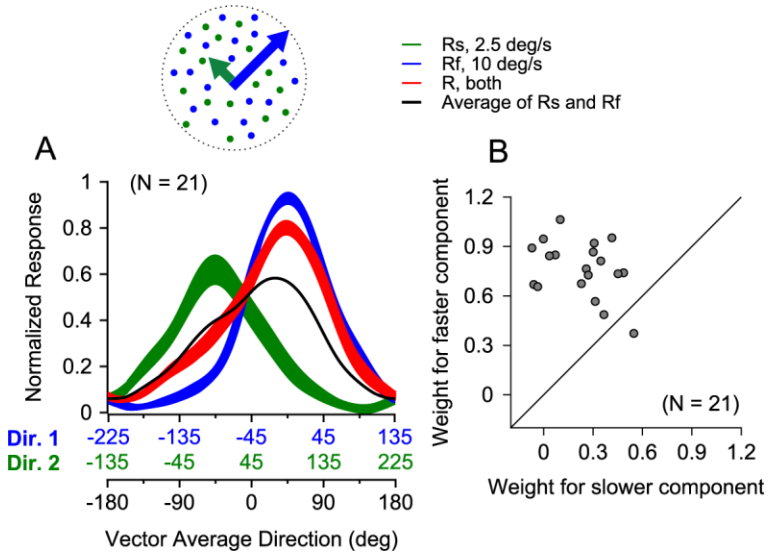


Figure 8. MT responses to two-speed stimuli that moved in different directions. A. Population-averaged direction tuning curves of 21 neurons in response to stimuli moving at two speeds and in two directions separated by 90º (red). The component direction Dir. 1 (blue) moved at 10º/s and the component 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 vector-average (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.

