

# Dissociation of Self-Motion and Object Motion by Linear Population Decoding That Approximates Marginalization

Ryo Sasaki,<sup>1</sup> Dora E. Angelaki,<sup>2,3</sup> and Gregory C. DeAngelis<sup>1</sup>

<sup>1</sup>Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York, 14627, <sup>2</sup>Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, and <sup>3</sup>Department of Electrical and Computer Engineering, Rice University, Houston, Texas 77005

We use visual image motion to judge the movement of objects, as well as our own movements through the environment. Generally, image motion components caused by object motion and self-motion are confounded in the retinal image. Thus, to estimate heading, the brain would ideally marginalize out the effects of object motion (or vice versa), but little is known about how this is accomplished neurally. Behavioral studies suggest that vestibular signals play a role in dissociating object motion and self-motion, and recent computational work suggests that a linear decoder can approximate marginalization by taking advantage of diverse multisensory representations. By measuring responses of MSTd neurons in two male rhesus monkeys and by applying a recently-developed method to approximate marginalization by linear population decoding, we tested the hypothesis that vestibular signals help to dissociate self-motion and object motion. We show that vestibular signals stabilize tuning for heading in neurons with congruent visual and vestibular heading preferences, whereas they stabilize tuning for object motion in neurons with discrepant preferences. Thus, vestibular signals enhance the separability of joint tuning for object motion and self-motion. We further show that a linear decoder, designed to approximate marginalization, allows the population to represent either self-motion or object motion with good accuracy. Decoder weights are broadly consistent with a readout strategy, suggested by recent computational work, in which responses are decoded according to the vestibular preferences of multisensory neurons. These results demonstrate, at both single neuron and population levels, that vestibular signals help to dissociate self-motion and object motion.

**Key words:** heading; marginalization; multisensory; object motion; optic flow

## Significance Statement

The brain often needs to estimate one property of a changing environment while ignoring others. This can be difficult because multiple properties of the environment may be confounded in sensory signals. The brain can solve this problem by marginalizing over irrelevant properties to estimate the property-of-interest. We explore this problem in the context of self-motion and object motion, which are inherently confounded in the retinal image. We examine how diversity in a population of multisensory neurons may be exploited to decode self-motion and object motion from the population activity of neurons in macaque area MSTd.

## Introduction

Under natural conditions in which many things may change simultaneously in the environment, incoming sensory signals re-

flect a combination of multiple environmental causes. However, for specific behavioral tasks, it may be necessary to make judgments about a single environmental cause of sensory input without being influenced by other competing causes (i.e., “nuisance variables”). In general, estimating one variable and ignoring others can be achieved by marginalization, a computation in which some variables are integrated out of a joint probability distribution.

Although marginalization is probably important for many cognitive processes, little is known about its underlying neural mechanisms (Beck et al., 2011). We recently proposed a method for determining a linear decoder that approximates marginaliza-

Received April 30, 2017; revised Oct. 2, 2017; accepted Oct. 6, 2017.

Author contributions: R.S., D.E.A., and G.C.D. designed research; R.S. performed research; R.S. analyzed data; R.S., D.E.A., and G.C.D. wrote the paper.

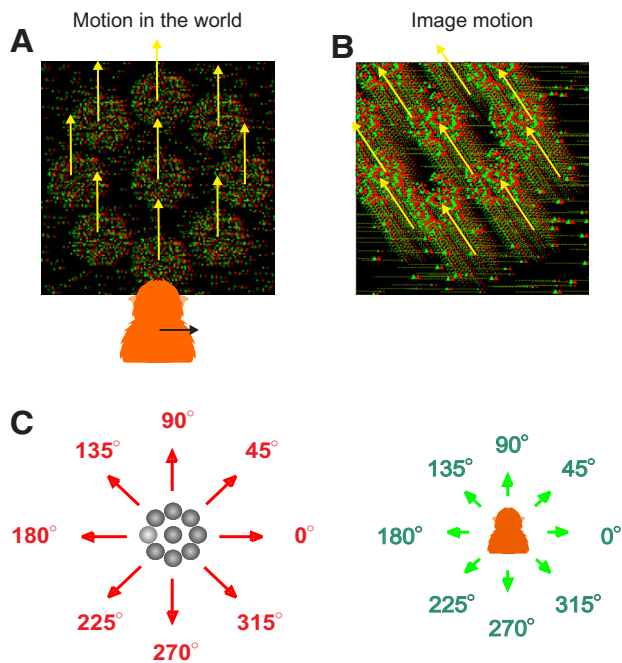
This work was supported by NIH Grants EY016178 (G.C.D.) and DC014678 (D.E.A.), The Uehara Memorial Foundation (R.S.), the Japan Society for the Promotion of Science (R.S.), and an NEI CORE Grant (EY001319). We thank Dina Knoedl, Swati Shimpi, and Emily Murphy for excellent technical support, and Johnny Wen for programming support.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Gregory C. DeAngelis, Department of Brain and Cognitive Sciences, Center for Visual Science, 310 Meliora Hall, University of Rochester, Rochester, NY 14627-0268. E-mail: gdeangelis@cvs.rochester.edu.

DOI:10.1523/JNEUROSCI.1177-17.2017

Copyright © 2017 the authors 0270-6474/17/3711204-16\$15.00/0



**Figure 1.** Schematic illustration of stimuli and experimental design. **A**, The monkey viewed a scene in which a multipart object moved in the world while the monkey was also translated. In this illustration, self-motion is rightward (0°) and the object moves upward (90°) in the world. The object and background were composed of random dots and rendered in 3D using disparity cues. **B**, Image motion resulting from the combination of self-motion and object motion depicted in **A**. Note that this is a planar image projection, and has not been flipped and inverted as it would be by the lens of the eye. **C**, Both the multipart object and the monkey moved in one of eight possible directions in the frontoparallel plane, as shown by red and green vectors, respectively. The speed of object motion was 50% faster than that of self-motion.

tion, and we showed through simulations that diverse multisensory representations could facilitate performance of the method (Kim et al., 2016). Here, we evaluate this method in the context of a naturalistic application: dissociation of self-motion and object motion.

For a moving observer, retinal image motion is a combination of components resulting from self-motion and object motion (Fig. 1A,B). Therefore, to accurately perceive self-motion, the brain needs to marginalize away the effects of object motion, and vice versa. Although some behavioral studies suggest that purely visual mechanisms exist to parse image motion into components related to self-motion and object motion (Rushton and Warren, 2005; Warren and Rushton, 2007, 2008; Matsumiya and Ando, 2009), other studies suggest that nonvisual (e.g., vestibular) signals make important contributions (Wexler, 2003; Wexler and van Boxtel, 2005; MacNeilage et al., 2012; Dupin and Wexler, 2013; Fajen and Matthis, 2013; Dokka et al., 2015a,b).

Little is known about neural computations that can dissociate self-motion and object motion. A few previous studies have examined interactions between visual motion patterns that simulate self-motion and object motion in area MSTd (Logan and Duffy, 2006; Sato et al., 2010; Kishore et al., 2012). These studies show that MSTd responses depend on self-motion and object motion in complex ways, but they did not systematically explore the joint tuning of MSTd neurons for heading and object direction, nor how nonvisual inputs influence the joint representation of self-motion and object motion. Given that vestibular signals contribute to perceptual judgements of self-motion and object motion (Fajen and Matthis, 2013; Dokka et al., 2015a,b), we hypothesized that the capacity of cortical neurons to distinguish

self-motion and object motion would be enhanced by vestibular cues.

Cortical neurons in multiple brain areas, including MSTd (Gu et al., 2006, 2008; Fetsch et al., 2011), ventral intraparietal (VIP; Bremmer et al., 2002; Chen et al., 2011b, 2013), visual posterior sylvian (VPS; Chen et al., 2011b), and frontal eye field (FEF; Gu et al., 2016), are selective for the direction of translation (heading) based on both optic flow and vestibular inputs. These areas contain “congruent” cells, which have matched vestibular and visual heading preferences, as well as “opposite” cells, which tend to prefer opposite directions of real and visually-simulated translation (Gu et al., 2006; Chen et al., 2011b). Whereas the activity of congruent cells is compatible with roles in cue integration (Gu et al., 2008; Chen et al., 2013) and reliability-based cue weighting (Fetsch et al., 2011), opposite cells are poorly suited to these computations. However, opposite cells may respond strongly when a moving object produces image motion that is inconsistent with self-translation. Thus, the combined activity of a mixture of congruent and opposite cells could allow the system to identify image motion that is inconsistent with self-motion. Indeed, simulations suggest that a properly trained linear decoder (Kim et al., 2016) can take advantage of this diversity to marginalize over object motion and estimate heading, but this approach has not been applied to real neural data.

We tested macaque MSTd neurons with many combinations of object motion and self-motion (including visual and vestibular cues to self-motion). We examined how vestibular signals modulate the joint tuning of single neurons for heading and object direction. In addition, we evaluated whether our method for approximate linear marginalization (ALM; Kim et al., 2016) could compute the probability distribution over heading while marginalizing out the effects of object motion, or vice versa. Our findings suggest that vestibular signals facilitate a robust neural representation of motion that can be used to estimate either heading or object motion.

## Materials and Methods

### Subjects and preparation

Two male rhesus monkeys (*Macaca mulatta*), weighing 12.8 and 11.5 kg, participated in this study. Both animals were 6–8 years of age during the course of the studies, and were pair-housed in a vivarium with a 12 h light cycle (6:00 A.M. to 6:00 P.M.). Under sterile conditions, monkeys were chronically implanted with a circular Delrin ring (diameter: 7 cm) for head stabilization, as described previously (Gu et al., 2006), as well as a scleral search coil for measuring eye movements. After recovery, animals were trained to fixate visual targets for fluid rewards using standard operant conditioning techniques.

For electrophysiological recording, a Delrin grid (2.5 × 4.5 × 0.5 cm) containing rows of holes was stereotactically secured to the skull inside the head-restraint ring and was positioned in the horizontal plane. The holes in the grid (0.8 mm spacing) allowed vertical penetration of microelectrodes into the brain via transdural guide tubes that were inserted through a small burr hole in the skull. Burr holes were made surgically under aseptic conditions while the subjects were anesthetized. The recording grid extended bilaterally from the midline to regions overlying area MST in both hemispheres. All experimental procedures conformed to National Institutes of Health guidelines and were approved by the University Committee on Animal Resources at the University of Rochester.

### Vestibular and visual stimuli

During experiments, monkeys were seated comfortably in a primate chair with their head restrained. The chair was securely attached to a 6 degree-of-freedom motion platform (MOOG 6DOF2000E, Moog) that was used to passively translate the animals. A field-coil system (CNC Electronics) was mounted on top of the motion platform, surrounding

the monkey's head and body, and was used to measure eye movements using the scleral coil technique. Visual stimuli were rear-projected onto a tangent screen by a three-chip digital light projector (Mirage S<sup>+</sup>3K, Christie Digital Systems), and the tangent screen was attached to the front of the field coil (Gu et al., 2006). The display screen measured 60 × 60 cm and was mounted at a viewing distance of ~30 cm in front of the monkey, thus subtending ~90 × 90° of visual angle. The sides and top of the field coil frame were covered with a matte black enclosure such that the animal could only see visual stimuli that were presented on the tangent screen; the room was not visible. Platform motion and visual stimuli were updated at 60 Hz and were precisely calibrated such that visual motion was synchronous with platform motion (Gu et al., 2006).

Visual stimuli simulated translational self-motion along one of eight directions, spaced 45° apart, in the frontoparallel (i.e., vertical) plane (Fig. 1C). We elected to restrict self-motion to the frontoparallel plane for two main reasons: (1) to reduce the dimensionality of the stimulus space, which was necessary because we test all combinations of self-motion and object motion directions; and (2) because it simplifies the types of visual motion interactions that take place within the receptive field (i.e., no radial motion), thus simplifying predictions for how vestibular signals may interact with visual motion.

The visual scene consisted of a three-dimensional (3D) field of stars. Each star was a triangle that measured 0.15 × 0.15 cm, and the cloud measured 100 cm wide by 100 cm tall by 40 cm deep with a star density of 0.01/cm<sup>3</sup>. Along the depth dimension, the volume of the star field was centered on the distance of the display screen and fixation target. Given that the viewing distance was ~30 cm, the star field ranged in depth from 10 to 50 cm in front of the monkey. To provide stereoscopic cues, the dot cloud was rendered as a red-green anaglyph and viewed through custom red-green goggles (Kodak Wratten2 filters #29 and #61). The optic flow field contained naturalistic cues mimicking translation of the observer through the 3D star field, including motion parallax, size variations, and binocular disparity. Visual stimuli were generated using the OpenGL libraries (under Microsoft Visual C++) and were rendered using an accelerated graphics card (Quadro FX4800), with anti-aliasing for smooth motion. Optic flow simulating self-motion was generated by translating the OpenGL cameras (1 for each eye) through the 3D star field. The star field always filled the entire visible video display; as dots disappeared from one edge of the display, other dots appeared on the opposite edge. This helped to enhance the perception that background dots were stationary in the world.

