

# 3D Visual Response Properties of MSTd Emerge from an Efficient, Sparse Population Code

 Michael Beyeler,<sup>1</sup> Nikil Dutt,<sup>1,2</sup> and Jeffrey L. Krichmar<sup>1,2</sup>

Departments of <sup>1</sup>Computer Science and <sup>2</sup>Cognitive Sciences, University of California, Irvine, Irvine, California 92697

Neurons in the dorsal subregion of the medial superior temporal (MSTd) area of the macaque respond to large, complex patterns of retinal flow, implying a role in the analysis of self-motion. Some neurons are selective for the expanding radial motion that occurs as an observer moves through the environment (“heading”), and computational models can account for this finding. However, ample evidence suggests that MSTd neurons exhibit a continuum of visual response selectivity to large-field motion stimuli. Furthermore, the underlying computational principles by which these response properties are derived remain poorly understood. Here we describe a computational model of macaque MSTd based on the hypothesis that neurons in MSTd efficiently encode the continuum of large-field retinal flow patterns on the basis of inputs received from neurons in MT with receptive fields that resemble basis vectors recovered with non-negative matrix factorization. These assumptions are sufficient to quantitatively simulate neurophysiological response properties of MSTd cells, such as 3D translation and rotation selectivity, suggesting that these properties might simply be a byproduct of MSTd neurons performing dimensionality reduction on their inputs. At the population level, model MSTd accurately predicts eye velocity and heading using a sparse distributed code, consistent with the idea that biological MSTd might be well equipped to efficiently encode various self-motion variables. The present work aims to add some structure to the often contradictory findings about macaque MSTd, and offers a biologically plausible account of a wide range of visual response properties ranging from single-unit selectivity to population statistics.

**Key words:** heading selectivity; MSTd; non-negative matrix factorization; optic flow; visual motion processing

## Significance Statement

Using a dimensionality reduction technique known as non-negative matrix factorization, we found that a variety of medial superior temporal (MSTd) neural response properties could be derived from MT-like input features. The responses that emerge from this technique, such as 3D translation and rotation selectivity, spiral tuning, and heading selectivity, can account for a number of empirical results. These findings (1) provide a further step toward a scientific understanding of the often nonintuitive response properties of MSTd neurons; (2) suggest that response properties, such as complex motion tuning and heading selectivity, might simply be a byproduct of MSTd neurons performing dimensionality reduction on their inputs; and (3) imply that motion perception in the cortex is consistent with ideas from the efficient-coding and free-energy principles.

## Introduction

Neurons in the dorsal subregion of the medial superior temporal (MSTd) area of monkey extrastriate cortex respond to relatively large and complex patterns of retinal flow, often preferring a mixture of translational, rotational, and, to a lesser degree, deformational flow components (Saito et al., 1986; Tanaka and Saito,

1989a; Duffy and Wurtz, 1991a,b; Orban et al., 1992; Lagae et al., 1994; Mineault et al., 2012). This has led researchers to suggest that MSTd might play a key role in visual motion processing for self-movement perception. In fact, one of the most commonly documented response properties of MSTd neurons is that of heading selectivity (Tanaka and Saito, 1989a; Duffy and Wurtz, 1995; Lappe et al., 1996; Britten and Van Wezel, 2002; Page and Duffy, 2003; Gu et al., 2006, 2012; Logan and Duffy, 2006), and computational models can account for this finding (Perrone, 1992; Zhang et al., 1993; Perrone and Stone, 1994, 1998; Lappe et al., 1996; Beintema and van den Berg, 1998, 2000; Zemel and Sejnowski, 1998). However, a number of recent studies have found that nearly all visually responsive neurons in macaque MSTd signal the three-dimensional (3D) direction of self-translation (i.e., “heading;” Gu et al., 2006, 2010; Logan and Duffy, 2006) and self-rotation (Takahashi et al., 2007) in re-

Received Feb. 3, 2016; revised June 7, 2016; accepted June 15, 2016.

Author contributions: M.B., N.D., and J.L.K. designed research; M.B. performed research; M.B., N.D., and J.L.K. analyzed data; M.B. and J.L.K. wrote the paper.

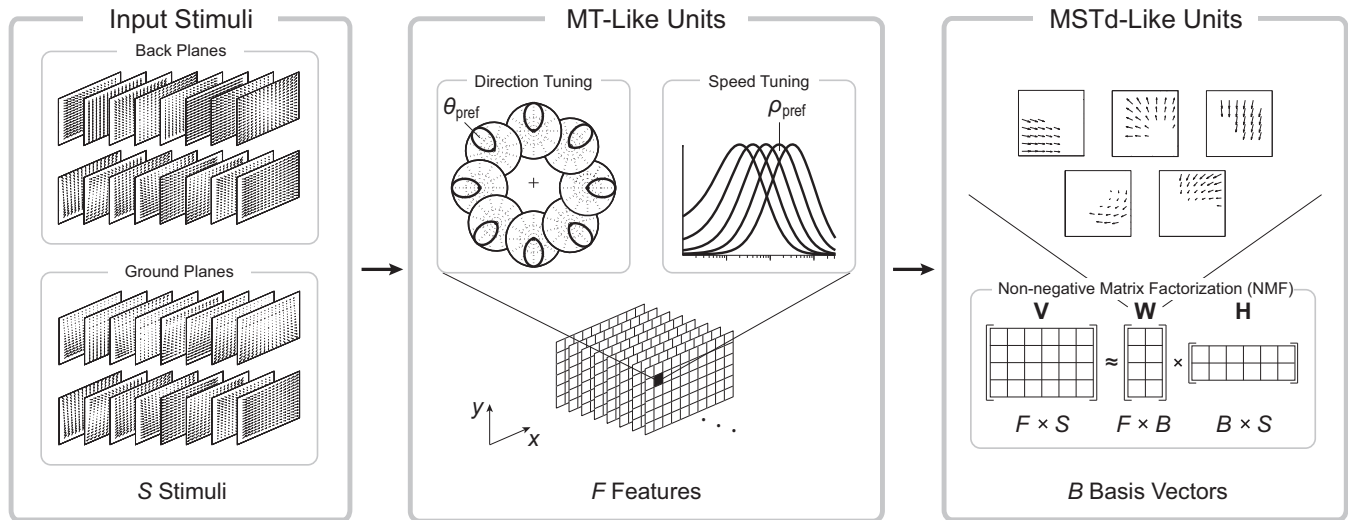
This research was supported by National Science Foundation Award IIS-1302125, Intel Corporation, and Northrop Grumman Aerospace Systems.