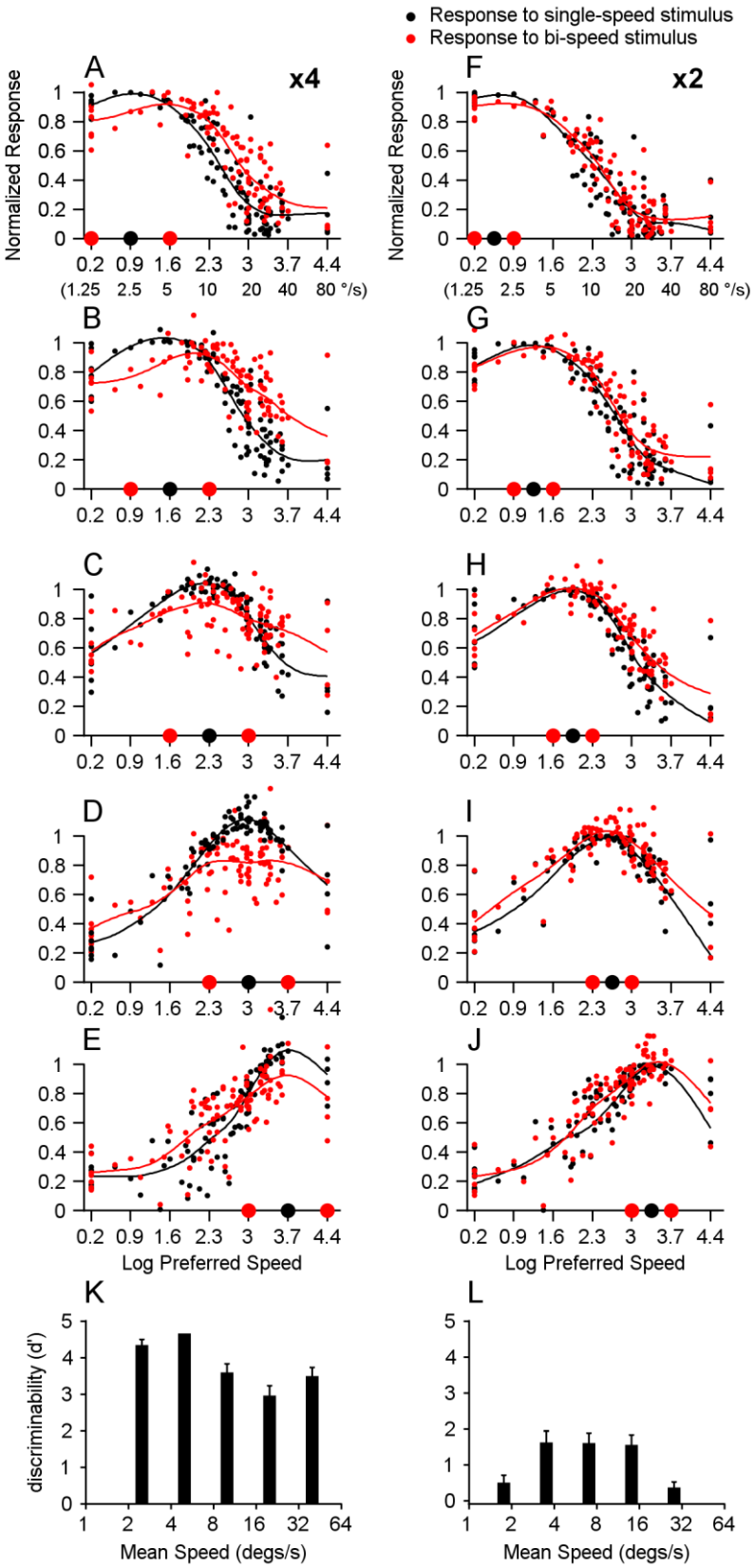


Figure 9. Population neural responses elicited by the bi-speed and single-speed stimuli and the performance of a linear classifier. A population of 100 neurons was constructed by pooling across recordings in different experimental sessions. Each neuron's response was averaged across experimental trials 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 preferred speed (PS) of the neuron in the natural logarithm scale. The curves represent the spline-fitted population neural responses. Red: response to the bi-speed stimuli; Black: the response to the corresponding single, log-mean speed. **A-E**. X4 speed separation. The speeds of the bi-speed stimuli are 1.25 and 5º/s (A), 2.5 and 10º/s (B), 5 and 20º/s (C), 10 and 40º/s (D), 20 and 80º/s (E). **F-J**. X2 speed separation. The speeds of the bi-speed stimuli are 1.25 and 2.5º/s (F), 2.5 and 5º/s (G), 5 and 10º/s (H), 10 and 20º/s (I), 20 and 40º/s (J). Two red dots on the X-axis indicate two stimulus speeds; the black dot indicates the log-mean speed. **K-L.** Performance of a linear classifier to discriminate the population neural responses to the bi-speed stimuli and the corresponding single log-mean speed. Error bars represent STE. **K.** X4 speed separation; **L.** X2 speed separation.