To explore the interaction between self-motion and object motion, a large, multipart object moved in the world while the monkey was translated. The object consisted of a cluster of spheres (diameter 15 cm each) that were composed of random dots, with a dot density (0.25/cm<sup>3</sup>) that is greater than that of the background dots, such that the object is easily detectable in the scene. Each sphere was separated from its neighbors by 22.5 cm and they were transparent such that they did not occlude background optic flow (Fig. 1A). The multipart object was centered in depth at the same distance from the monkey (~30 cm) as the fixation point, which was in the plane of the display screen. Note that the entire outer boundary of the group of objects moved relative to the screen boundaries, facilitating the distinction between motion of the object and motion of the (stationary) background dots. After exploring various options including single spheres of different sizes, we elected to use the multipart object for the following reasons. First, because object location was not tailored to the receptive field of each neuron (to facilitate decoding analyses), we wanted object motion to cover a substantial portion of the visual field such that most receptive fields would be stimulated by the object. Second, using a single very large object may create figure-ground ambiguity that we wanted to avoid. Using a cluster of transparent moving spheres as the object allows decent coverage of the visual field while still allowing the moving objects to be easily segmented from the background.

As for self-motion, object motion was along one of eight directions (45° apart) within the frontoparallel plane. All self-motion and object motion trajectories were straight and had a Gaussian velocity profile with a SD of 1/3 s (Gu et al., 2006). The total excursion (0.15 m) and peak velocity (0.45 m/s) of object motion was somewhat greater than that for self-motion (0.10 m and 0.30 m/s, respectively), such that there was

always movement of the object relative to the background even when the background and the object moved in the same direction.

### Electrophysiology procedures

We recorded action potentials extracellularly from area MSTd of two monkeys. A tungsten microelectrode (Frederic Haer; tip diameter 3 μm, impedance 1–2 MΩ at 1 kHz) was advanced into the cortex through a transdural guide tube, using a hydraulic micromanipulator (Narishige) mounted on top of the head-restraint ring. Action potentials were amplified and isolated using a head-stage preamplifier, a bandpass eight-pole filter (Krohn-Hite, model 3384; 400–4000 Hz), and a dual voltage-time window discriminator (Bak Electronics, model RP-1). The times of occurrence of action potentials and all behavioral events were recorded with 1 ms resolution by the data acquisition computer. Raw neural signals were also digitized at 25 kHz and stored to disk for off-line spike sorting and additional analyses. Experimental control and data acquisition were coordinated by scripts written with TEMPO software (Reflective Computing).

Area MSTd was localized with aid from structural MRI scans and a standard macaque atlas (Van Essen et al., 2001). MSTd was typically identified as a region centered ~15 mm lateral to the midline and ~3–6 mm posterior to the interaural plane. Electrode penetrations were also guided by the pattern of background activity as the electrode traversed through gray and white matter, as well as the response properties of neurons to visual stimuli. MSTd was usually the first gray matter encountered, ~6–10 mm below the cortical surface, which exhibited prominent response modulation to flashing random-dot stimuli and direction-selective responses to motion of the dots. We were careful to distinguish MSTd from the lateral subdivision of area MST (MSTl). To do this, we also carefully mapped the portions of area MT that were found beneath MSTd, in the posterior bank of the superior temporal sulcus. We mainly targeted regions of area MSTd that were located postero-medially, overlying portions of area MT that had receptive fields with moderate to large eccentricities.

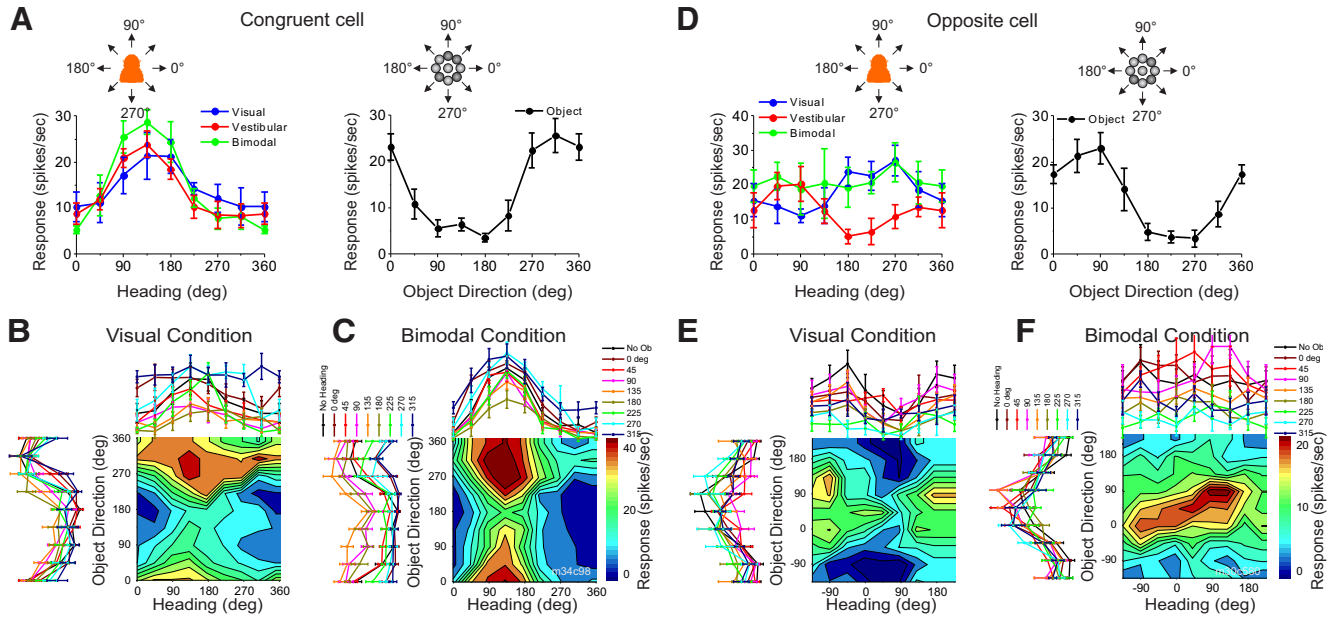
### Experimental protocol

We searched for neurons while presenting a large pattern of flickering or drifting random dots. After isolating the action potential of an MSTd neuron, we used custom software to perform an initial characterization of response properties, making use of a graphical interface that controlled the position, size, and velocity of a patch of dots. After hand-mapping the receptive field, a reverse-correlation procedure was used to obtain a quantitative map of the receptive field, as described in detail previously (Chen et al., 2008). We also measured heading tuning within the frontoparallel plane by presenting eight directions of motion, 45° apart, defined by optic flow alone (visual condition), platform motion alone (vestibular condition), and congruent visual and vestibular inputs (bimodal condition).

After this initial characterization of response properties, the main experimental protocol was run. This protocol included a fully crossed design in which eight directions of self-motion in the frontoparallel plane were crossed with eight directions of object motion. Two full sets of these 8 × 8 stimulus conditions were presented, a visual condition in which self-motion was visually simulated by optic flow while the animal remained stationary, and a bimodal condition in which self-motion was provided by congruent optic flow and vestibular signals. In addition, the main protocol included three self-motion control conditions in which self-motion was presented without any moving object: a visual, a vestibular, and a bimodal condition. The protocol also included an object-motion control condition in which only the moving object was presented, with no self-motion, and a null condition in which no stimuli were presented except for the fixation target. In total, 161 unique stimulus conditions were randomly interleaved, requiring 805 trials to complete 5 repetitions of each distinct stimulus.

### Data analysis

The response of a neuron to each stimulus was computed as the firing rate over a time window from 500 to 2000 ms following stimulus onset, as this time window was previously found to contain most of the neural response to these stimuli (Gu et al., 2006, their Fig. 1C). The joint tuning



**Figure 2.** Interactions between self-motion and object motion for two example MSTd neurons. **A, D**, Left, Heading tuning curves in the absence of object motion (visual: blue; vestibular: red; bimodal: green). Right, Object motion tuning curve in the absence of self-motion (black). Data are shown for a congruent cell (**A**) and an opposite cell (**D**). **B, E**, Joint heading-object tuning profiles are shown for the visual condition in which self-motion is simulated by optic flow. Responses to combinations of heading and object motion are shown as a color contour map, and tuning curves along the margins show heading tuning for each object motion direction (top) and object tuning for each heading (left). **C, F**, Joint heading-object tuning profiles for the bimodal condition in which self-motion is indicated by both optic flow and vestibular signals.

of MSTd neurons for the 64 combinations of heading and object direction was plotted as a color map for visualization (Fig. 2C). In addition, data from the control conditions involving only self-motion or only object motion were plotted as tuning curves (Fig. 2A).

**Direction discrimination indices.** To quantify the strength of directional tuning exhibited by a neuron, we computed a well established metric of neural selectivity (Prince et al., 2002; DeAngelis and Uka, 2003) called the direction discrimination index (DDI):

$$DDI = \frac{R_{\max} - R_{\min}}{R_{\max} - R_{\min} + 2\sqrt{SSE/(N - M)}} \quad (1)$$

In this formulation,  $R_{\max}$  and  $R_{\min}$  denote the mean responses to the most effective and least effective stimulus directions, respectively, SSE is the sum squared error around the mean responses,  $N$  is the number of observations (trials), and  $M$  is the number of stimulus values tested.

In our application of DDI, we wanted to quantify the strength of neural selectivity for heading (while pooling across object directions) or the strength of selectivity for object direction (pooling across headings). Thus, we computed two DDI metrics.  $DDI_{\text{heading}}$  was computed after responses were pooled across the eight possible object motion directions, and  $DDI_{\text{object}}$  was computed after responses were pooled across the eight possible headings (see Fig. 4). These pooled DDI metrics therefore reflect the consistency of tuning for one variable (e.g., heading) across variations in the other variable (e.g., object motion). A neuron could have a low DDI value because it is not tuned at all or because its tuning for one variable is not consistent across variations in the other variable. These pooled DDI metrics are useful to evaluate whether the addition of vestibular signals makes tuning more consistent or less consistent.

To further investigate whether the interactions between self-motion and object motion in MSTd responses can be described by baseline response shifts, peak shifts, and/or gain changes, we calculated DDI metrics that are compensated for baseline shifts (mean-compensated DDI), peak shifts (shift-compensated DDI), and gain modulations (gain-compensated DDI). To compute the mean-compensated  $DDI_{\text{heading}}$ , mean responses were equated across object directions before computing  $DDI_{\text{heading}}$ . To compute the shift-compensated  $DDI_{\text{heading}}$ , the peaks of heading tuning curves were aligned across object directions before computing  $DDI_{\text{heading}}$ . To compute the gain-compensated  $DDI_{\text{heading}}$ , the

peak-trough amplitudes of heading tuning curves were equated across object directions before computing  $DDI_{\text{heading}}$ . Analogous operations were performed to compute the compensated values of  $DDI_{\text{object}}$ , but instead equating mean values, peak locations, or peak-trough amplitudes across headings.

**Direction separability index.** If heading tuning is independent of object direction (and vice versa), then we may expect the joint tuning profile for heading and object motion to be multiplicatively separable. In contrast, if heading tuning depends on object direction (or vice versa), then the joint heading/object tuning profile may have a somewhat slanted structure (see Fig. 7A). To quantify the separability of joint heading/object tuning, singular value decomposition (SVD) was applied to the joint tuning profile of each neuron. This approach represents the empirically measured joint tuning profile as a weighted sum of multiplicatively-separable components. Thus, if the joint tuning is multiplicatively separable, then the first singular value will be large and all of the other singular values will be relatively small. To quantify separability, we compute a direction separability index (DSI), which depends on the relative magnitude of the first singular value to the sum of all singular values (Depireux et al., 2001; Mazer et al., 2002):

$$DSI = 1 - \frac{\lambda(1)^2}{\sum_i \lambda(i)^2} \quad (2)$$

where  $\lambda(i)$  is the  $i$ th singular value. If the joint tuning is separable, then DSI would be close to zero, and greater values indicate increasing inseparability.

**Population decoding by likelihood computation.** Maximum likelihood decoding can be used to estimate heading or object direction from a sample of activity of our population of MSTd neurons. To compute an estimate of the log likelihood over heading  $[\log L(\theta)]$  or object direction  $[\log L(\phi)]$ , each neuron's response ( $r_i$ ) to a particular stimulus was multiplied by the log of its tuning function  $[\log f_i(\theta) \text{ or } \log f_i(\phi)]$  and the result was summed across neurons (Eqs. 3, 4; Dayan and Abbott, 2001; Jazayeri and Movshon, 2006). The second term in these equations compensates for biases caused by a nonuniform distribution of direction preferences.

$$\log L(\theta) = \sum_{i=1}^N r_i \log f_i(\theta) - \sum_{i=1}^N f_i(\theta), \quad (3)$$

$$\log L(\varphi) = \sum_{i=1}^N r_i \log f_i(\varphi) - \sum_{i=1}^N f_i(\varphi). \quad (4)$$

In these formulations,  $r_i$  represents the response of the  $i$ th neuron to a particular combination of heading and object motion in either the visual or bimodal condition. For heading estimation,  $f_i(\theta)$  could be either the visual or vestibular heading tuning function for each neuron, without object motion. For object motion estimation,  $f_i(\varphi)$  is the object motion tuning curve measured without self-motion. The peak of the resulting distribution represents the maximum-likelihood estimate of either heading or object motion direction. This approach assumes that neurons have independent Poisson spiking statistics.