The authors declare no competing financial interests.

Correspondence should be addressed to Michael Beyeler, Department of Computer Science, University of California, Irvine, Irvine, CA 92697. E-mail: mbeyeler@uci.edu.

DOI:10.1523/JNEUROSCI.0396-16.2016

Copyright © 2016 the authors 0270-6474/16/360001-17\$15.00/0



**Figure 1.** Overall model architecture. A number  $S$  of 2D flow fields depicting observer translations and rotations in a 3D world were processed by an array of  $F$  MT-like motion sensors, each tuned to a specific direction and speed of motion. MT-like activity values were then arranged into the columns of a data matrix,  $\mathbf{V}$ , which served as input for NMF. The output of NMF was two reduced-rank matrices,  $\mathbf{W}$  (containing  $B$  non-negative basis vectors) and  $\mathbf{H}$  (containing hidden coefficients). Columns of  $\mathbf{W}$  (basis vectors) were then interpreted as weight vectors of MSTd-like model units.

sponse to both simulated and actual motion. In contrast to earlier studies that concentrated on a small number of MSTd neurons preferring fore–aft (i.e., forward and backward) motion directions (Duffy and Wurtz, 1995), these more recent studies demonstrated that heading preferences in MSTd were distributed throughout the spherical stimulus space, and that there was a significant predominance of cells preferring lateral as opposed to fore–aft headings. However, little is known about the underlying computational principles by which these response properties are derived.

In this article, we describe a computational model of MSTd based on the hypothesis that neurons in MSTd efficiently encode the continuum of large-field retinal flow patterns encountered during self-movement on the basis of inputs received from neurons in MT. Non-negative matrix factorization (NMF; Paatero and Tapper, 1994; Lee and Seung, 1999, 2001), a linear dimensionality reduction technique, is used to find a set of non-negative basis vectors, which are interpreted as MSTd-like receptive fields. NMF is similar to principal component analysis (PCA) and independent component analysis (ICA), but is unique among these dimensionality reduction techniques in that it can recover representations that are often sparse and “parts based,” much like the intuitive notion of combining parts to form a whole (Lee and Seung, 1999). We aim to show that this computational principle can account for a wide range of MSTd visual response properties, ranging from single-unit selectivity to population statistics (e.g., 3D translation and rotation selectivity, spiral tuning, and heading selectivity), and that seemingly contradictory findings in the literature can be resolved by eliminating differences in neuronal sampling procedures.

## Materials and Methods

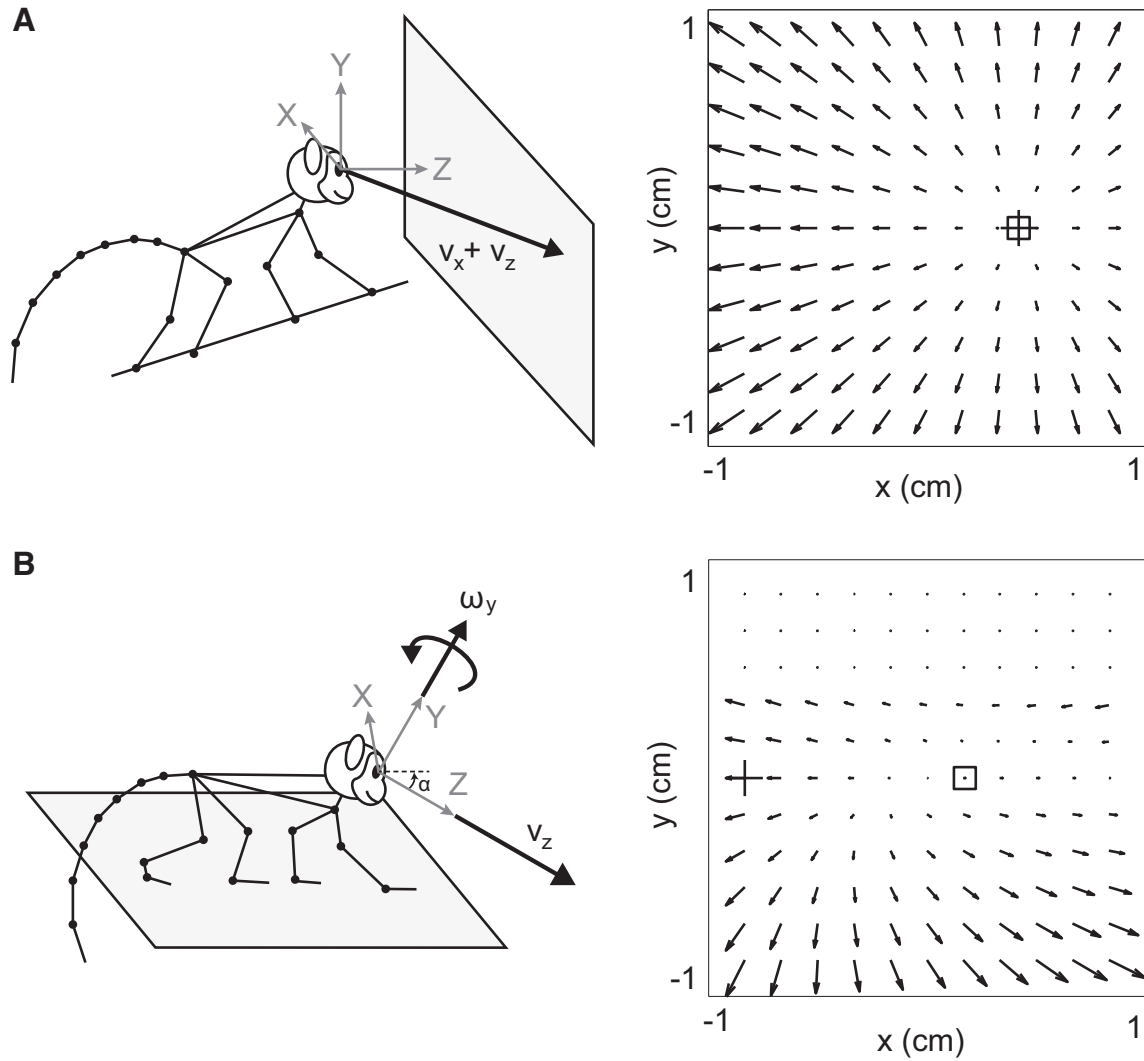
The overall architecture of the model is depicted in Figure 1. Visual input to the system encompassed a range of idealized two-dimensional (2D) flow fields caused by observer translations and rotations in a 3D world. We sampled flow fields that mimic natural viewing conditions during locomotion over ground planes and toward back planes located at various depths, with various linear and angular observer velocities, to yield a total of  $S$  flow fields, comprising the input stimuli. Each flow field was processed by an array of  $F$  MT-like motion sensors (MT-like model units), each tuned to a specific direction and speed of motion. The activ-