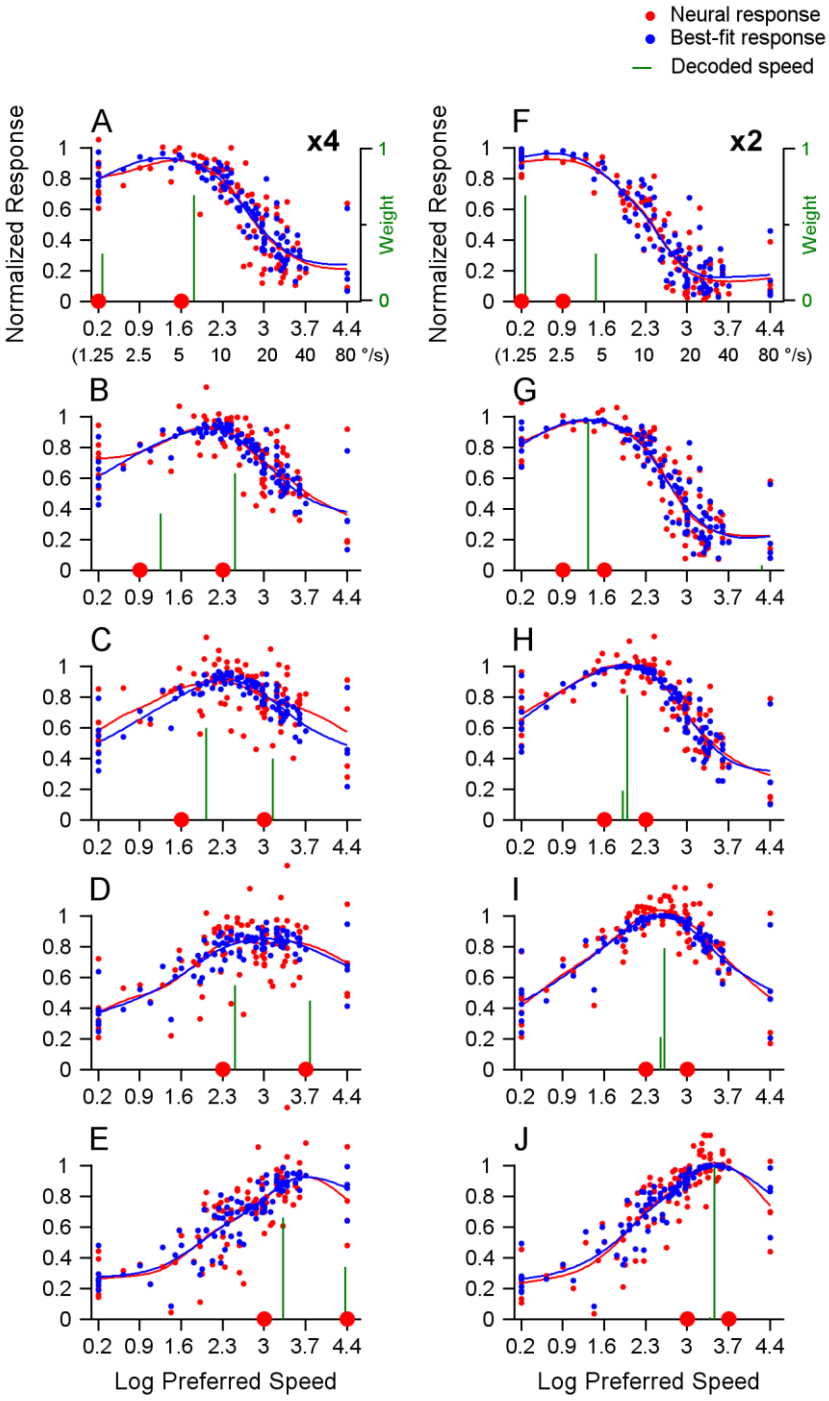


Figure 10. Illustration of the decoding procedure and extraction of speeds from trial-averaged population neural response to the bi-speed stimuli. A-E. X4 speed separation. F-J. X2 speed separation. The neural population contains 100 neurons as in Figure 9. Each red dot represents the trial-averaged response from one neuron plotted as the PS of the neuron in the natural logarithm scale. The red curve represents the spline-fitted population neural response. The decoder found either one speed or two speeds with different weights (green bars on the X-axis) giving rise to the estimated population response (blue curve) that best fits the recorded population neural response (red curve). Each blue dot represents the estimated response from one neuron and the blue curve represents the spline-fitted estimated population response. Two red dots on the X-axis indicate the stimulus speeds. The Y-axis on the right side shows the weight of the readout speed.