It should be emphasized that the distributions obtained from Equations 3 and 4 will only equal the corresponding log likelihoods over heading and object direction if the shape of the population activity is not influenced by nuisance variables such as object motion and self-motion, respectively. If these conditions are not met (which will generally be the case in our applications, since tuning for heading is altered by object motion and vice versa), then Equations 3 and 4 do not compute true log-likelihoods but may instead be referred to as recognition models (Hinton and Dayan, 1996). Although we do not expect this simple form of population decoding to effectively perform marginalization, it provides an instructive comparison to an optimal linear decoder that does approximate marginalization, as described below.

In addition, we also computed the joint likelihood over both heading and object direction by applying the same approach to the joint tuning profiles for heading and object motion:

$$\log L(\theta, \varphi) = \sum_{i=1}^N r_i \log f_i(\theta, \varphi) - \sum_{i=1}^N f_i(\theta, \varphi). \quad (5)$$

Here,  $f_i(\theta, \varphi)$  represents the joint tuning profile for heading and object direction (Fig. 2C) for the  $i$ th neuron.

**Approximate linear marginalization.** As described above, the responses of MSTd neurons will generally be a function of both heading and object direction,  $f(\theta, \varphi)$ , and the joint posterior over both variables,  $P(\theta, \varphi | \mathbf{r})$ , can be computed from the neural responses,  $\mathbf{r}$ , and the joint tuning profiles (assuming a flat prior; see Eq. 9). However, in many situations, it is not necessary to compute the joint posterior, and it may be advantageous to directly estimate the marginal posterior over heading or object direction,  $P(\theta | \mathbf{r})$  or  $P(\varphi | \mathbf{r})$ , respectively. Mathematically, this involves integrating one of these variables out of the joint posterior (see Eq. 10), a computation known as marginalization. But how can the marginal posterior be computed from neural activity? Is it possible to directly decode neural activity and obtain a reasonable approximation of the marginal posterior?

As described in detail recently (Kim et al., 2016), we sought to find a linear transformation of neural activity that can perform near-optimal marginalization; for example, to compute an approximate marginal posterior distribution over heading that discounts object motion. To perform what we call ALM, we assume that the marginal posterior distribution over heading is approximated by a member ( $Q$ ) of the exponential family with linear sufficient statistics (Ma et al., 2006):

$$Q(\theta | \mathbf{r}; \mathbf{h}, \mathbf{g}) = \frac{1}{Z} e^{\mathbf{h}(\theta) \cdot \mathbf{r} + \mathbf{g}(\theta)}, \quad (6)$$

where

$$Z(\mathbf{r}; \mathbf{h}, \mathbf{g}) = \sum_{\theta} e^{\mathbf{h}(\theta) \cdot \mathbf{r} + \mathbf{g}(\theta)}. \quad (7)$$

We then optimize the parameters,  $\mathbf{h}(\theta)$  and  $\mathbf{g}(\theta)$ , to best explain data drawn from the true marginal posterior distribution,  $P(\theta | \mathbf{r})$ . This optimization, performed using multinomial logistic regression (Bishop, 2006), maximizes the likelihood of the parameters and minimizes the Kullback–Leibler divergence between the true marginal distribution  $P$  and the approximation  $Q$ . Given  $K$  samples from the joint distribution

$(\theta_k, \varphi_k, \mathbf{r}_k) \sim P(\theta, \varphi, \mathbf{r})$  for  $k = 1 \dots K$ , where  $\theta$  denotes heading,  $\varphi$  denotes object motion direction, and  $\mathbf{r}$  denotes the neural population response, the log-likelihood of the model parameters (for heading estimation) is given by the following:

$$\begin{aligned} \log L(\mathbf{h}, \mathbf{g}) &= \sum_k \log Q(\theta_k | \mathbf{r}_k; \mathbf{h}, \mathbf{g}) \\ &= \sum_k [\mathbf{h}(\theta_k) \cdot \mathbf{r}_k + \mathbf{g}(\theta_k) - \log Z(\mathbf{r}_k, \mathbf{h}, \mathbf{g})], \end{aligned} \quad (8)$$

where  $\mathbf{h} = [h_i(\theta)]$  denotes a matrix of weights specifying how the response  $r_i$  of neuron  $i$  influences the log-probability of heading  $\theta$ ,  $\mathbf{g}(\theta)$  represents a set of bias parameters for each heading, and the constant  $Z$  ensures that  $Q(\theta_k | \mathbf{r}_k; \mathbf{h}, \mathbf{g})$  is properly normalized. These parameters were obtained by performing multinomial logistic regression using the `glmtrain()` function in the Netlab toolbox for MATLAB (Nabney, 2002; Bishop, 2006), specifically with the algorithmic variant based on an approximate Hessian. For our purposes,  $\mathbf{h}(\theta)$  represents the key quantity that is obtained by ALM: it contains the optimal decoding weights for each neuron that provide the best approximation to the marginal posterior over heading that can be obtained by a linear transformation of responses. The equations above describe computation of the ALM weights for the case of estimating heading, but the procedure is completely analogous for object direction (substituting  $\varphi$  for  $\theta$ ), yielding decoding weights given by  $\mathbf{h}(\varphi)$ .

ALM was trained on a stimulus set including 500 simulated trials for each distinct stimulus condition. For each simulated trial, a pseudo-population response was created by sampling a response from each neuron (with replacement). Once ALM was trained, its performance was tested on a different set of pseudo-population responses, again comprising 500 simulated trials for each distinct stimulus condition. The stimulus conditions in the training set included all combinations of eight self-motion directions and eight object motion directions, as well as conditions involving self-motion only or object motion only. Specifically, when ALM was trained to decode heading, 30% of trials in the training set were self-motion only conditions (10% vestibular, 10% visual, and 10% bimodal) and 70% of trials involved combinations of self-motion and object motion (evenly split between visual and bimodal self-motion cues). Similarly, when ALM was trained to decode object motion direction, 30% of trials in the training set were object-only conditions, and 70% of trials involved combinations of self-motion and object motion (exactly the same as for decoding heading). Conditions involving only self-motion or only object motion were included in the training sets such that the resulting decoder should be able to estimate heading or object motion across a range of stimulus conditions.

### Experimental design and statistical analysis

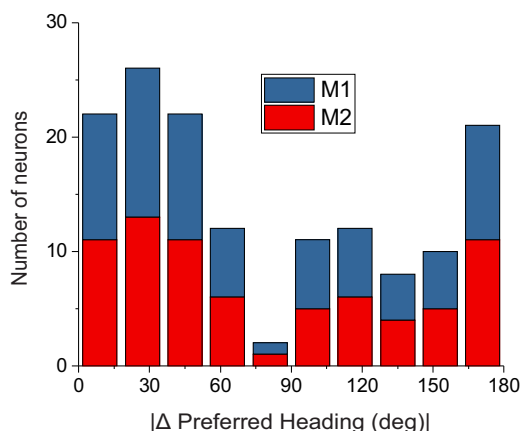
Data were collected from two male animals such that sex was not a variable in the analysis. For each animal, we collected data from between 70 and 100 neurons, which is commensurate with the standards of the field for single-unit studies in awake macaques.

The design of the experiments was such that each neuron was tested with an identical set of stimulus conditions. All stimulus conditions to be directly compared in the main experimental protocol were randomly interleaved in a block of trials for each neuron, such that there were no temporal sequence effects to consider.

Because all neurons were tested with the same stimulus set, statistical analyses focus on examining comparisons of quantitative metrics across the population. For this purpose, data from the two monkeys were pooled together, and there were no between-subject factors in the analysis. All statistical analyses were performed using the Statistics toolbox in MATLAB. Parametric or nonparametric statistics were used, as appropriate to each particular comparison, as described in detail in the Results section.

## Results

To investigate how visual motion signals related to self-motion and object motion interact, and how these interactions are modulated by vestibular self-motion signals, we recorded from 164



**Figure 3.** Summary of relative heading preferences for visual and vestibular inputs. The histogram shows the distribution of the absolute difference in preferred heading,  $|\Delta \text{Preferred Heading}|$ , between the visual and vestibular conditions. Data are shown for all 164 neurons, and heading preferences were estimated by computing the vector sum of responses to eight directions of motion in the frontoparallel plane. Colors denote data for the two animals.

MSTd neurons in two animals (73 from Monkey D and 91 from Monkey N). Monkeys were required to maintain fixation on a head-fixed target while a large multipart object moved in one of eight directions within the frontoparallel plane (Fig. 1). In addition, self-motion along eight possible directions within the frontoparallel plane was simulated by either full-field optic flow presented alone (visual condition) or by a combination of optic flow and whole-body translation (bimodal condition). All 64 combinations of 8 directions of self-motion and 8 directions of object motion ( $45^\circ$  apart) were presented (Fig. 1C), along with controls in which either self-motion or object motion was presented alone.

We attempted to record from any MSTd neuron that exhibited clear visual responses to moving and/or flickering random-dot stimuli. We classified cells as congruent ( $n = 82$ ) if they had a Pearson correlation coefficient between visual and vestibular heading tuning that was significantly  $>0.2$  (95% confidence interval not including 0.2, bootstrap;  $n = 1000$ ). Heading tuning curves for an example congruent cell are shown in Figure 2A. Similarly, we classified cells as opposite ( $n = 29$ ) if the correlation coefficient between visual and vestibular heading tuning was significantly  $<-0.2$  (95% CI not including  $-0.2$ , bootstrap). Data from an example opposite cell are shown in Figure 2D. The remaining neurons were denoted as unclassified ( $n = 53$ ). We found that these criteria produced a categorization of congruent and opposite cells that agreed well with classification by eye.

For comparison with previous studies (Gu et al., 2006), Figure 3 shows the distribution of differences in heading preference ( $|\Delta \text{Preferred Heading}|$ ) between the visual and vestibular self-motion conditions. The distribution is bimodal, indicating substantial proportions of both congruent and opposite cells. Although the bimodality is not as strong as in previous studies of MSTd (Gu et al., 2006), this may be due to the fact that heading tuning was measured in the frontoparallel plane in the present study, whereas Gu et al. (2006) measured this relationship in the horizontal plane.

### Interactions between self-motion and object motion in single-unit responses

We now examine how self-motion and object motion interact in the responses of MSTd neurons. Joint tuning profiles for heading

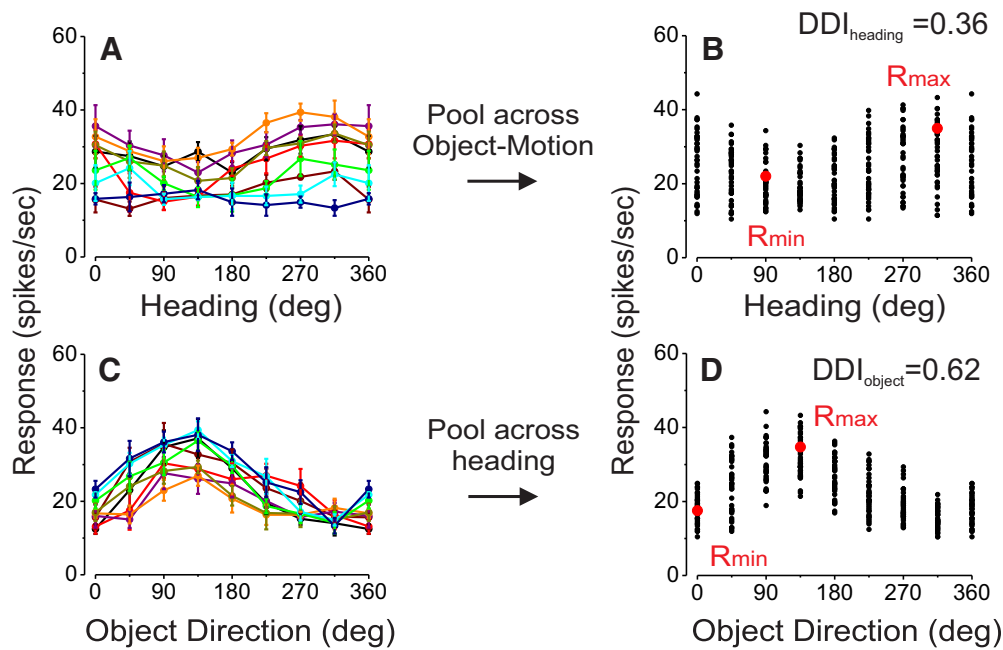
and object direction are shown in Figure 2, B, C, and E, F, for the example congruent and opposite cells, respectively. When self-motion and object-motion are presented simultaneously, both clearly influence the joint tuning of the example neurons. As a result, heading tuning curves are generally altered by object motion and object tuning curves are altered by heading. Interestingly, for the example congruent neuron (Fig. 2B, C), heading tuning curves are more consistent across object directions in the bimodal condition than in the visual condition, whereas object tuning curves are more consistent across headings in the visual condition than in the bimodal condition. In contrast, for the example opposite cell (Fig. 2E, F), the opposite pattern was observed: heading tuning is more consistent in the visual condition than the bimodal condition, and object tuning is more consistent in the bimodal condition than in the visual condition. In other words, vestibular self-motion signals appear to stabilize heading tuning for congruent cells, whereas they stabilize object tuning for opposite cells. This makes sense, intuitively, because vestibular heading tuning aligns with visual heading tuning for congruent cells, whereas it aligns with object tuning for opposite cells.