ity values of the MT-like model units were then arranged into the columns of an  $F \times S$  matrix,  $\mathbf{V}$ , which served as input for NMF (Paatero and Tapper, 1994; Lee and Seung, 1999, 2001). NMF belongs to a class of dimensionality reduction methods that can be used to linearly decompose a multivariate data matrix,  $\mathbf{V}$ , into an inner product of two reduced-rank matrices,  $\mathbf{W}$  and  $\mathbf{H}$ , such that  $\mathbf{V} \approx \mathbf{W}\mathbf{H}$ . The first of these reduced-rank matrices,  $\mathbf{W}$ , contains as its columns a total of  $B$  non-negative basis vectors of the decomposition. The second matrix,  $\mathbf{H}$ , contains as its rows the contribution of each basis vector in the input vectors (the hidden coefficients). These two matrices are found by iteratively reducing the residual between  $\mathbf{V}$  and  $\mathbf{W}\mathbf{H}$  using an alternating non-negative least-squares method. In our experiments, the only open parameter of the NMF algorithm was the number of basis vectors,  $B$ . We interpreted the columns of  $\mathbf{W}$  as the weight vectors of a total of  $B$  MSTd-like model units. Each weight vector had  $F$  elements representing the synaptic weights from the array of MT-like model units onto a particular MSTd-like model unit. The response of an MSTd-like model unit was given as the dot product of the  $F$  MT-like unit responses to a particular input stimulus and the corresponding non-negative synaptic weight vector,  $\mathbf{W}$ . Crucially, once the weight matrix  $\mathbf{W}$  was found, all parameter values remained fixed across all experiments. The following subsections explain the model in detail.

*Optic flow stimuli.* Input to the model was a computer-generated  $15 \times 15$  pixel array of simulated optic flow. We simulated the apparent motion on the retina that would be caused by an observer undergoing translations and rotations in a 3D world using the motion field model (Longuet-Higgins and Prazdny, 1980), where a pinhole camera with focal length,  $f$ , is used to project 3D real-world points,  $\vec{P} = [X, Y, Z]^t$ , onto a 2D image plane,  $\vec{p} = [x, y]^t = fZ[X, Y]^t$  (i.e., the retina). Local motion at a particular position  $\vec{P}$  on the image plane was specified by a vector,  $\dot{\vec{p}} = [\dot{x}, \dot{y}]^t$ , with local direction and speed of motion given as  $\tan^{-1}(\dot{y}/\dot{x})$  and  $\|\dot{\vec{x}}\|$ , respectively. The vector  $\dot{\vec{p}}$  was expressed by the sum of a translational flow component,  $\dot{\vec{x}}_T = [\dot{x}_T, \dot{y}_T]^t$ , and a rotational flow component,  $\dot{\vec{x}}_R = [\dot{x}_R, \dot{y}_R]^t$ :

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} \dot{x}_T \\ \dot{y}_T \end{bmatrix} + \begin{bmatrix} \dot{x}_R \\ \dot{y}_R \end{bmatrix} \quad (1)$$

where the translational component depended on the linear velocity of the camera,  $\vec{v} = [v_x, v_y, v_z]^t$ , and the rotational component depended on the angular velocity of the camera,  $\vec{\omega} = [\omega_x, \omega_y, \omega_z]^t$ :