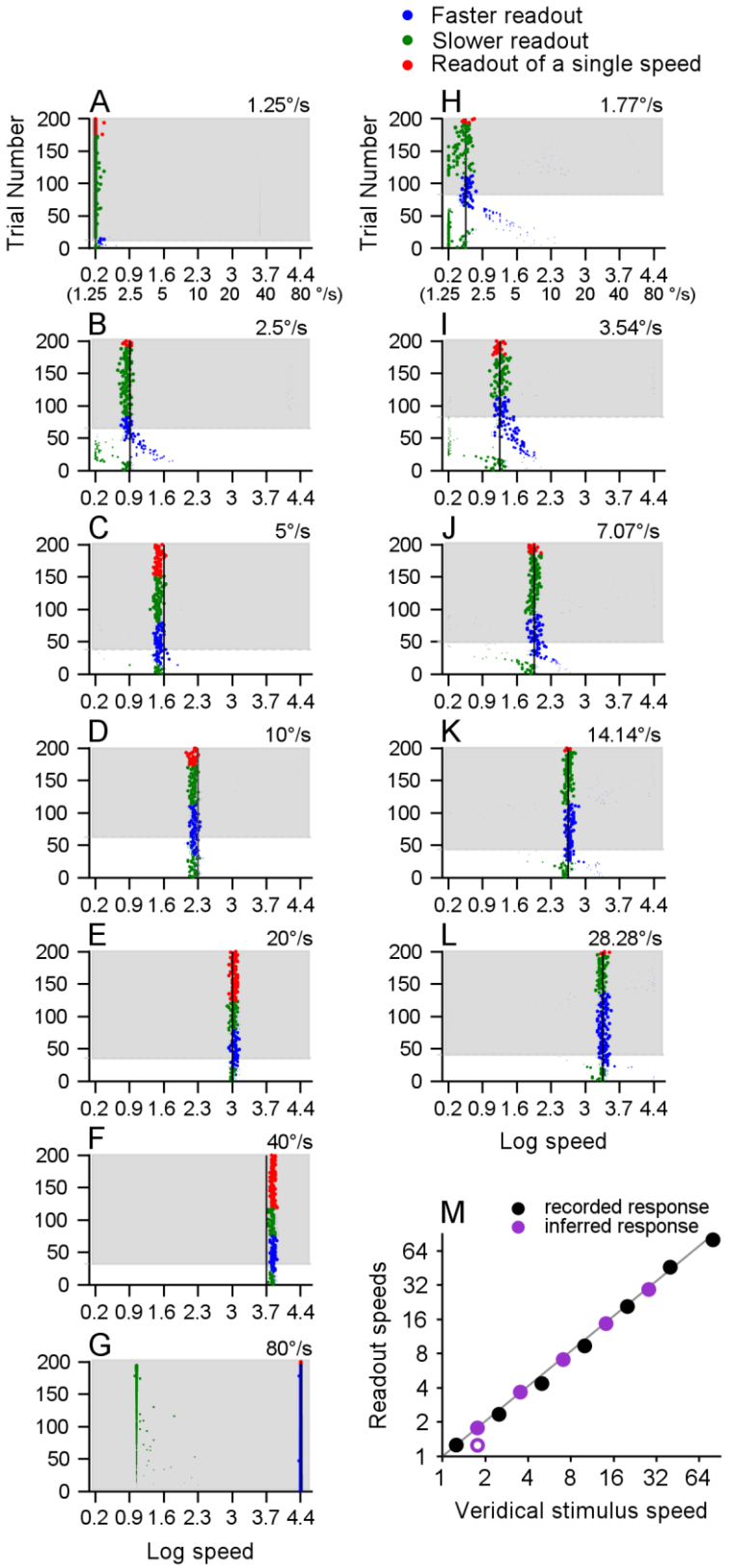


Figure 11. Trial-by-trial readout speeds decoded from population neural responses to single speeds. The neural population contained 100 neurons as in Figure 9. The trial-by-trial responses were randomly generated based on a Poisson process with the mean set to the spike count averaged across the experimentally recorded trials. Each row shows the readout speed(s) from one trial and the size of each dot is proportional to the weight of the readout speed. If only one speed is decoded in a trial, that readout speed is shown in red. In trials that have two readout speeds, the slower and faster readout speeds are shown in green and blue, respectively. The white background indicates trials that have a weight difference between two readout speeds less than 0.7 and are considered to have two readout speeds. The gray background indicates trials that have a weight difference greater than 0.7 and are considered to have only one readout speed. The vertical black line and the speed marked in each panel indicate the stimulus speed. **A-G.** Speeds decoded from recorded population neural responses to single speeds from 1.25 to 80º/s. **H-L.** Speeds decoded from inferred population neural response to single speeds, which are the log-mean speed of the bi-speed stimuli that have x2 speed separation. **M.** Comparison of the readout speeds and the stimulus speeds. The diagonal line is the unity line. The ordinate represents the speed at the peak of the readout speed distribution. At the stimulus speed of 1.77º/s (H), the distribution of the readout speed has two peaks, indicated by a solid circle (at 1.77º/s) and an open circle (at 1.25º/s).