To quantify these observations, we used a DDI to measure the overall strength of tuning relative to response variability (Prince et al., 2002; DeAngelis and Uka, 2003). In our application, we wish to quantify both the overall strength of tuning for one variable (heading or object direction) and how it is influenced by variations in the other variable. Thus, we computed a DDI for heading ( $\text{DDI}_{\text{heading}}$ ) by pooling responses across all object directions (see Materials and Methods; Fig. 4A), such that  $\text{DDI}_{\text{heading}}$  reflects the consistency of heading tuning across different directions of object motion. Analogously, we computed a DDI for object motion ( $\text{DDI}_{\text{object}}$ ) by pooling responses across headings (see Materials and Methods; Fig. 4B).

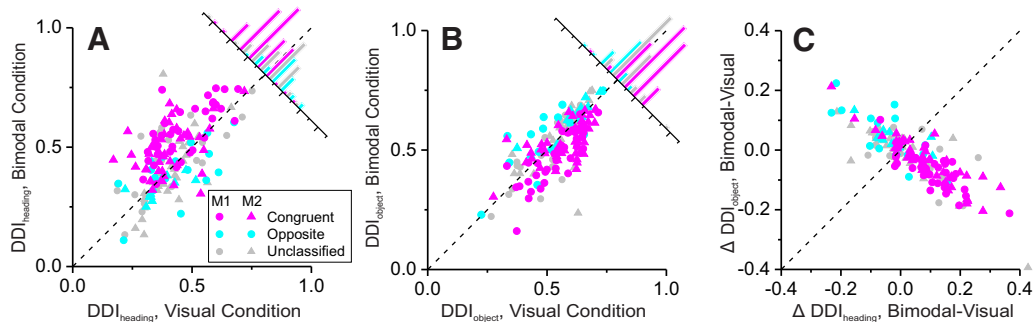
Results from our sample of MSTd neurons (Fig. 5) confirm the observations made for the example neurons of Figure 2. For congruent cells (magenta;  $n = 82$ ),  $\text{DDI}_{\text{heading}}$  is significantly greater in the bimodal condition than in the visual condition (paired  $t$  test,  $t_{(81)} = -8.29$ ,  $p = 2.1 \times 10^{-12}$ ), whereas for opposite cells (cyan;  $n = 29$ ),  $\text{DDI}_{\text{heading}}$  is significantly greater in the visual condition (paired  $t$  test,  $t_{(28)} = 1.98$ ,  $p = 0.0057$ ; Fig. 5A). Comparing the two types of cells, the median difference in  $\text{DDI}_{\text{heading}}$  between the bimodal and visual conditions was significantly greater for congruent cells than opposite cells (Wilcoxon rank sum test,  $p = 9.7 \times 10^{-8}$ ). Thus, vestibular signals stabilize the heading tuning of congruent cells, and weaken the heading tuning of opposite cells.

The opposite pattern of results was found for tuning to the direction of object motion (Fig. 5B). For congruent cells,  $\text{DDI}_{\text{object}}$  is significantly greater in the visual condition than in the bimodal condition (paired  $t$  test,  $t_{(81)} = 6.20$ ,  $p = 2.2 \times 10^{-8}$ ), whereas for opposite cells,  $\text{DDI}_{\text{object}}$  is greater in the bimodal condition (paired  $t$  test,  $t_{(28)} = -3.30$ ,  $p = 0.0026$ ). The median difference in  $\text{DDI}_{\text{object}}$  between bimodal and visual conditions is significantly less for congruent cells than opposite cells (Wilcoxon rank sum test,  $p = 8.3 \times 10^{-8}$ ). Hence vestibular signals stabilize the object tuning of opposite cells but substantially disrupt the object tuning of congruent cells.

To further summarize these effects, we plotted the difference in  $\text{DDI}_{\text{object}}$  between visual and bimodal conditions against the difference in  $\text{DDI}_{\text{heading}}$  between these stimulus conditions (Fig. 5C). Strikingly, across the entire population, there is a very strong inverse correlation between these variables (Spearman's rank  $r = -0.84$ ,  $p = 2.80 \times 10^{-45}$ ;  $n = 164$ ), with congruent and opposite cells being pretty well segregated by the unity-slope diagonal.



**Figure 4.** Illustration of the computation of the DDI. **A**, Heading tuning curves (mean responses  $\pm$  SD) are shown for each of eight directions of object motion for an example neuron. **B**, Responses of the same neuron are plotted as a function of heading after pooling across object motion directions.  $DDI_{\text{heading}}$  is computed from these pooled responses (see Materials and Methods). **C**, Object direction tuning curves of an example neuron are shown for each of the eight headings. Format as in **A**. **D**,  $DDI_{\text{object}}$  is computed after object motion responses are pooled across headings.



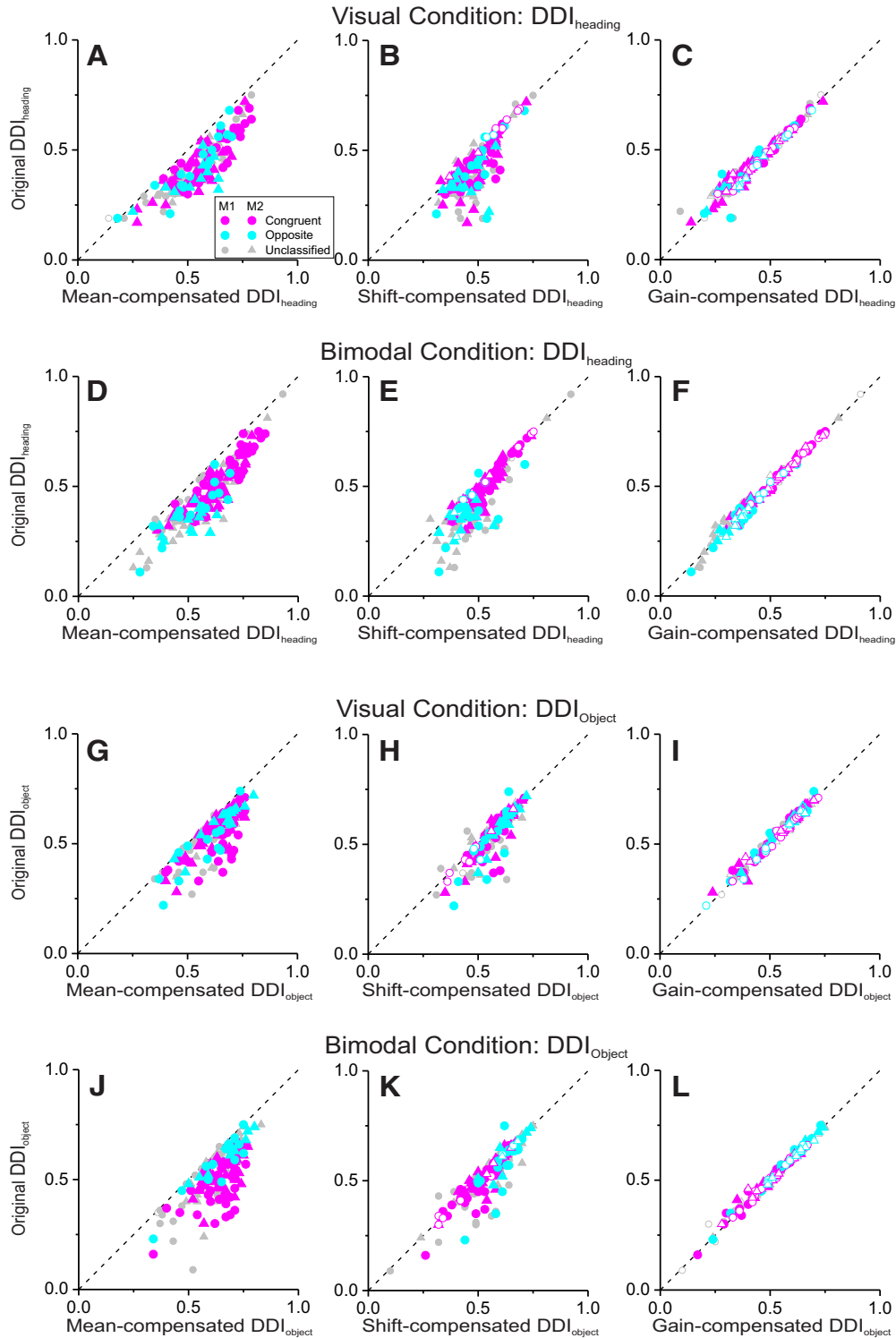
**Figure 5.** Summary of the effect of vestibular signals on heading and object tuning. **A**,  $DDI_{\text{heading}}$  for the bimodal condition is plotted against  $DDI_{\text{heading}}$  for the visual condition. For congruent cells (magenta), vestibular signals increase heading DDI, whereas vestibular signals decrease heading DDI for opposite cells (cyan). Gray symbols denote unclassified neurons. Circles and triangles indicate data for the two animals (M1 and M2). **B**,  $DDI_{\text{object}}$  values for the visual and bimodal conditions are plotted in the same format as **A**. Vestibular signals weaken object tuning for congruent cells and strengthen it for opposite cells. **C**, The difference in  $DDI_{\text{object}}$  between bimodal and visual conditions (ordinate) is plotted against the difference in  $DDI_{\text{heading}}$  between these two conditions (abscissa).

Note that there is also a strong negative correlation in Figure 5C for neurons that are not statistically categorized as congruent or opposite cells (gray symbols;  $r = -0.78$ ,  $p = 5.6 \times 10^{-12}$ ;  $n = 53$ ), indicating that the effect of vestibular signals on joint tuning for heading and object motion extends to most neurons in the population.

One may expect that neurons with stronger vestibular inputs would have greater differences in  $DDI_{\text{heading}}$  or  $DDI_{\text{object}}$  between visual and bimodal conditions. Indeed, this was the case for congruent cells: the difference in  $DDI_{\text{heading}}$  between bimodal and visual conditions is positively correlated with strength of vestibular tuning as quantified by computing DDI for the vestibular heading tuning curve ( $r = 0.53$ ,  $p = 4.1 \times 10^{-7}$ ;  $n = 82$ , data not shown). Analogously, the difference in  $DDI_{\text{object}}$  between bimodal and visual conditions is negatively correlated with strength of vestibular tuning for congruent cells ( $r = -0.37$ ,  $p = 0.050$ ;  $n = 82$ , data not shown). For opposite cells, vestibular tuning strength is significantly negatively correlated with the difference in  $DDI_{\text{heading}}$  between bimodal and visual conditions ( $r = -0.51$ ,  $p = 1.6 \times 10^{-6}$ ;

$n = 29$ ), and significantly positively correlated with the difference in  $DDI_{\text{object}}$  ( $r = 0.56$ ,  $p = 0.0019$ ;  $n = 29$ , data not shown). There was no significant difference overall, however, between the strength of vestibular heading tuning of congruent and opposite cells (Wilcoxon rank sum test,  $p = 0.22$ ).

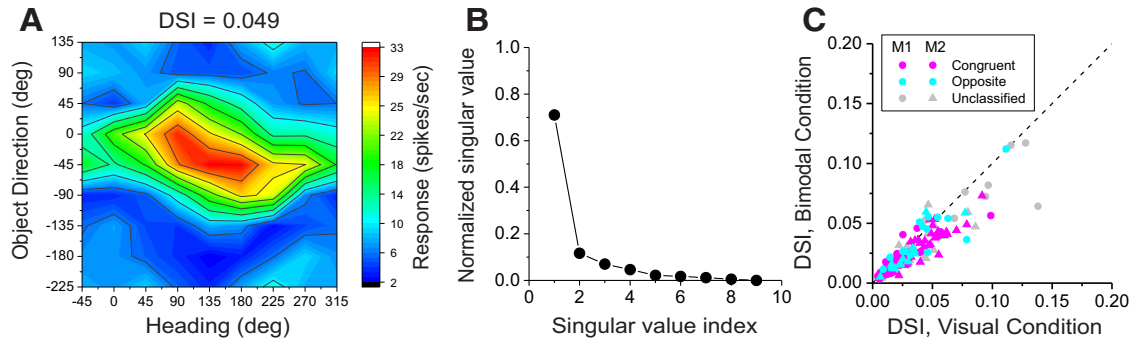
DDI quantifies the overall strength of tuning, but it does not indicate which aspects of heading tuning are modulated by object motion and vice versa. For example, different directions of object motion could change the response gain, shift the peak, or vertically shift the baseline of heading tuning curves (Fig. 2). We explored this issue by equalizing one of these aspects of tuning (e.g., peak location) and recomputing DDI values (Fig. 6; see Materials and Methods for details). For heading tuning, original  $DDI_{\text{heading}}$  values are significantly less than mean-compensated values (Wilcoxon signed rank test: Visual:  $p = 1.2 \times 10^{-28}$ , Bimodal:  $p = 1.6 \times 10^{-28}$ ; Fig. 6A,D) and shift-compensated values (Wilcoxon signed rank test: Visual:  $p = 1.6 \times 10^{-21}$ , Bimodal:  $p = 2.2 \times 10^{-18}$ ; Fig. 6B,E), but are not significantly different from gain-compensated  $DDI_{\text{heading}}$  values (Wilcoxon signed rank test:



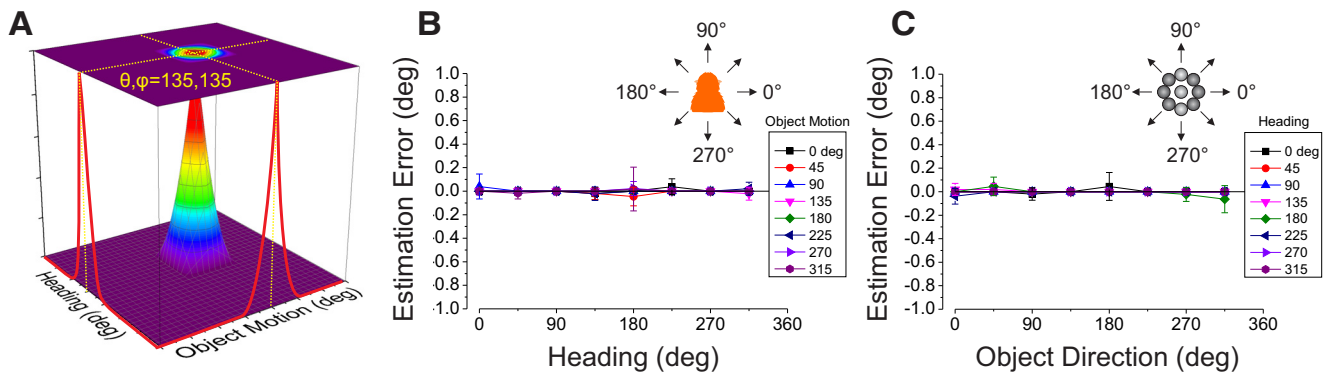
**Figure 6.** Summary of the nature of interactions between heading and object direction tuning. **A**, Original  $DDI_{\text{heading}}$  versus mean-compensated  $DDI_{\text{heading}}$  for the visual condition. Neurons with original and compensated values that are significantly different are plotted as filled symbols (bootstrap test,  $p < 0.05$ ); the remainder is shown as open symbols. Magenta, cyan, and gray symbols denote congruent, opposite, and unclassified cells, respectively. Circles and triangles indicate data for the two animals (M1 and M2). **B**, Original  $DDI_{\text{heading}}$  versus shift-compensated  $DDI_{\text{heading}}$  for the visual condition. Format as in **A**. **C**, Original  $DDI_{\text{heading}}$  versus gain-compensated  $DDI_{\text{heading}}$  for the visual condition. **D–F**, Same as **A–C** but for the bimodal condition. **G–I**, Same as **A–F**, except that original  $DDI_{\text{object}}$  values are plotted against compensated  $DDI_{\text{object}}$  values.

Visual:  $p = 0.19$ , Bimodal:  $p = 0.49$ ; Fig. 6C,F). This indicates that object direction mainly affects  $DDI_{\text{heading}}$  values through baseline and peak shifts, with relatively little effect of gain changes. Similar results are obtained for object direction tuning (Fig. 6G–L; mean-compensated  $DDI_{\text{object}}$ : Visual:  $p = 5.2 \times$

$10^{-28}$ , Bimodal:  $p = 1.8 \times 10^{-28}$ ; shift-compensated  $DDI_{\text{object}}$ : Visual:  $p = 1.3 \times 10^{-9}$ , Bimodal:  $p = 1.5 \times 10^{-6}$ ; gain-compensated  $DDI_{\text{object}}$ : Visual:  $p = 0.17$ , Bimodal:  $p = 0.58$ ). These findings suggest that variations in heading tuning or object tuning are mainly caused by vertical and horizontal shifts of the



**Figure 7.** Separability analysis of joint heading/object tuning profiles. **A**, A joint tuning profile for heading and object direction for an example neuron. Inseparability can be observed in the “tilt” of the long axis. **B**, SVD analysis was performed on the example tuning profile from **A**, and the first nine singular values are plotted here. The singular values are normalized such that they sum to unity. **C**, DSI for the bimodal condition is plotted against DSI for the visual condition. For both congruent (magenta) and opposite (cyan) cells, vestibular signals enhance the separability (reduce DSI) of joint tuning for heading and object motion. Circles, data for Monkey 1 (M1); triangles, data for Monkey 2 (M2).



**Figure 8.** Decoding heading and object direction by computing and directly marginalizing the joint posterior. **A**, An example joint posterior distribution, computed from a single trial of activity from our population of MSTd neurons. The red curves show the true marginal posteriors for heading and object direction. In this example, the peak of the joint posterior lies at the true stimulus values: 135° heading and 135° object motion. **B**, Heading estimation errors (obtained from the true marginal posterior over heading) are plotted as a function of heading for each object motion direction. Error bars indicate the SD over 500 simulations for each stimulus. **C**, Object motion estimation errors (obtained from the true marginal posterior over object direction) as a function of object motion direction. Format as in **B**.

tuning curves, with less contribution from variations in response gain.

The fact that heading tuning curves show horizontal and vertical shifts as a function of object direction (and vice versa) implies that joint tuning profiles are generally not multiplicatively separable. To quantify inseparabilities in joint tuning, we used SVD analysis to compute a DSI (see Materials and Methods; Depireux et al., 2001; Mazer et al., 2002). For neurons with joint tuning profiles that are multiplicatively separable, the first singular value should dominate, and the DSI value will be close to zero. For neurons with inseparable joint tuning, such as the slanted structure visible in the example of Figure 7A, the second and higher singular values will have non-negligible values (Fig. 7B), and the DSI value will be greater. The key question is whether inseparability is reduced by the inclusion of vestibular signals. Indeed, we found that DSI values for MSTd neurons were significantly smaller (closer to 0) in the bimodal condition than in the visual condition (Wilcoxon signed rank test,  $p = 1.3 \times 10^{-12}$ ;  $n = 164$ ). There was no significant difference in DSI between congruent, opposite and unclassified cells (ANCOVA, main effect of congruency,  $p = 0.082$ ) and no significant interaction between congruency and the effect of vestibular signals (ANCOVA, interaction effect,  $p = 0.069$ ). These results indicate that vestibular signals reduce the interdependencies of heading and object direction in the responses of MSTd neurons.

### Likelihood-based decoding of MSTd population responses

Given that self-motion and object motion interact in the responses of MSTd neurons in rather diverse ways, the question arises as to whether it is possible to decode heading or object direction accurately from a population of such neurons. One possibility is that the brain computes the joint posterior probability distribution over both heading ( $\theta$ ) and object direction ( $\varphi$ ) from the neural population response ( $\mathbf{r}$ ). To illustrate this, we assumed that the prior distributions over heading and object direction are flat (as in our experiment) and that population activity follows independent Poisson statistics, even though the latter is not accurate for real MSTd neurons (Gu et al., 2010, 2011). With these assumptions, we computed the joint posterior as follows (Dayan and Abbott, 2001):

$$P(\theta, \varphi | \mathbf{r}) \propto \prod_i \frac{e^{-f_i(\theta, \varphi)} f_i(\theta, \varphi)^{r_i}}{r_i!}, \quad (9)$$

where  $f_i(\theta, \varphi)$  denotes the joint tuning profile (Fig. 2B,C) of the  $i$ th neuron and  $r_i$  denotes the response of the  $i$ th neuron on a particular trial.

An example joint posterior, computed from a single-trial sample of activity from all neurons in our population (Fig. 8A), is sharply peaked around the true stimulus ( $\theta = 135^\circ$ ,  $\varphi = 135^\circ$ ). If the brain actually computes the joint posterior over both heading

and object direction, then it should be possible to estimate both variables accurately (note that this requires the decoder to know the joint heading-object tuning profile of each neuron). Indeed, we found that performance of this decoder was quite accurate and precise (Fig. 8B,C), with errors that were close to zero.

It may not be practical, in general, for the brain to compute the joint posterior because it could have several dimensions. For example, if there were multiple moving objects in a scene, then it would be necessary to compute a multidimensional posterior over heading and each object's motion. Alternatively, for tasks that involve judging heading, object motion can be treated as a "nuisance variable" and an ideal strategy would be for the brain to marginalize over object motion to estimate heading. In other words, it may be necessary to compute the marginal posterior over heading:

$$P(\theta|\mathbf{r}) = \sum_{\varphi} P(\theta, \varphi|\mathbf{r}). \quad (10)$$

Analogously, the brain could marginalize over heading to compute the marginal posterior for object motion. Indeed, marginalization is a general operation that is likely to be performed by the brain for a variety of forms of probabilistic inference.

Mathematically, marginalization simply involves integrating one (or more) variables out of a joint posterior, but how can such an operation be implemented neutrally? More specifically, how might the responses of MSTd neurons be decoded to marginalize over object motion and estimate heading, or vice versa? In general, performing optimal marginalization requires nonlinear operations (Beck et al., 2011); thus, we considered whether an approximate solution might be achieved through some form of linear decoding. We first considered the possibility that an approximation to marginalization could be achieved through a standard form of linear decoding that involves estimating the log-likelihood function as a response-weighted sum of tuning curves (maximum likelihood decoding; Dayan and Abbott, 2001). In simulations, we have reported recently that heading could be estimated robustly, in the presence of object motion, by computing approximate log-likelihood functions from a mixed population of congruent and opposite cells based on their vestibular tuning curves (Kim et al., 2016). This approach was quite successful in idealized populations of neurons but substantially less effective in the presence of tuning curve diversity (Kim et al., 2016); hence, it was not clear if it would be successful when applied to real data. Thus, we decoded the responses of real MSTd neurons by computing log-likelihoods based on either visual or vestibular tuning curves (see Materials and Methods; Eqs. 3, 4).

We found that decoding the bimodal responses of MSTd neurons using unimodal tuning curves generally produced large errors (Fig. 9A,B). We summarized these results by computing the root mean square error (RMSE) for estimates of heading (pooled across object directions) and the RMSE for estimates of object direction (pooled across headings). For heading estimation in the bimodal condition, RMSE averaged  $73.2^\circ \pm 0.98^\circ$  when heading was decoded based on visual heading tuning curves (Fig. 9C, filled blue bar), and  $77.6^\circ \pm 6.2^\circ$  when based on vestibular tuning curves (Fig. 9C, filled red bar). Errors were comparable or larger for the visual condition (Fig. 9C, open blue and red bars). Thus, the strategy of estimating log-likelihood functions over heading from unimodal tuning curves is unsuccessful in the presence of object motion. This strategy likely fails due to the considerable diversity of tuning properties across the MSTd population (Kim et al., 2016).

Analogous results are obtained for decoding object direction based on object-only tuning (Fig. 9D, orange bars), which is also

unsuccessful. Thus, it does not appear possible for the brain to effectively dissociate heading and object direction by estimating log-likelihood functions from response-weighted sums of unimodal tuning curves. However, this does not rule out the possibility that other forms of linear decoding might better approximate marginalization.

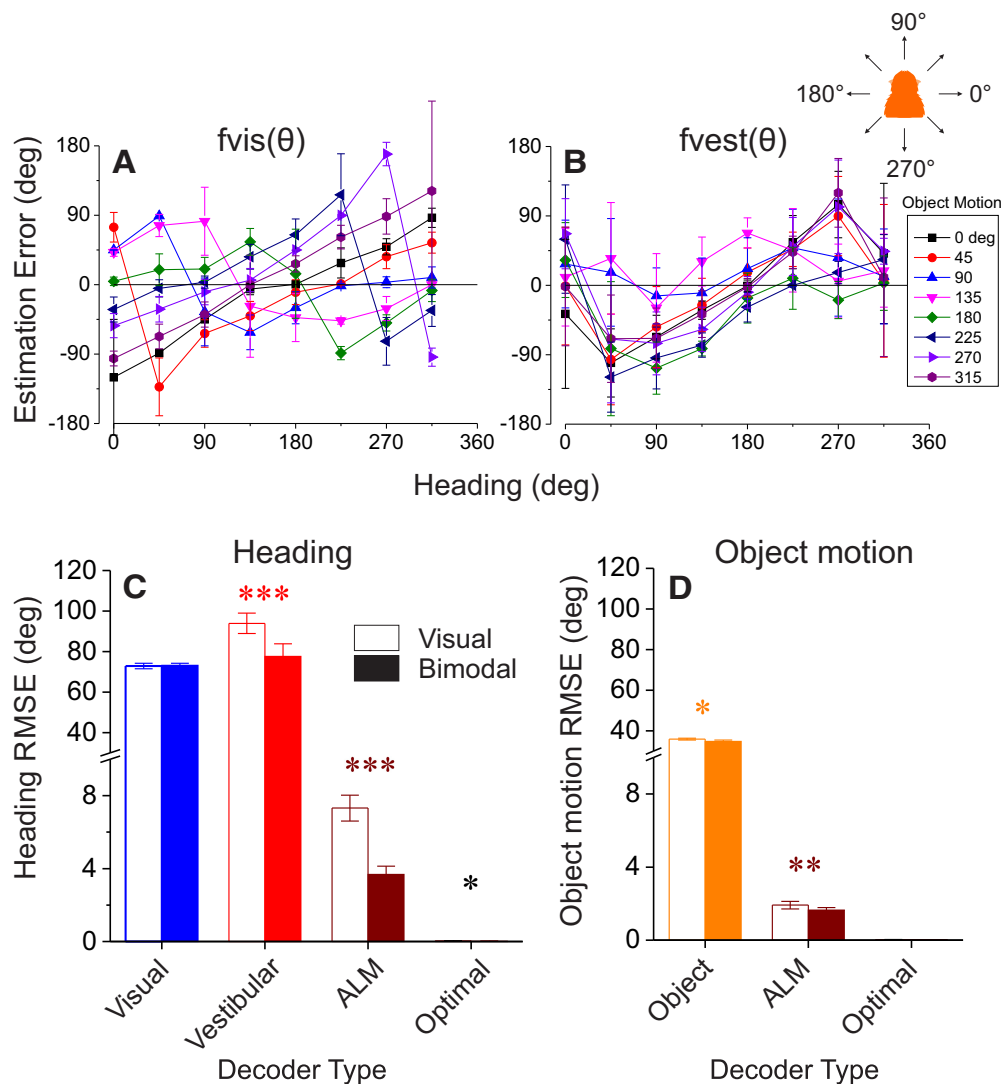
### Approximate linear marginalization

We have recently shown, through biologically-constrained simulations, that it is possible to learn a linear decoder of neural responses that can provide a good approximation to the marginal posterior (Kim et al., 2016). We now apply this ALM approach to population responses from area MSTd to determine whether it is capable of dissociating the effects of self-motion and object motion.