**Figure 2.** Example flow fields generated with the motion field model (Longuet-Higgins and Prazdny, 1980; figure is a derivative of FigExamplesOfOpticFlow by Raudies (2013), used under CC BY). **A, B.** We sampled flow fields that mimic natural viewing conditions during upright locomotion toward a back plane (**A**) and over a ground plane (**B**), generated from a pinhole camera with image plane  $x, y \in \{-1 \text{ cm}, 1 \text{ cm}\}$  and focal length  $f = 1 \text{ cm}$ . Gray arrows indicate the axes of the 3D coordinate system, and bold black arrows indicate self-movement (translation, straight arrows; rotation, curved arrows). Crosses indicate the direction of self-movement (i.e., heading), and squares indicate the COM. In the absence of rotation, the COM indicates heading (**B**). **A.** Example of forward/sideward translation ( $v_x = 0.45 \text{ m/s}, v_z = 0.89 \text{ m/s}$ ) toward a back plane located at a distance of  $Z(x, y) = 10 \text{ m}$ . **B.** Example of curvilinear motion ( $v_x = v_z = 0.71 \text{ m/s}$ ) and yaw rotation ( $\omega_y = 3^\circ/\text{s}$ ) over a ground plane located at distance  $Z(y) = d/f(y \cos(\alpha) + f \sin(\alpha))$ , where  $d = -10 \text{ m}$  and  $\alpha = -30^\circ$ .

$$\begin{bmatrix} \dot{x}_T \\ \dot{y}_T \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} -f & 0 & x \\ 0 & -f & y \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} \dot{x}_R \\ \dot{y}_R \end{bmatrix} = \frac{1}{f} \begin{bmatrix} xy & -(f^2 + x^2) & fy \\ (f^2 + y^2) & -xy & -fx \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}. \quad (3)$$

In our simulations, we set  $f = 0.01 \text{ m}$  and  $x, y \in [-0.01 \text{ m}, 0.01 \text{ m}]$  (Raudies and Neumann, 2012). The  $15 \times 15$  optic flow array thus subtended  $90^\circ \times 90^\circ$  of the visual angle.

We sampled a total of  $S = 6000$  flow fields that mimic natural viewing conditions during locomotion over a ground plane (tilted  $\alpha = -30^\circ$  down from the horizontal) and toward a back plane (examples are shown in Fig. 2). Linear velocities corresponded to comfortable walking speeds  $\|\vec{v}\| = \{0.5, 1, 1.5\} \text{ m/s}$ , and angular velocities corresponded to common eye rotation velocities for gaze stabilization  $\|\vec{\omega}\| = \{0, \pm 5, \pm 10\} \text{ }^\circ/\text{s}$  (Perrone and Stone, 1994). Movement directions were uniformly sampled from all possible 3D directions (i.e., including backward translations). Back and ground planes were located at distances  $d = \{2, 4, 8, 16, 32\} \text{ m}$

from the observer. This interval of depths was exponentially sampled due to the reciprocal dependency between depth and the length of vectors of the translational visual motion field (Eq. 2; Perrone and Stone, 1994).

Note that  $\dot{x}_T$  depends on the distance to the point of interest ( $Z$ ; Eq. 2), but  $\dot{x}_R$  does not (Eq. 3). The point at which  $\dot{x}_T = 0$  is referred to as the epipole or center of motion (COM). If the optic flow stimulus is radially expanding, as is the case for translational forward motion, the COM is called the focus of expansion (FOE). In the absence of rotational flow, the FOE coincides with the direction of travel, or heading (Fig. 2A). However, in the presence of rotational flow, the FOE appears shifted with respect to the true direction of travel (Fig. 2B), and this discrepancy increases with increased eye rotation velocity.

**MT stage.** Each virtual flow field was processed by an array of idealized MT-like model units, each selective to a particular direction of motion,  $\theta_{\text{pref}}$  and a particular speed of motion,  $\rho_{\text{pref}}$  at a particular spatial location,  $(x, y)$ . The activity of each unit,  $r_{\text{MT}}$ , was given as follows:

$$r_{\text{MT}}(x, y; \theta_{\text{pref}}, \rho_{\text{pref}}) = d_{\text{MT}}(x, y; \theta_{\text{pref}}) s_{\text{MT}}(x, y; \rho_{\text{pref}}), \quad (4)$$

where  $d_{\text{MT}}$  was the direction response of the unit and  $s_{\text{MT}}$  was the speed response of the unit.

The direction tuning of a unit was given as a von Mises function based on the difference between the local direction of motion at a particular spatial location,  $\theta(x, y)$ , and the preferred direction of motion of the unit,  $\theta_{\text{pref}}$  (Mardia, 1972; Swindale, 1998; Jazayeri and Movshon, 2006):

$$d_{\text{MT}}(x, y; \theta_{\text{pref}}) = \exp(\sigma_{\theta}(\cos(\theta(x, y) - \theta_{\text{pref}}) - 1)) \quad (5)$$

where the bandwidth parameter was  $\sigma_{\theta} = 3$ , so that the resulting tuning width (full-width at half-maximum) was approximately  $90^{\circ}$ , consistent with reported values in the literature (Britten and Newsome, 1998).

The speed tuning of a unit was given as a log-Gaussian function of the local speed of motion,  $\rho(x, y)$ , relative to the preferred speed of motion of the unit,  $\rho_{\text{pref}}$  (Nover et al., 2005):

$$s_{\text{MT}}(x, y; \rho_{\text{pref}}) = \exp\left(-\frac{\log^2\left(\frac{\rho(x, y) + s_0}{\rho_{\text{pref}} + s_0}\right)}{2\sigma_{\rho}^2}\right), \quad (6)$$

where the bandwidth parameter was  $\sigma_{\rho} = 1.16$ , and the speed offset parameter was  $s_0 = 0.33$ , both of which corresponded to the medians of physiological recordings (Nover et al., 2005). Note that the offset parameter,  $s_0$ , was necessary to keep the logarithm from becoming undefined as stimulus speed approached zero.