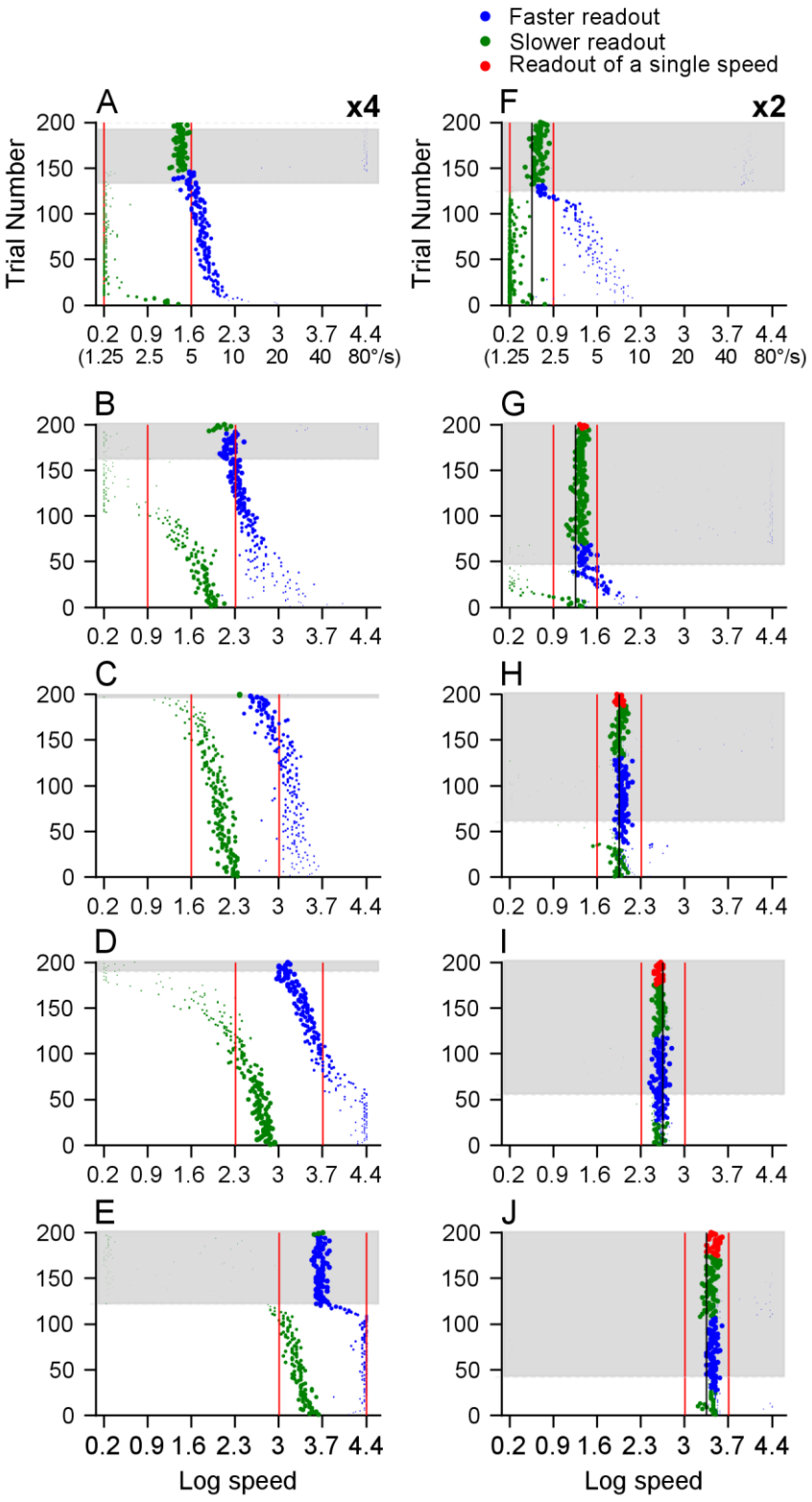


Figure 12. Trial-by-trial readout speeds decoded from population neural responses to the bi-speed stimuli. The neural population contains 100 neurons and the trial-by-trial responses are randomly generated based on a Poisson process. The convention is the same as in Figure 11. **A-E. Speeds decoded from population responses to x4 speed separation.** The vertical red lines indicate two component speeds, which are 1.25 and 5º/s (A), 2.5 and 10º/s (B), 5 and 20º/s (C), 10 and 40º/s (D), 20 and 80º/s (E). **F-J. Speeds decoded from population responses to x2 speed separation.** The red vertical line indicates two component speeds and the black vertical line indicates the log mean speed. The component speeds are 1.25 and 2.5º/s (F), 2.5 and 5º/s (G), 5 and 10º/s (H), 10 and 20º/s (I), 20 and 40º/s (J).