Similar to the likelihood computation described above, ALM involves a weighted linear decoding of neural responses. However, unlike the likelihood computation in which the weight profiles are constrained to be the neuron's tuning curve, ALM finds a weight profile for each neuron that achieves the best approximation to the marginal posterior. More specifically, ALM seeks to compute an approximate version of the marginal posterior (over heading, for example) that can be described as a member of the exponential family with linear sufficient statistics (see Materials and Methods; Eq. 6). We therefore attempted to find the parameters,  $\mathbf{h}(\theta)$  and  $g(\theta)$ , that best approximate the true marginalized posterior ( $P$ ) by the approximate form  $Q$ . In this formulation (Eq. 6),  $\mathbf{h}(\theta)$  represents a matrix of decoding weights, in which each neuron (row) has a weight corresponding to each possible heading (column), whereas  $g(\theta)$  represents a response-independent bias parameter for each heading.

We have demonstrated previously (Kim et al., 2016) that the desired quantities,  $\mathbf{h}(\theta)$  and  $g(\theta)$ , can be computed by applying multinomial logistic regression to minimize the Kullback–Leibler divergence between  $P$  and  $Q$  (Fig. 10A). We refer to the linear transformation that best approximates the marginalized posterior as ALM. The result of this analysis is a set of decoding weights for each neuron, given by  $\mathbf{h}(\theta)$ , that describe how much each neuron contributes to the representation of each possible heading,  $\theta$ , in the approximate marginal posterior. The resulting decoding weights provide the best approximation to the marginalized posterior over heading that can be achieved by a linear decoder in the exponential family. Analogously, we can also learn a set of decoding weights,  $\mathbf{h}(\varphi)$ , to estimate the marginal posterior over object direction (see Materials and Methods for details).

We applied ALM to responses of a subpopulation of 58 MSTd neurons: all 29 opposite cells and a subset of 29 congruent cells that were randomly chosen from our sample of 82 congruent cells. Results were obtained for 10 different randomly-chosen subsets of congruent cells, and the outcomes were averaged. This was done such that the contributions of congruent and opposite cells to the population were balanced. Our results demonstrate that ALM is quite accurate in extracting estimates of heading that are robust to object motion and vice versa (compare Figs. 10B,C and 9A,B). For estimating heading, RMSEs were  $7.32^\circ$  and  $3.68^\circ$  for the visual and bimodal conditions (Fig. 9C, brown bars), values that are dramatically smaller than the RMSE values obtained when heading is decoded by computing log-likelihoods based on vestibular tuning (Fig. 9C, red bars; visual:  $93.9^\circ$ , bimodal:  $77.6^\circ$ ) or visual tuning (Fig. 9C, blue bars; visual:  $72.9^\circ$ , bimodal:  $73.2^\circ$ ). Indeed, the median RMS heading error for ALM was highly significantly less than those obtained by decoding with visual or vestibular tuning curves (Wilcoxon signed rank tests,  $p < 10^{-12}$  for both visual and bimodal conditions). Importantly,



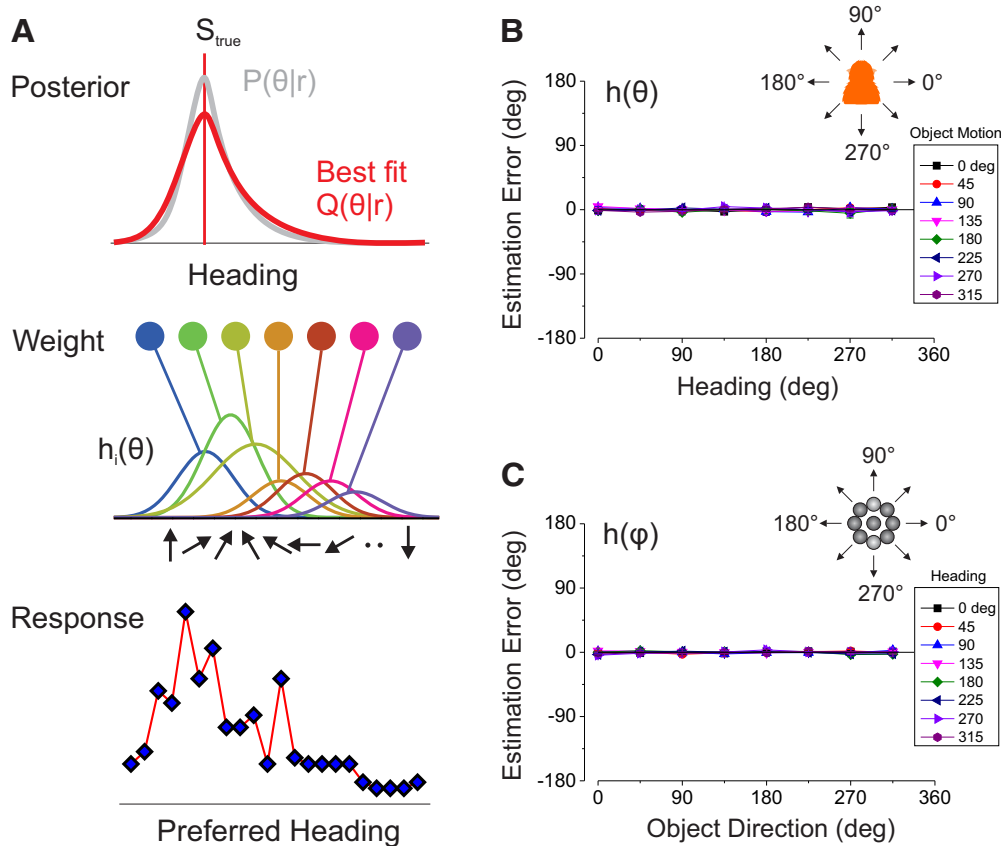
**Figure 9.** Summary of heading and object motion estimation errors for different decoding schemes. Results are based on decoding balanced populations of 29 congruent and 29 opposite cells. Results are shown averaged across 10 different populations of neurons comprising the 29 opposite cells and different random subsets of 29 congruent cells (chosen from the total of 82 congruent cells). **A**, Heading estimation errors are plotted as a function of true heading when heading is decoded by computing likelihood functions based on the visual heading tuning of each neuron. Each color represents a different object motion direction. For each of the 10 different populations of neurons, errors were averaged across 500 simulated trials for each distinct stimulus. Error bars represent SD of the averaged errors across the 10 simulated populations. **B**, Heading errors are plotted for the case in which decoding is based on the vestibular tuning curve of each neuron. Format as in **A**. **C**, RMSE averaged across all different headings and object motion directions is shown for the following methods of decoding: likelihood computation according to visual heading tuning (blue), likelihood computation according to vestibular tuning (red), ALM (brown), and direct marginalization of the joint posterior (black; but not visible because errors are essentially 0). Filled bars represent results for the bimodal condition and open bars represent results for the visual condition. Error bars represent 95% confidence intervals of RMSE across the 10 different populations with different random subsets of congruent cells. **D**, RMSE for object motion estimation based on likelihood computation using object tuning curves (orange), ALM (brown), and direct marginalization of joint posterior (black). Format as in **C**. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

the RMSE for ALM is significantly smaller in the bimodal condition ( $3.68^\circ$ ) than the visual condition ( $7.32^\circ$ ; Wilcoxon signed rank test,  $p = 1.5 \times 10^{-5}$ ; Fig. 9C, brown bars), indicating that vestibular signals significantly improve the accuracy with which the marginal posterior can be estimated from MSTd activity. This benefit of vestibular signals on ALM performance is roughly similar to recent behavioral results from monkeys (Dokka et al., 2015a), as discussed further below.

Analogous results were obtained when ALM was used to estimate object motion and marginalize over heading (Figs. 10C, Fig. 9D). RMSEs for estimating object direction were greatly reduced compared with when object direction was estimated by computing likelihoods based on object-only tuning (Fig. 9D; visual:  $p = 1.6 \times 10^{-10}$ , bimodal:  $p = 6.1 \times 10^{-12}$ ). Furthermore, the addition of vestibular signals modestly but significantly reduced the

RMSE in object direction from  $1.91^\circ$  in the visual condition to  $1.64^\circ$  in the bimodal condition (Fig. 9D, brown bars; Wilcoxon signed rank test,  $p = 0.0032$ ).

Although the results of Figure 9, C and D, were obtained from balanced populations of congruent and opposite cells, we also found that qualitatively similar results were obtained when ALM was trained to decode the responses of all MSTd neurons. For decoding heading, RMSE in the bimodal condition ( $0.37^\circ$ ) was significantly less than RMSE in the visual condition ( $0.81^\circ$ ), with the difference being statistically significant ( $p = 0.019$ , Wilcoxon signed rank test). For decoding object motion, the RMSE for the bimodal condition ( $0.28^\circ$ ) was also significantly less than the RMSE in the visual condition ( $0.36^\circ$ ;  $p = 2.3 \times 10^{-4}$ , Wilcoxon signed rank test). Thus, the effect of vestibular signals on estimation of heading and object motion can be observed when all



**Figure 10.** Schematic illustration of ALM and results for heading and object motion. **A**, Schematic illustration of ALM. The response of each neuron is multiplied by a weight profile,  $h_i(\theta)$ , which is learned through multinomial logistic regression. **B**, Heading estimation errors, derived from ALM, are plotted as a function of true heading for each object motion direction (color coded). For each of 10 different randomized populations of neurons, errors were averaged across 500 simulated trials for each distinct stimulus. Error bars represent SD of the averaged errors across the 10 simulated populations. **C**, Object motion estimation errors derived from ALM are plotted against true object direction for each heading. Format as in **B**.

neurons are decoded, although the differential effects become smaller as the errors approach zero.

Together the results of Figures 9 and 10 show that it is possible to linearly transform MSTd activity and obtain robust estimates of either heading or object direction; moreover, vestibular input significantly improves the accuracy of these estimates. These findings qualitatively mirror the effects of vestibular signals observed in recent behavioral studies (Fajen and Matthis, 2013; Dokka et al., 2015a,b), although one should exercise caution in comparing absolute errors between ALM and behavior (see Discussion).

### Comparison of ALM weights and neural tuning curves

Our previous computational study (Kim et al., 2016) showed that the decoding weight profiles learned by ALM for heading estimation were roughly similar to the vestibular tuning curves of model neurons, but not to the visual tuning curves (for opposite cells). To determine whether similar relationships hold for real MSTd neurons, we compared the learned weight profile for each neuron,  $h_i(\theta)$  or  $h_i(\varphi)$ , with the unimodal heading tuning curves and the object-only tuning curve. As described above, we averaged results across 10 balanced populations of 29 congruent and 29 opposite cells.

Results from an example congruent cell (Fig. 11A) reveal that the ALM weights for estimating heading,  $h_i(\theta)$ , have a profile similar in shape to the vestibular and visual heading tuning curves of this neuron. The ALM weights for estimating object motion,  $h_i(\varphi)$ , are roughly similar in shape to the object-only

tuning curve but not the visual heading tuning curve. For the example opposite cell (Fig. 11B), the results are quite different. Here,  $h_i(\theta)$  resembles the vestibular heading tuning curve, but not the visual heading tuning curve, and  $h_i(\varphi)$  resembles the object-only tuning curve and not the visual heading tuning. Notably,  $h_i(\theta)$  resembles the vestibular heading tuning for both the congruent cell and the opposite cell in Figure 11, A and B, similar to the findings of Kim et al. (2016) for simulated neural populations.

The pattern of results exhibited by these example neurons is largely confirmed across the population of MSTd neurons, as assessed by computing correlation coefficients between ALM weights and tuning curves. For congruent cells (Fig. 11C),  $h_i(\theta)$  is generally well correlated with both vestibular and visual heading tuning curves: the median correlation coefficients are 0.69 and 0.65, respectively. Both of these values are significantly different from zero (sign tests: vestibular:  $p = 1.6 \times 10^{-76}$ , visual:  $p = 5.6 \times 10^{-61}$ ), and the median correlation with vestibular heading tuning is significantly greater than the correlation with visual tuning (Wilcoxon signed rank test,  $p = 4.1 \times 10^{-4}$ ). By comparison,  $h_i(\varphi)$  for congruent cells is modestly, but significantly, correlated with object-only tuning (median correlation coefficient = 0.30; sign test,  $p = 2.6 \times 10^{-11}$ ).