As a result, the population prediction of speed discrimination thresholds obeyed Weber's law for speeds larger than  $\sim 5^{\circ}/s$  (Nover et al., 2005). We chose five octave-spaced bins and a uniform distribution between 0.5 and  $32^{\circ}/s$  (Nover et al., 2005); that is,  $\rho_{\text{pref}} = \{2, 4, 8, 16, 32\}^{\circ}/s$ .

At each spatial location, there were a total of 40 MT-like model units (selective for eight directions times five speeds of motion), yielding a total of  $F = 15 \times 15 \times 8 \times 5 = 9000$  units. The activity pattern of these 40 units/pixel thus acted as a population code for the local direction and speed of motion. We assumed that the receptive fields of all MT-like model units had a single pixel radius (subtending  $\sim 3^{\circ}$  of visual angle), which is comparable to receptive field sizes near the fovea as found in macaque MT (Raiguel et al., 1995).

**Non-negative matrix factorization.** We hypothesized that appropriate synaptic weights of the feedforward projections from MT to MSTd could be learned with NMF (Paatero and Tapper, 1994; Lee and Seung, 1999, 2001). NMF belongs to a class of methods that can be used to decompose multivariate data into an inner product of two reduced-rank matrices, where one matrix contains non-negative basis vectors and the other contains non-negative activation vectors (hidden coefficients). The non-negativity constraints of NMF enforce the combination of different basis vectors to be additive, leading to representations that are often parts based and sparse (Lee and Seung, 1999). For example, when applied to images of faces, NMF recovers basis images that resemble facial features, such as eyes, mouths, and noses. When applied to neural networks, these non-negativity constraints correspond to the notion that neuronal firing rates are never negative and that synaptic weights are either excitatory or inhibitory, but they do not change sign (Lee and Seung, 1999).

Assume that we observe data in the form of a large number of independent and identically distributed random vectors,  $\vec{v}^{(i)}$ , such as the neuronal activity of a population of MT neurons in response to a visual stimulus, where  $i$  is the sample index. When these vectors are arranged into the columns of a matrix,  $\mathbf{V}$ , linear decompositions describe these data as  $\mathbf{V} \approx \mathbf{WH}$ , where  $\mathbf{W}$  is a matrix that contains as its columns the basis vectors of the decomposition. The rows of  $\mathbf{H}$  contain the corresponding hidden coefficients that give the contribution of each basis vector in the input vectors. Like PCA, the goal of NMF is then to find a linear decomposition of the data matrix  $\mathbf{V}$ , with the additional constraint that all elements of the matrices  $\mathbf{W}$  and  $\mathbf{H}$  be non-negative. In contrast to ICA, NMF does not make any assumptions about the statistical dependencies of  $\mathbf{W}$  and  $\mathbf{H}$ . The resulting decomposition is not exact;  $\mathbf{WH}$  is a lower-rank approximation of  $\mathbf{V}$ , and the difference between  $\mathbf{WH}$  and  $\mathbf{V}$  is termed the reconstruction error. Perfect accuracy is only possible when the number of basis vectors approaches infinity, but good approximations can usually be obtained with a reasonably small number of basis vectors (Pouget and Sejnowski, 1997).

We used the standard `nnmf` function provided by MATLAB R2014a (MathWorks), which, using an alternating least-squares algorithm, aims to iteratively minimize the root mean square residual  $D$  between  $\mathbf{V}$  and  $\mathbf{WH}$ , as follows:

$$D = \frac{\|\mathbf{V} - \mathbf{WH}\|}{FS}, \quad (7)$$

where  $F$  is the number of rows in  $\mathbf{W}$  and  $S$  is the number of columns in  $\mathbf{H}$  (Fig. 1).  $\mathbf{W}$  and  $\mathbf{H}$  were normalized so that the rows of  $\mathbf{H}$  had unit length. The output of NMF is not unique because of random initial conditions (i.e., random weight initialization).

The only open parameter was the number of basis vectors,  $B$ , whose value had to be determined empirically. In our simulations, we examined a range of values ( $B = 2^i$ , where  $i = \{4, 5, 6, 7, 8\}$ ; see Efficient encoding of multiple perceptual variables), but found that  $B = 64$  led to basis vectors that most resembled receptive fields found in macaque MSTd.