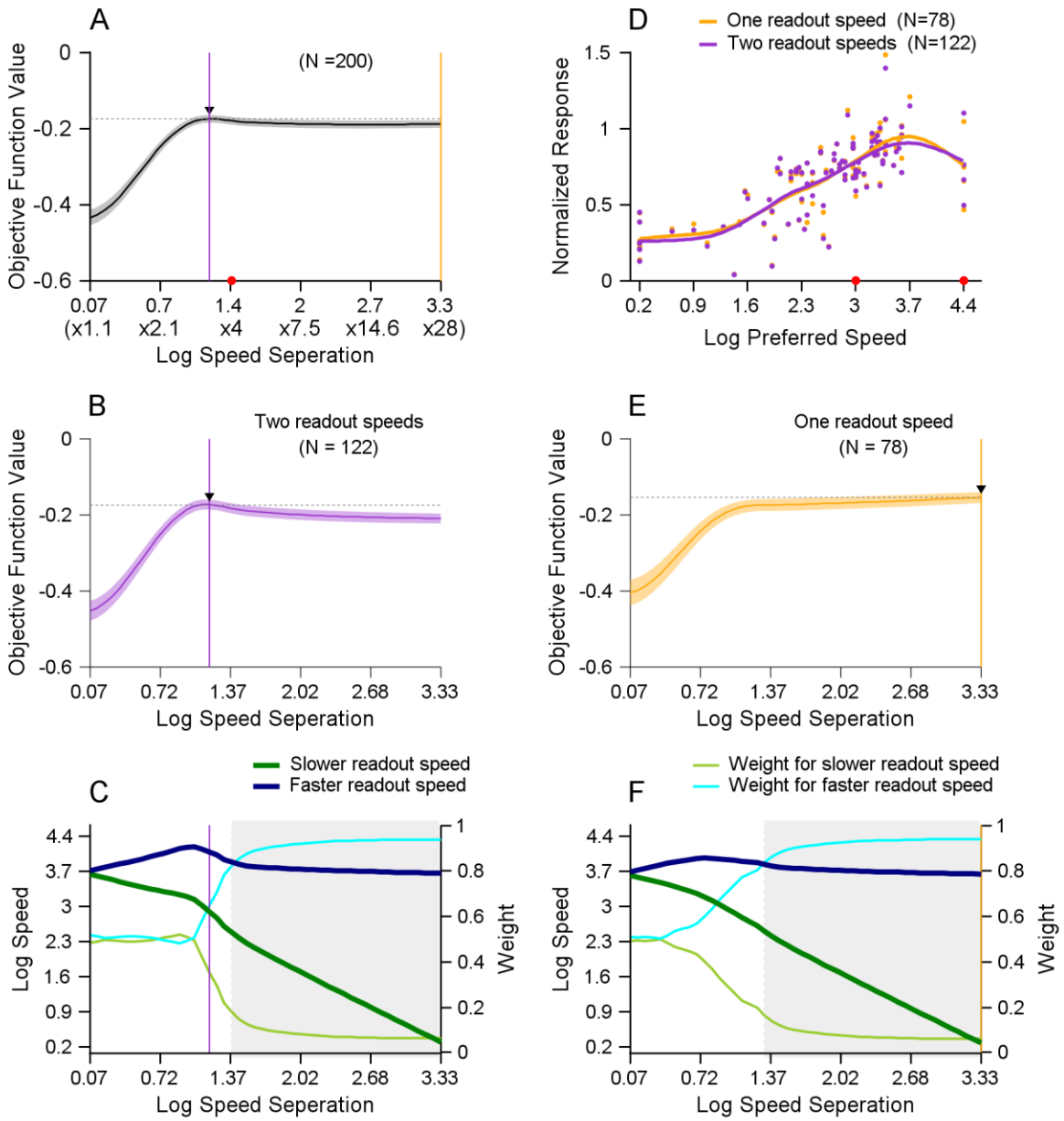


Figure 13. Analysis of decoding the speeds of the bi-speed stimulus at the fastest speeds of 20 and 80º/s. A. Evolution of the objective function averaged across all 200 trials as the decoder searched through different speed separations. The red dot on the X-axis indicates the speed separation of the stimulus speeds. **B, E.** Evolution of the objective functions averaged across trials considered to have two (B) and one (E) readout speed(s). In A, B, and E, the error bands indicate ±STE. The black arrow indicates the speed separation where the objective function reaches its peak. The horizontal dotted line indicates the peak value of the objective function. **C, F.** Evolution of the readout speeds (darker and thick lines) and their weights (lighter and thin lines) as the decoder searched through different speed separations in trials considered to have two (C) and one (F) readout speed(s). **D.** Population neural responses averaged across trials that are considered to have two readout speeds (purple) and one readout speed (orange). Each dot represents the trial-averaged response of one neuron. The curves represent the spline-fitted population neural responses. The two red dots on the X-axis indicate stimulus speeds of 20 and 80º/s.