For opposite cells (Fig. 11D), a different pattern emerges. We find that  $h_i(\theta)$  is generally positively correlated with vestibular heading tuning (median correlation coefficient = 0.39; sign test,  $p = 5.6 \times 10^{-18}$ ) but shows no systematic correlation with visual heading tuning (median correlation coefficient = -0.20; sign test,  $p = 0.80$ ). Comparing ALM heading weights across congruent

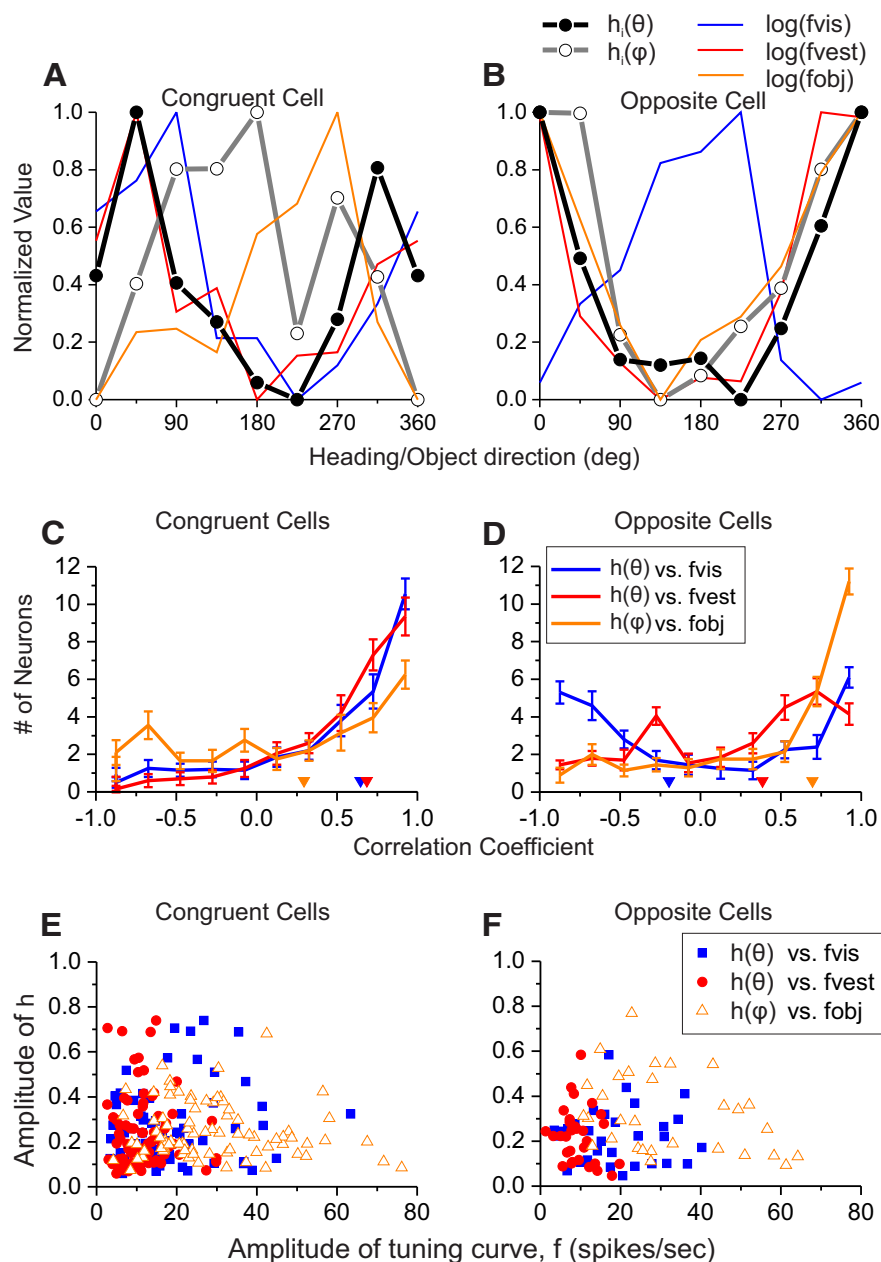
ent and opposite cells, we see that ALM weights tend to be positively correlated with vestibular heading tuning for both types of neurons, roughly consistent with the findings of Kim et al. (2016) for model neurons.

For opposite cells, ALM weights for estimating object direction,  $h_i(\varphi)$ , are strongly correlated with object-only tuning (median correlation coefficient = 0.70; sign test,  $p = 5.3 \times 10^{-49}$ ), and this correlation is significantly stronger than the corresponding relationship for congruent cells (Wilcoxon rank sum test,  $p = 1.3 \times 10^{-13}$ ). This finding is consistent with the idea that opposite cells may contribute more substantially to estimating object direction than congruent cells.

The results of Figure 11, C and D, suggest that the weight profiles learned by ALM have some systematic relationships with the unimodal tuning curves for heading and object motion. We also examined whether the amplitudes of the ALM weight profiles,  $h_i(\theta)$  and  $h_i(\varphi)$ , are correlated with the amplitudes of the unimodal tuning curves. We found no significant correlation between the amplitude of  $h_i(\theta)$  and the amplitude of visual or vestibular heading tuning curves, for either congruent or opposite cells (Fig. 11E,F; Spearman rank correlations,  $p > 0.10$ ). Similarly, we found no correlation between the amplitude of  $h_i(\varphi)$  and object tuning curves (Spearman rank correlations,  $p > 0.66$ ). Thus, ALM does not weight the responses of MSTd neurons according to the strength of their unimodal tuning curves (unlike the log-likelihood computation), even though there are systematic relationships between the shapes of ALM weights and tuning curves. This suggests that the superior performance of ALM over the log-likelihood computation arises, at least in part, from applying a gain factor to neurons that is not predictable from their tuning strength.

## Discussion

Our findings show that vestibular signals contribute to generating robust representations of self-motion and object motion in area MSTd. Specifically, vestibular signals enhance the separability of the joint heading/object tuning profiles of MSTd neurons. As a result, vestibular signals enhance the representation of heading in congruent cells and the representation of object motion in opposite cells. We demonstrate that a linear transformation of MSTd responses (ALM) can allow fairly accurate decoding of either heading or object motion while marginalizing over variations in the other variable, and that vestibular signals facilitate this marginalization computation. Together, our findings demonstrate a fundamental role for di-



**Figure 11.** Summary of the relationships between learned decoding weights and tuning curves. **A**, Comparison between heading weights [ $h(\theta)$ ; black filled symbols], object direction weights [ $h(\varphi)$ ; open gray symbols], visual heading tuning (blue curve), vestibular heading tuning (red curve), and object-only tuning (orange curve) for an example congruent cell. **B**, Same as **A**, for an example opposite cell. **C**, Distributions of correlation coefficients between ALM weights and tuning curves for congruent cells: heading weights versus visual tuning (blue), heading weights versus vestibular tuning (red), and object weights versus object-only tuning (orange). Arrowheads denote median values. **D**, Distributions of correlation coefficients for opposite cells. Format as in **C**. Error bars represent 95% confidence intervals across 10 different populations with different random subsets of congruent cells. **E**, The peak-trough amplitude of the ALM-derived heading weight function,  $h(\theta)$ , is plotted against the peak-trough amplitude of the visual (blue) and vestibular (red) heading tuning curves. Similarly, the amplitude of the object decoding weight profile,  $h(\varphi)$ , is plotted against amplitude of the object-only tuning function (orange) for each neuron. Data are shown for congruent cells ( $n = 82$ ). **F**, Amplitude comparison for opposite cells ( $n = 29$ ; format as in **E**).

verse multisensory representations in performing neural computations that parse sensory inputs into components that reflect separate physical causes.

## Limitations of stimulus design

Our exploration of interactions between self-motion and object motion was limited to the frontoparallel plane, because it was not practically feasible to record from neurons long enough to study

the interactions in all three dimensions. We chose the frontoparallel plane, instead of the horizontal plane, because there are fewer visual cues to distinguish object motion from self-motion, thus making the marginalization problem more difficult. In the horizontal plane, there would also be looming and changing disparity cues that might be exploited to dissociate self-motion and object motion. Previous work (Logan and Duffy, 2006) suggests that the relationship between visual heading tuning and object direction tuning is similar in the frontoparallel and horizontal planes for MSTd neurons; thus, we would expect similar results in the horizontal plane. We recognize, however, that the additional visual cues that would be available in the horizontal plane might also aid the dissociation of self-motion and object motion. Thus, there might be less need for marginalization operations when self-motion occurs in the horizontal plane.

Another limitation is that we used a large, multipart object to elicit robust responses of MSTd neurons to object motion. Although our object is probably larger than most objects encountered in natural scenes, this choice was driven by the fact that most MSTd neurons do not respond well to small objects (Komatsu and Wurtz, 1988a,b), as well as the need to have the object easily visually segmented from the background dots.

### Approximate linear marginalization

Marginalization operations are of considerable importance because they allow the brain to estimate a variable of interest and integrate out the effects of nuisance variables. Marginalization can be easily accomplished with a simple linear decoder if nuisance parameters simply change the gain of neural responses (Ma et al., 2006; Beck et al., 2011). However, the interactions between self-motion and object motion in MSTd responses are generally not well described as gain modulations (Figs. 2, 6). Under these conditions, marginalization generally requires nonlinear transformations of neural responses (Beck et al., 2011) that may be difficult for the brain to implement.

We recently demonstrated that it is possible to approximate marginalization using a linear transformation of neural responses that we call ALM (Kim et al., 2016). ALM was successful in biologically-constrained simulations, even when the joint heading-object tuning of model neurons was strongly inseparable. Our present findings show that ALM performs quite well when applied to responses of real MSTd neurons, even when the neural populations are modest in size (~60 neurons). Our results suggest that simple linear transformations could help to solve a variety of computational problems that require marginalizing out the effects of nuisance parameters. Further work is needed to determine how well ALM performs when multiple nuisance variables are present, but the approach is general and should be applicable to larger-scale problems.

### Neural substrates for dissociating heading and object motion

In macaque monkeys, visual motion processing involves a network of cortical areas, including MT, MST, VIP, and FST (Orban et al., 2004). Although many studies have been conducted with a variety of stimuli, including plaids and random-dot patterns that contain multiple velocity components (Treue et al., 2000; Rust et al., 2006; McDonald et al., 2014), only a few previous studies have tested neurons with combinations of visual motion cues that simulate both self-motion and object motion (Logan and Duffy, 2006; Sato et al., 2010; Kishore et al., 2012). Consistent with our findings, these previous studies report that interactions between heading and object motion can be complex, showing a variety of subadditive and superadditive interactions (Sato et al., 2010). A

population vector decoder failed to recover heading accurately when object motion was opposite to self-motion (Logan and Duffy, 2006), consistent with our finding that likelihood functions computed as response-weighted sums of tuning curves fail to approximate marginalization. Importantly, these previous studies only examined interactions between visual cues to self-motion and object motion; our study is the first to examine whether vestibular signals help to dissociate self-motion from object motion in neural responses.

Our findings suggest that area MSTd is well suited to play a role in dissociating self-motion and object motion through multisensory integration. However, MSTd neurons generally do not respond well to small moving stimuli (Komatsu and Wurtz, 1988a,b). Indeed, this was also our experience in preliminary experiments, leading us to adopt a stimulus design in which a large multipart object moved independently in the virtual environment (see Materials and Methods). The role of MSTd in representing object motion may therefore be limited to situations in which moving objects subtend a substantial portion of the visual field, and other areas may represent the motion of small objects during self-motion. One candidate area is MSTl, where neurons have smaller receptive fields and frequently exhibit strong surround suppression (Eifuku and Wurtz, 1998). It is unknown, however, whether MSTl neurons carry vestibular signals that could aid in dissociating self-motion and object motion. Another candidate area is V6 (Galletti et al., 1999, 2001), based on human imaging studies (Pitzalis et al., 2010, 2012). To our knowledge, a direct examination of interactions between object motion and self-motion has not been conducted in V6 at the single-neuron level. However, recent results (Fan et al., 2015) indicate that monkey V6 neurons do not carry vestibular signals regarding body translation. Other visual-vestibular areas that may contribute include the VIP area (Chen et al., 2011b), VPS area (Chen et al., 2011a), and the FEF (Gu et al., 2016).

It is important to emphasize that multisensory mechanisms for dissociating self-motion and object motion may operate in parallel with purely visual mechanisms, such as those involving center-surround mechanisms (Anstis and Reinhardt-Rutland, 1976; Frost and Nakayama, 1983). Indeed, computational models have demonstrated that such visual strategies can be effective (Royden, 2002; Royden and Holloway, 2014). Our results suggest that multisensory mechanisms may augment these visual strategies.

Until recently, the functional role of opposite cells has remained unclear. Previous work showed that opposite cells are ill-suited for cue integration (Gu et al., 2008, 2014) and cue-weighting (Fetsch et al., 2011) in heading perception. Choice-related activity suggested that opposite cells might be decoded according to their vestibular heading preferences (Gu et al., 2008, 2014), and simulations showed this to be an effective strategy for estimating heading during object motion (Kim et al., 2016). Our present results are broadly consistent with this idea, as the weights learned by ALM to estimate heading tend to resemble vestibular tuning for opposite cells. Together, these findings suggest that opposite cells make important contributions to dissociating self-motion and object motion; moreover, the computational strategy may be applicable to source separation problems involving other sensory modalities.