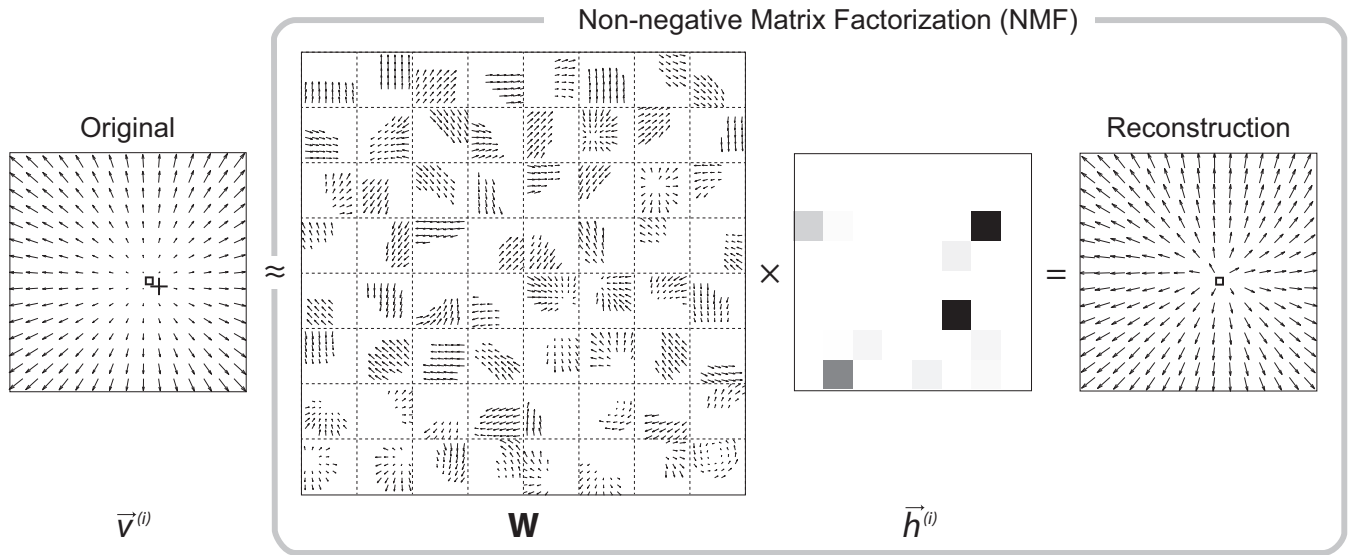
To get a better intuitive understanding of the decomposition process, an example reconstruction with  $B = 64$  is shown in Figure 3. The database of  $S = 6000$  virtual flow fields generated above (see Optic flow stimuli) was parsed by an array of  $F = 15 \times 15 \times 8 \times 5 = 9000$  idealized MT-like model (see MT stage), yielding an  $F \times S$  data matrix,  $\mathbf{V}$ . As illustrated in Figure 1 above, NMF then found a factorization of the form  $\mathbf{V} \approx \mathbf{WH}$ , where the columns of  $\mathbf{W}$  corresponded to  $B = 64$  basis vectors, and the columns of  $\mathbf{H}$  contained 64 corresponding hidden coefficients. By calculating the population vector for the direction (vector orientation) and speed of motion vector magnitude) at each of the  $15 \times 15$  pixel locations (Georgopoulos et al., 1982), the 64 columns of  $\mathbf{W}$  could themselves be represented by a set of basis flow fields, shown in a  $8 \times 8$  montage in Figure 3. These basis flow fields, each of them limited to a subregion of the visual field and preferring a mixture of translational and rotational flow components, can be thought of as an efficient "toolbox" from which to generate more complex flow fields, similar to the set of basis face images described by Lee and Seung (1999), their Figure 1. A particular instance of an input stimulus (corresponding to the  $i$ th column of  $\mathbf{V}$ ,  $\vec{v}^{(i)}$ ), shown on the left in Figure 3, could then be approximated by a linear superposition of basis flow fields. The coefficients of the linear superposition (corresponding to the  $i$ th column of  $\mathbf{H}$ ,  $\vec{h}^{(i)}$ ) were arranged into an  $8 \times 8$  matrix, shown next to  $\mathbf{W}$ , where darker colors corresponded to higher activation values. The reconstructed stimulus, given by the product  $\mathbf{W}\vec{h}^{(i)}$ , can again be visualized as a flow field via population vector decoding (shown on the right in Fig. 3). The reconstruction error was 0.0824 for this particular stimulus and  $D = 0.143$  across all stimuli (see Eq. 7). This particular input stimulus was made of both translational and rotational components ( $\vec{v} = [0.231, -0.170, 0.958]^t$  m/s and  $\vec{\omega} = [0.481, 0.117, 9.99]^t$   $^{\circ}/s$ ; see Eqs. 2 and 3) toward a back plane located at  $d = 2$  m from the observer, with heading indicated by a small cross and the COM indicated by a small square.

To facilitate statistical comparisons between model responses and biological data, NMF with  $B = 64$  was repeated 14 times with different random initial conditions to generate a total of  $N = 896$  MSTd-like model units. These units were generated only once, after which the model remained fixed across all subsequent experiments and analyses. If an experiment required a different number of model units, these units were sampled randomly from the original population of  $N = 896$  model units.

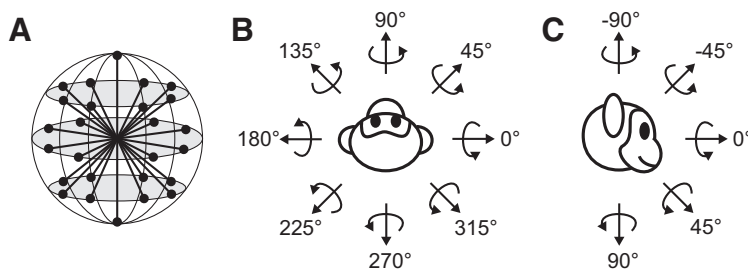
**MSTd stage.** We interpreted the resulting columns of  $\mathbf{W}$  as the weight vectors from the population of  $F$  MT-like model units onto a population of  $B$  MSTd-like model units. The activity of the  $b$ th MSTd-like unit,  $r_{\text{MSTd}}^b$ , was given as the dot product of all  $F$  MT-like responses to a particular input stimulus,  $i$ , and the corresponding non-negative weight vector of the unit, as follows:

$$r_{\text{MSTd}}^b(i) = \vec{v}^{(i)} \cdot \vec{w}^{(b)}, \quad (8)$$

where  $\vec{v}^{(i)}$  was the  $i$ th column of  $\mathbf{V}$ , and  $\vec{w}^{(b)}$  was the  $b$ th column of  $\mathbf{W}$ . This is in agreement with the finding that MSTd responses are approximately linear in terms of their feedforward input from area MT (Tanaka et al., 1989b; Duffy and Wurtz, 1991b), which provides one of the strongest projections to MST in the macaque (Boussaoud et al., 1990). In contrast to other models (Lappe et al., 1996; Zemel and Sejnowski, 1998;