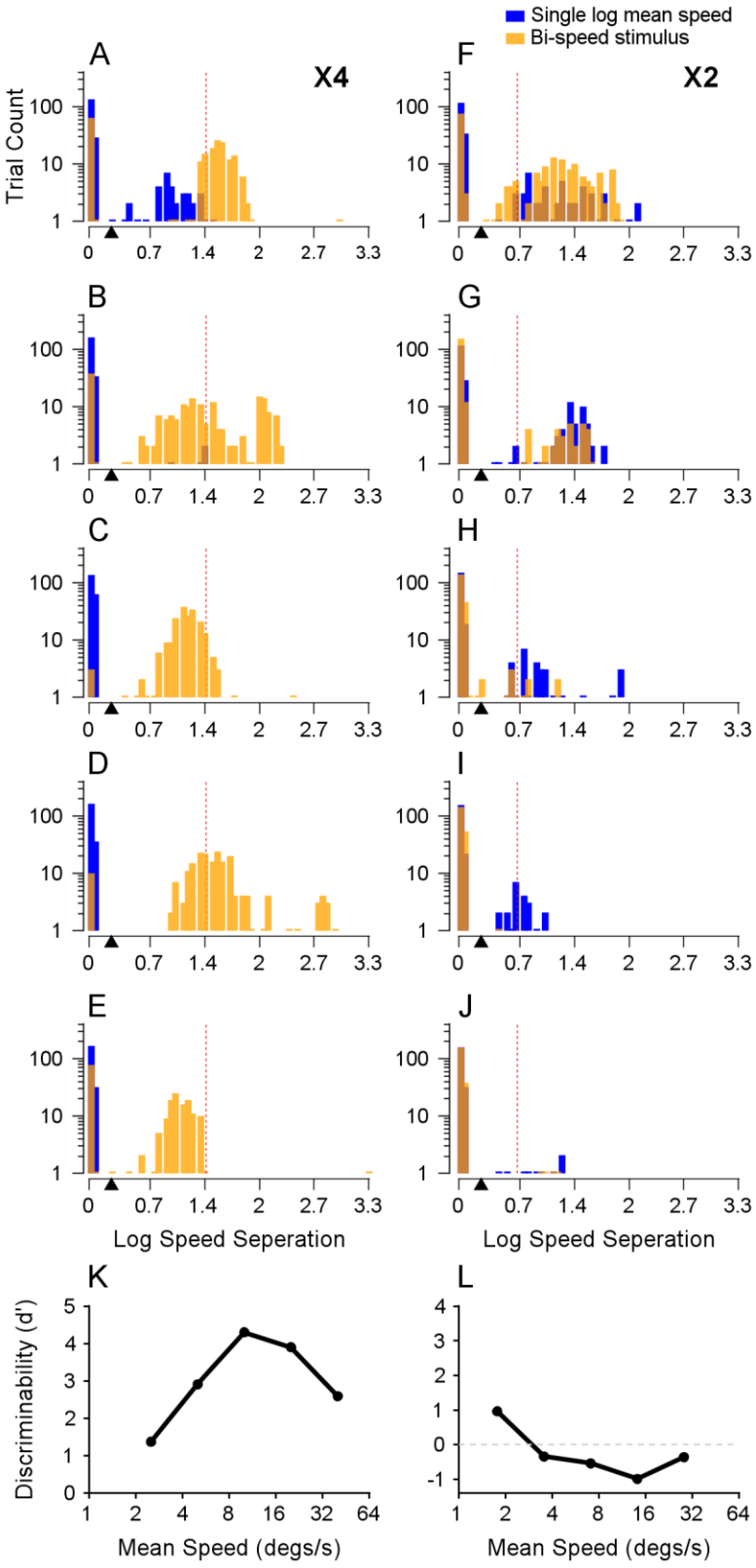


Figure 14. Discrimination between single- and bi-speed stimuli based on decoded speeds. A-J. The distributions of the speed separation between two readout speeds in each trial for the bi-speed stimuli (yellow) and the single, log-mean speed (blue). The bin width is 0.05. The abscissa is shown in natural logarithm scale. The red dotted line indicates veridical speed separation. **A-E. X4 speed separation.** The speeds of the bi-speed stimuli are 1.25 and 5º/s (A), 2.5 and 10º/s (B), 5 and 20º/s (C), 10 and 40º/s (D), 20 and 80º/s (E). **F-J. X2 speed separation.** The speeds of the bi-speed stimuli are 1.25 and 2.5º/s (F), 2.5 and 5º/s (G), 5 and 10º/s (H), 10 and 20º/s (I), 20 and 40º/s (J). **K-L.** Performance of discriminating a bi-speed stimulus from the corresponding log-mean speed based on the speed separation of the decoded speeds. **K.** X4 speed separation; **L.** X2 speed separation. The black triangles in A-J indicate the speed separation threshold of x1.3 (0.26 on the log scale) used for classifying bi-speed and single-speed stimuli.

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. (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., & Ringach, D. L. (1997). Predictions of a recurrent model of orientation selectivity. *Vision Res*, *37*(21), 3061-3071.

Caruso, V. C., Mohl, J. T., Glynn, C., Lee, J., Willett, S. M., Zaman, A., Ebihara, A. F., Estrada, R., Freiwald, W. A., Tokdar, S. T., & Groh, J. M. (2018). Single neurons may encode simultaneous stimuli by switching between activity patterns. *Nat Commun*, *9*(1), 2715.

Churchland, M. M., & Lisberger, S. G. (2001). Shifts in the population response in the middle temporal visual area parallel perceptual and motor illusions produced by apparent motion. *J Neurosci*, *21*(23), 9387-9402.

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., & 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.

Huang, X., MacEvoy, S. P., & Paradiso, M. A. (2002). Perception of brightness and brightness illusions in the macaque monkey. *J Neurosci*, *22*(21), 9618-9625.

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.

Krekelberg, B., van Wezel, R. J., & Albright, T. D. (2006b). Interactions between speed and contrast tuning in the middle temporal area: implications for the neural code for speed. *J Neurosci*, *26*(35), 8988-8998.

Li, K., Kozyrev, V., Kyllingsbæk, S., Treue, S., Ditlevsen, S., Bundesen, C. (2016). Neurons in Primate Visual Cortex Alternate between Responses to Multiple Stimuli in Their Receptive Field. *Frontiers in Computational Neuroscience* 10: 141.

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.

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.

Morgan, M. L., Deangelis, G. C., & Angelaki, D. E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, *59*(4), 662-673.

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.

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.

Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annu Rev Neurosci*, *26*, 381-410.

Pouget, A., & Snyder, L. H. (2000). Computational approaches to sensorimotor transformations. *Nat Neurosci*, *3 Suppl*, 1192-1198.