### Relationship to behavioral studies

Psychophysical studies have explored whether the visual system can parse retinal image motion into components related to self-motion and object motion (Rushton and Warren, 2005; Warren

and Rushton, 2007, 2008, 2009; Matsumiya and Ando, 2009). These “flow parsing” studies show that global patterns of optic flow consistent with self-motion can alter the perceived velocity of small moving objects, consistent with discounting global flow, even when the object is located in the hemi-field opposite to optic flow (Warren and Rushton, 2009). However, it is not clear from these studies whether the visual system, by itself, can fully dissociate self-motion and object motion, nor whether vestibular signals may contribute to flow-parsing.

A few recent psychophysical studies have examined perception of object motion during both real and simulated self-motion (MacNeilage et al., 2012; Fajen and Matthis, 2013; Dokka et al., 2015b). Consistent with our findings from decoding MSTd neurons, Dokka et al. (2015b) report that vestibular signals reduce biases in perceived object direction during self-motion. Dokka et al. (2015b) found substantially larger biases than we predict from ALM, however, suggesting that neural processing in humans may not achieve near-optimal marginalization. However, the species and stimulus conditions are different, making direct comparisons difficult. Using a different paradigm, Fajen et al. (2013) also found evidence that vestibular signals contribute to human judgements of object direction during self-motion, indicating that multisensory mechanisms play important roles in dissociating self-motion and object motion.

One can also consider the complementary question of how moving objects affect judgements of heading. Observers can exhibit substantial biases when they judge heading in the presence of objects that move independently in the world, especially when those objects overlap the focus of expansion in optic flow (Warren and Saunders, 1995; Royden and Hildreth, 1996, 1999; Fajen and Kim, 2002; Mapstone and Duffy, 2010). It is generally difficult to compare our population decoding results from MSTd to these previous studies because stimulus conditions were quite different. Our results regarding vestibular signals are most comparable to a recent study of monkey heading perception (Dokka et al., 2015a), despite substantial differences in object size and directions of self-motion. Dokka et al. (2015a) reported that addition of vestibular self-motion signals largely eliminated heading biases caused by object motion, indicating that vestibular signals contribute to dissociating self-motion and object motion. Notably, the relative magnitudes of heading biases found by Dokka et al. (2015a) in the visual and bimodal conditions (see their Fig. 4B) are approximately comparable to our results from applying ALM to MSTd responses (Fig. 9C, brown bars).

In closing, in conjunction with recent computational work (Kim et al., 2016), the present findings suggest that a linear transformation of the responses of a diverse population of multisensory neurons may be sufficient to dissociate self-motion and object motion. Future studies can build on this foundation by recording and manipulating neural activity while animals dissociate object motion and self-motion perceptually, and these efforts will be important to tease apart the circuitry underlying these fundamental computations.

## References

- Anstis SM, Reinhardt-Rutland AH (1976) Interactions between motion aftereffects and induced movement. *Vision Res* 16:1391–1394. [CrossRef Medline](#)
- Beck JM, Latham PE, Pouget A (2011) Marginalization in neural circuits with divisive normalization. *J Neurosci* 31:15310–15319. [CrossRef Medline](#)
- Bishop CM (2006) Pattern recognition and machine learning. New York: Springer.
- Bremmer F, Klam F, Duhamel JR, Ben Hamed S, Graf W (2002) Vestibular interactive responses in the macaque ventral intraparietal area (VIP). *Eur J Neurosci* 16:1569–1586. [CrossRef Medline](#)
- Chen A, Gu Y, Takahashi K, Angelaki DE, DeAngelis GC (2008) Clustering of self-motion selectivity and visual response properties in macaque area MSTd. *J Neurophysiol* 100:2669–2683. [CrossRef Medline](#)
- Chen A, DeAngelis GC, Angelaki DE (2011a) Convergence of vestibular and visual self-motion signals in an area of the posterior sylvian fissure. *J Neurosci* 31:11617–11627. [CrossRef Medline](#)
- Chen A, DeAngelis GC, Angelaki DE (2011b) Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex. *J Neurosci* 31:12036–12052. [CrossRef Medline](#)
- Chen A, DeAngelis GC, Angelaki DE (2013) Functional specializations of the ventral intraparietal area for multisensory heading discrimination. *J Neurosci* 33:3567–3581. [CrossRef Medline](#)
- Dayan P, Abbott LF (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA: MIT.
- DeAngelis GC, Uka T (2003) Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J Neurophysiol* 89:1094–1111. [CrossRef Medline](#)
- Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85:1220–1234. [Medline](#)
- Dokka K, DeAngelis GC, Angelaki DE (2015a) Multisensory integration of visual and vestibular signals improves heading discrimination in the presence of a moving object. *J Neurosci* 35:13599–13607. [CrossRef Medline](#)
- Dokka K, MacNeilage PR, DeAngelis GC, Angelaki DE (2015b) Multisensory self-motion compensation during object trajectory judgments. *Cereb Cortex* 25:619–630. [CrossRef Medline](#)
- Dupin L, Wexler M (2013) Motion perception by a moving observer in a three-dimensional environment. *J Vis* 13(2):15–14. [CrossRef Medline](#)
- Eifuku S, Wurtz RH (1998) Response to motion in extrastriate area MSTl: center-surround interactions. *J Neurophysiol* 80:282–296. [Medline](#)
- Fajen BR, Kim NG (2002) Perceiving curvilinear heading in the presence of moving objects. *J Exp Psychol Hum Percept Perform* 28:1100–1119. [CrossRef Medline](#)
- Fajen BR, Matthis JS (2013) Visual and non-visual contributions to the perception of object motion during self-motion. *PLoS One* 8:e55446. [CrossRef Medline](#)
- Fan RH, Liu S, DeAngelis GC, Angelaki DE (2015) Heading tuning in macaque area V6. *J Neurosci* 35:16303–16314. [CrossRef Medline](#)
- Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE (2011) Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci* 15:146–154. [CrossRef Medline](#)
- Frost BJ, Nakayama K (1983) Single visual neurons code opposing motion independent of direction. *Science* 220:744–745. [CrossRef Medline](#)
- Galletti C, Fattori P, Gamberini M, Kutz DF (1999) The cortical visual area V6: brain location and visual topography. *Eur J Neurosci* 11:3922–3936. [CrossRef Medline](#)
- Galletti C, Gamberini M, Kutz DF, Fattori P, Luppino G, Matelli M (2001) The cortical connections of area V6: an occipito-parietal network processing visual information. *Eur J Neurosci* 13:1572–1588. [CrossRef Medline](#)
- Gu Y, Watkins PV, Angelaki DE, DeAngelis GC (2006) Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J Neurosci* 26:73–85. [CrossRef Medline](#)
- Gu Y, Angelaki DE, DeAngelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci* 11:1201–1210. [CrossRef Medline](#)
- Gu Y, Fetsch CR, Adeyemo B, DeAngelis GC, Angelaki DE (2010) Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron* 66:596–609. [CrossRef Medline](#)
- Gu Y, Liu S, Fetsch CR, Yang Y, Fok S, Sunkara A, DeAngelis GC, Angelaki DE (2011) Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* 71:750–761. [CrossRef Medline](#)
- Gu Y, Angelaki DE, DeAngelis GC (2014) Contribution of correlated noise and selective decoding to choice probability measurements in extrastriate visual cortex. *eLife* 3:e02670. [CrossRef Medline](#)
- Gu Y, Cheng Z, Yang L, DeAngelis GC, Angelaki DE (2016) Multisensory convergence of visual and vestibular heading cues in the pursuit area of the frontal eye field. *Cereb Cortex* 26:3785–3801. [CrossRef Medline](#)
- Hinton GE, Dayan P (1996) Varieties of Helmholtz machine. *Neural Netw* 9:1385–1403. [CrossRef Medline](#)

- Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci* 9:690–696. [CrossRef Medline](#)
- Kim HR, Pitkow X, Angelaki DE, DeAngelis GC (2016) A simple approach to ignoring irrelevant variables by population decoding based on multi-sensory neurons. *J Neurophysiol* 116:1449–1467.
- Kishore S, Hornick N, Sato N, Page WK, Duffy CJ (2012) Driving strategy alters neuronal responses to self-movement: cortical mechanisms of distracted driving. *Cereb Cortex* 22:201–208. [CrossRef Medline](#)
- Komatsu H, Wurtz RH (1988a) Relation of cortical areas MT and MST to pursuit eye movements: I. Localization and visual properties of neurons. *J Neurophysiol* 60:580–603. [Medline](#)
- Komatsu H, Wurtz RH (1988b) Relation of cortical areas MT and MST to pursuit eye movements: III. Interaction with full-field visual stimulation. *J Neurophysiol* 60:621–644. [Medline](#)
- Logan DJ, Duffy CJ (2006) Cortical area MSTd combines visual cues to represent 3-D self-movement. *Cereb Cortex* 16:1494–1507. [CrossRef Medline](#)
- Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9:1432–1438. [CrossRef Medline](#)
- MacNeilage PR, Zhang Z, DeAngelis GC, Angelaki DE (2012) Vestibular facilitation of optic flow parsing. *PLoS One* 7:e40264. [CrossRef Medline](#)
- Mapstone M, Duffy CJ (2010) Approaching objects cause confusion in patients with Alzheimer's disease regarding their direction of self-movement. *Brain* 133:2690–2701. [CrossRef Medline](#)
- Matsumiya K, Ando H (2009) World-centered perception of 3D object motion during visually guided self-motion. *J Vis* 9(1):15 11–13. [CrossRef Medline](#)
- Mazer JA, Vinje WE, McDermott J, Schiller PH, Gallant JL (2002) Spatial frequency and orientation tuning dynamics in area V1. *Proc Natl Acad Sci U S A* 99:1645–1650. [CrossRef Medline](#)
- McDonald JS, Clifford CW, Solomon SS, Chen SC, Solomon SG (2014) Integration and segregation of multiple motion signals by neurons in area MT of primate. *J Neurophysiol* 111:369–378. [CrossRef Medline](#)
- Nabney IT (2002) *Netlab: algorithms for pattern recognition*. London: Springer.
- Orban GA, Van Essen D, Vanduffel W (2004) Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn Sci* 8:315–324. [CrossRef Medline](#)
- Pitzalis S, Sereno MI, Committeri G, Fattori P, Galati G, Patria F, Galletti C (2010) Human v6: the medial motion area. *Cereb Cortex* 20:411–424. [CrossRef Medline](#)
- Pitzalis S, Fattori P, Galletti C (2012) The functional role of the medial motion area V6. *Front Behav Neurosci* 6:91. [CrossRef Medline](#)
- Prince SJ, Pointon AD, Cumming BG, Parker AJ (2002) Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. *J Neurophysiol* 87:191–208. [CrossRef Medline](#)
- Royden CS (2002) Computing heading in the presence of moving objects: a model that uses motion-opponent operators. *Vision Res* 42:3043–3058. [CrossRef Medline](#)
- Royden CS, Hildreth EC (1996) Human heading judgments in the presence of moving objects. *Percept Psychophys* 58:836–856. [CrossRef Medline](#)
- Royden CS, Hildreth EC (1999) Differential effects of shared attention on perception of heading and 3-D object motion. *Percept Psychophys* 61:120–133. [CrossRef Medline](#)
- Royden CS, Holloway MA (2014) Detecting moving objects in an optic flow field using direction- and speed-tuned operators. *Vision Res* 98:14–25. [CrossRef Medline](#)
- Rushton SK, Warren PA (2005) Moving observers, relative retinal motion and the detection of object movement. *Curr Biol* 15:R542–R543. [CrossRef Medline](#)
- Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9:1421–1431. [CrossRef Medline](#)
- Sato N, Kishore S, Page WK, Duffy CJ (2010) Cortical neurons combine visual cues about self-movement. *Exp Brain Res* 206:283–297. [CrossRef Medline](#)
- Treue S, Hol K, Rauber HJ (2000) Seeing multiple directions of motion-physiology and psychophysics. *Nat Neurosci* 3:270–276. [CrossRef Medline](#)
- Van Essen DC, Lewis JW, Drury HA, Hadjikhani N, Tootell RB, Bakircioglu M, Miller MI (2001) Mapping visual cortex in monkeys and humans using surface-based atlases. *Vision Res* 41:1359–1378. [CrossRef Medline](#)
- Warren PA, Rushton SK (2007) Perception of object trajectory: parsing retinal motion into self and object movement components. *J Vis* 7(11):2 1–11. [CrossRef Medline](#)
- Warren PA, Rushton SK (2008) Evidence for flow-parsing in radial flow displays. *Vision Res* 48:655–663. [CrossRef Medline](#)
- Warren PA, Rushton SK (2009) Optic flow processing for the assessment of object movement during ego movement. *Curr Biol* 19:1555–1560. [CrossRef Medline](#)
- Warren WH Jr, Saunders JA (1995) Perceiving heading in the presence of moving objects. *Perception* 24:315–331. [CrossRef Medline](#)
- Wexler M (2003) Voluntary head movement and allocentric perception of space. *Psychol Sci* 14:340–346. [CrossRef Medline](#)
- Wexler M, van Boxtel JJ (2005) Depth perception by the active observer. *Trends Cogn Sci* 9:431–438. [CrossRef Medline](#)