**Figure 3.** Applying NMF to MT-like patterns of activity led to a sparse parts-based representation of retinal flow, similar to the parts-based representation of faces in Lee and Seung (1999), their Figure 1. Using NMF, a particular instance of an input stimulus (corresponding to the  $i$ th column of data matrix  $\mathbf{V}$ ), shown on the left, can be approximated by a linear superposition of basis vectors (i.e., the columns of  $\mathbf{W}$ ), here visualized as basis flow fields using population vector decoding (Georgopoulos et al., 1982) in a  $8 \times 8$  montage. The coefficients of the linear superposition (corresponding to the  $i$ th column of  $\mathbf{H}$ ) are shown next to  $\mathbf{W}$  as an  $8 \times 8$  matrix, where darker colors correspond to higher activation values. The reconstructed stimulus (with a reconstruction error of 0.0824), is shown on the right. The input stimulus was made of both translational and rotational components ( $\vec{v} = [0.231, -0.170, 0.958]^T$  m/s and  $\vec{\omega} = [0.481, 0.117, 9.99]^T$  °/s) toward a back plane located at  $d = 2$  m from the observer, with heading indicated by a small cross, and the COM indicated by a small square.



**Figure 4.** Schematic of the 26 translational and rotational directions used to test MSTd-like model units (modified with permission from Takahashi et al., 2007, their Figure 2). **A**, Illustration of the 26 movement vectors, spaced  $45^\circ$  apart on the sphere, in both azimuth and elevation. **B**, Top view: definition of azimuth angles. **C**, Side view: definition of elevation angles. Straight arrows illustrate the direction of movement in the translation protocol, whereas curved arrows indicate the direction of rotation (according to the right-hand rule) about each of the movement vectors.

Grossberg et al., 1999; Mineault et al., 2012), no additional nonlinearities were required to fit the data presented in this article.

**3D translation/rotation protocol.** To determine 3D translation and rotation tuning profiles, the MSTd-like model units generated above were probed with two sets of virtual optic flow stimuli, as described by Takahashi et al. (2007).

The “translation protocol” consisted of straight translational movements along 26 directions in 3D space, corresponding to all combinations of azimuth and elevation angles in increments of  $45^\circ$  (Fig. 4A). This included all combinations of movement vectors having eight different azimuth angles ( $0, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$  and  $315^\circ$ ; Fig. 4B), each of which was presented at three different elevation angles (Fig. 4C):  $0^\circ$  (the horizontal plane) and  $\pm 45^\circ$  (for a sum of  $8 \times 3 = 24$  directions). Two additional directions were specified by elevation angles of  $-90^\circ$  and  $+90^\circ$ , which corresponded to upward and downward movement directions, respectively.

The “rotation protocol” consisted of rotations about the same 26 directions, which now represented the corresponding axes of rotation according to the right-hand rule (Fig. 4B, C). For example, azimuths of  $0$  and  $180^\circ$  (elevation,  $0$ ) corresponded to pitch-up and pitch-down rotations, respectively. Azimuths of  $90^\circ$  and  $270^\circ$  (elevation,  $0$ ) corresponded to roll rotations (right-ear down and left-ear down, respectively). Finally,

elevation angles of  $-90^\circ$  and  $+90^\circ$  corresponded to leftward or rightward yaw rotation, respectively.

**Heading tuning index.** To quantify the strength of heading tuning of a model unit, we followed a procedure described by Gu et al. (2006) to compute a heading tuning index (HTI). In each trial, the activity of a model unit was considered to represent the magnitude of a 3D vector whose direction was defined by the azimuth and elevation angles of the respective movement trajectory (Gu et al., 2006). An HTI was then computed for each model unit using the following equation:

$$HTI^b = \frac{|\sum_{i=1}^n r_{MSTd}^b(i) \vec{e}_i|}{\sum_{i=1}^n |r_{MSTd}^b(i) \vec{e}_i|}, \quad (9)$$

where  $r_{MSTd}^b(i)$  was the activity of the  $b$ th model unit for the  $i$ th stimulus,  $\vec{e}_i$  was a unit vector indicating the 3D Cartesian heading direction of the  $i$ th stimulus,  $|\cdot|$  denoted the vector magnitude, and  $n = 26$  corresponded to the number of different heading directions tested (see 3D translation/rotation protocol). The HTI ranged from 0 to 1, where 0 indicated weak direction tuning and 1 indicated strong direction tuning. The preferred heading direction for each stimulus was then computed from the azimuth and elevation of the vector sum of the individual responses (numerator in Eq. 9).

**Uniformity test.** To determine whether a measured distribution was significantly different from uniform, we performed a resampling analysis as described by Takahashi et al. (2007). First, we calculated the sum-squared error (across bins) between the measured distribution and an ideal uniform distribution containing the same number of observations. This calculation was repeated for 1000 different random distributions generated using the `unifrnd` function provided by MATLAB R2014a (MathWorks) to produce a distribution of sum-squared error values that represent random deviations from an ideal uniform distribution. If the sum-squared error for the experimentally measured distribution lay outside the 95% confidence interval of values from the randomized distributions, the measured distribution was considered to be significantly different from uniform ( $p < 0.05$ ; Takahashi et al., 2007).

**Decoding population activity.** We tested whether perceptual variables, such as heading or angular velocity, could be decoded from the population activity of MSTd-like model units generated above using simple linear regression. We assembled a dataset consisting of  $10^4$  flow fields with randomly selected headings, which depicted linear observer movement (velocities sampled uniformly between 0.5 and 2 m/s; no eye rotations) toward a back plane located at various distances  $d = \{2, 4, 8, 16, 32\}$  m away. As part of a 10-fold cross-validation procedure, stimuli were split repeatedly into a training set containing 9000 stimuli and a test set containing 1000 stimuli. Using linear regression, we then obtained a set of  $N \times 2$  linear weights used to decode MSTd population activity in response to samples from the training set and monitored performance on the test set (Picard and Cook, 1984).