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. (2004). Estimating target speed from the population response in visual area MT. *J Neurosci*, *24*(8), 1907-1916.

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.

Vintch, B., & Gardner, J. L. (2014). Cortical correlates of human motion perception biases. *J Neurosci*, *34*(7), 2592-2604.

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.

Yang, J., & Lisberger, S. G. (2009). Relationship between adapted neural population responses in MT and motion adaptation in speed and direction of smooth-pursuit eye movements. *J Neurophysiol*, *101*(5), 2693-2707.

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural*

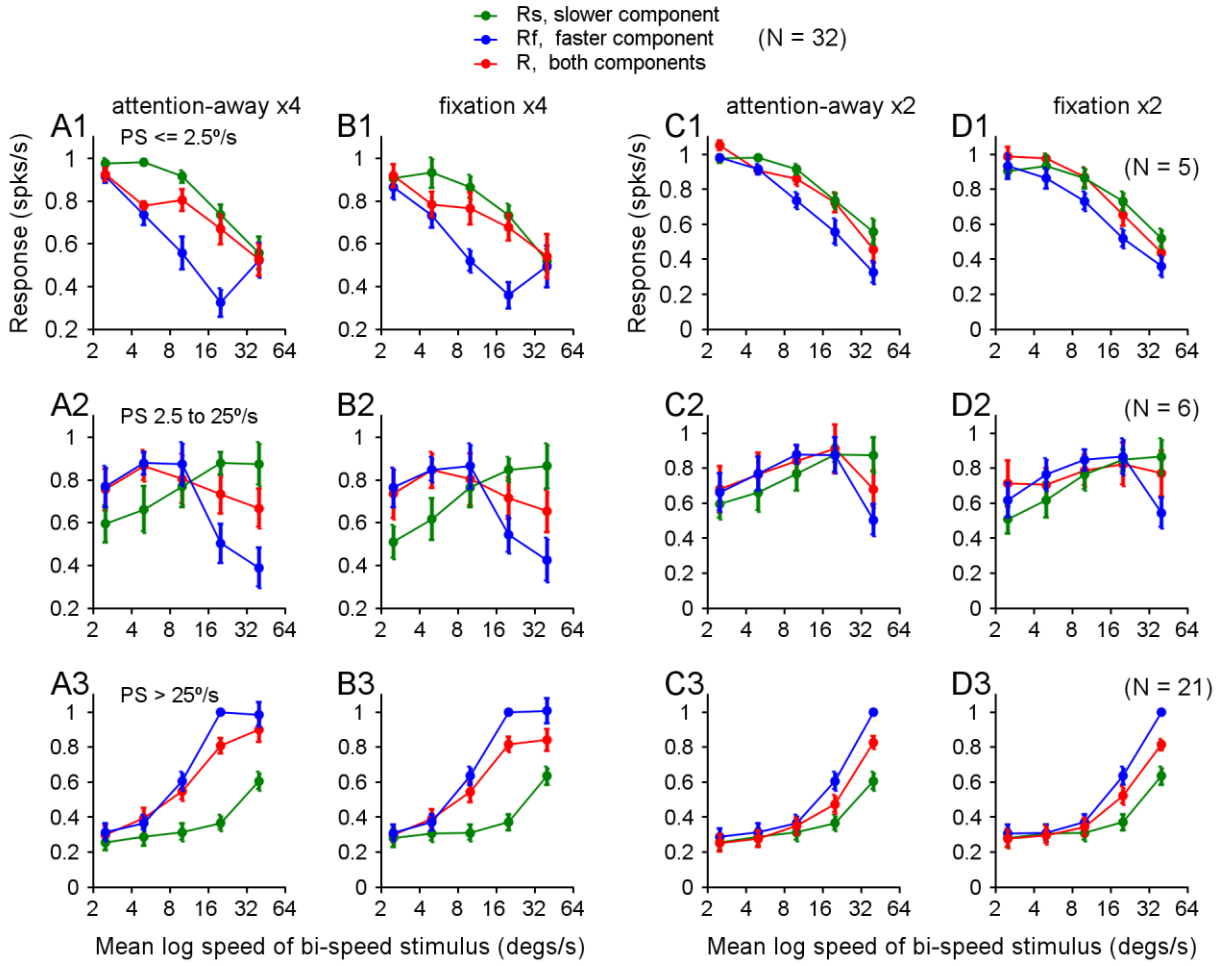
Comput, *10*(2), 403-430.

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.

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*(6485), 140-143.

Neural coding 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 fixation paradigm. Speed tuning curves from one monkey (RG) averaged across **A1-D1.** 5 neurons that had PS lower than 2.5º/s. **A2-D2.** 6 neurons that had PS between 2.5 and 25º/s. **A3-D3.** 21 neurons that had PS greater than 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.