For forward headings and in the absence of eye rotations, heading coincides with the location of the FOE on the retina (see Optic flow stimuli). Conceptually, the 2D Cartesian coordinates of the FOE can be found by setting  $\dot{x} = 0$  in Equation 2, and solving for  $(x, y)$ . However, because these coordinates could approach infinity for lateral headings, we only considered cases of headings (in spherical coordinates) with azimuth angles between  $45^\circ$  and  $135^\circ$ , as well as elevation angles between  $-45^\circ$  and  $+45^\circ$ .

Similarly, we obtained a set of  $N \times 2$  linear weights to predict the 2D Cartesian components of angular velocity,  $\dot{x}_R$ . In this case, the dataset contained stimuli consisting only of rotational flow components (i.e.,  $\dot{x}_T = 0$  in Eq. 1). To compare model performance to neurophysiological studies (Ben Hamed et al., 2003), we randomly selected  $N = 144$  model units from the population and limited rotations to the  $x$ - $z$ -plane (i.e.,  $\omega_y = 0$  in Eq. 3).

**Sparseness.** We computed a sparseness metric for the modeled MSTd population activity according to the definition of sparseness by Vinje and Gallant (2000), as follows:

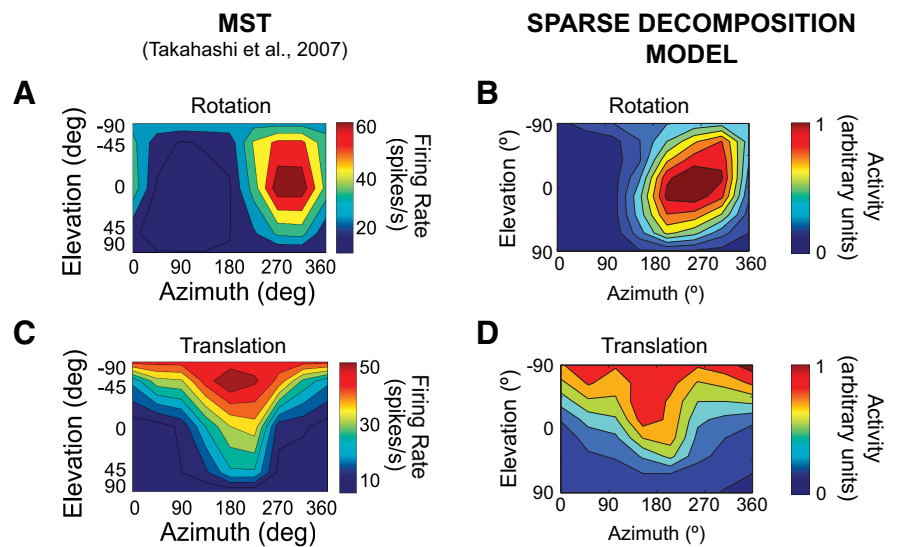
$$s = \left( 1 - \frac{1}{N} \frac{(\sum_i r_i)^2}{\sum_i r_i^2} \right) / \left( 1 - \frac{1}{N} \right). \quad (10)$$

Here,  $s \in [0, 1]$  is a measure of sparseness for a signal  $r$  with  $N$  sample points, where  $s = 1$  denotes maximum sparseness and is indicative of a local code, and  $s = 0$  is indicative of a dense code. To measure how many MSTd-like model units were activated by any given stimulus (population sparseness),  $r_i$  was the response of the  $i$ th neuron to a particular stimulus, and  $N$  was the number of model units. To determine how many stimuli any given model unit responded to (lifetime sparseness),  $r_i$  was the response of a unit to the  $i$ th stimulus, and  $N$  was the number of stimuli. Population sparseness was averaged across stimuli, and lifetime sparseness was averaged across units.

**Fisher information analysis.** To investigate whether simulated MSTd population activity could account for the dependence of psychophysical thresholds on reference heading, we followed a procedure described by Gu et al. (2010) to compute Fisher information,  $I_F$ , which provides an upper limit on the precision with which an unbiased estimator can discriminate small variations in a variable ( $x$ ) around a reference value ( $x_{\text{ref}}$ ), as follows:

$$I_F = \sum_{i=1}^N \frac{R'_i(x_{\text{ref}})^2}{\sigma_i(x_{\text{ref}})^2}, \quad (11)$$

where  $N$  denotes the number of neurons in the population,  $R'_i(x_{\text{ref}})$  denotes the derivative of the tuning curve for the  $i$ th neuron at  $x_{\text{ref}}$ , and



**Figure 5.** A–D, Example of 3D direction tuning for an MSTd neuron (rotation, A; translation, C), reprinted with permission from Takahashi et al. (2007), their Fig. 4, and a similarly tuned MSTd-like model unit (rotation, B; translation, D). Color contour maps show the mean firing rate or model activation as a function of azimuth and elevation angles. Each contour map shows the Lambert cylindrical equal-area projection of the original data, where the abscissa corresponds to the azimuth angle and the ordinate is a sinusoidally transformed version of elevation angle (Snyder, 1987).

$\sigma_i(x_{\text{ref}})$  is the variance of the response of the  $i$ th neuron at  $x_{\text{ref}}$ . Neurons with steeply sloped tuning curves and small variance will contribute most to Fisher information. In contrast, neurons having the peak of their tuning curve at  $x_{\text{ref}}$  will contribute little.

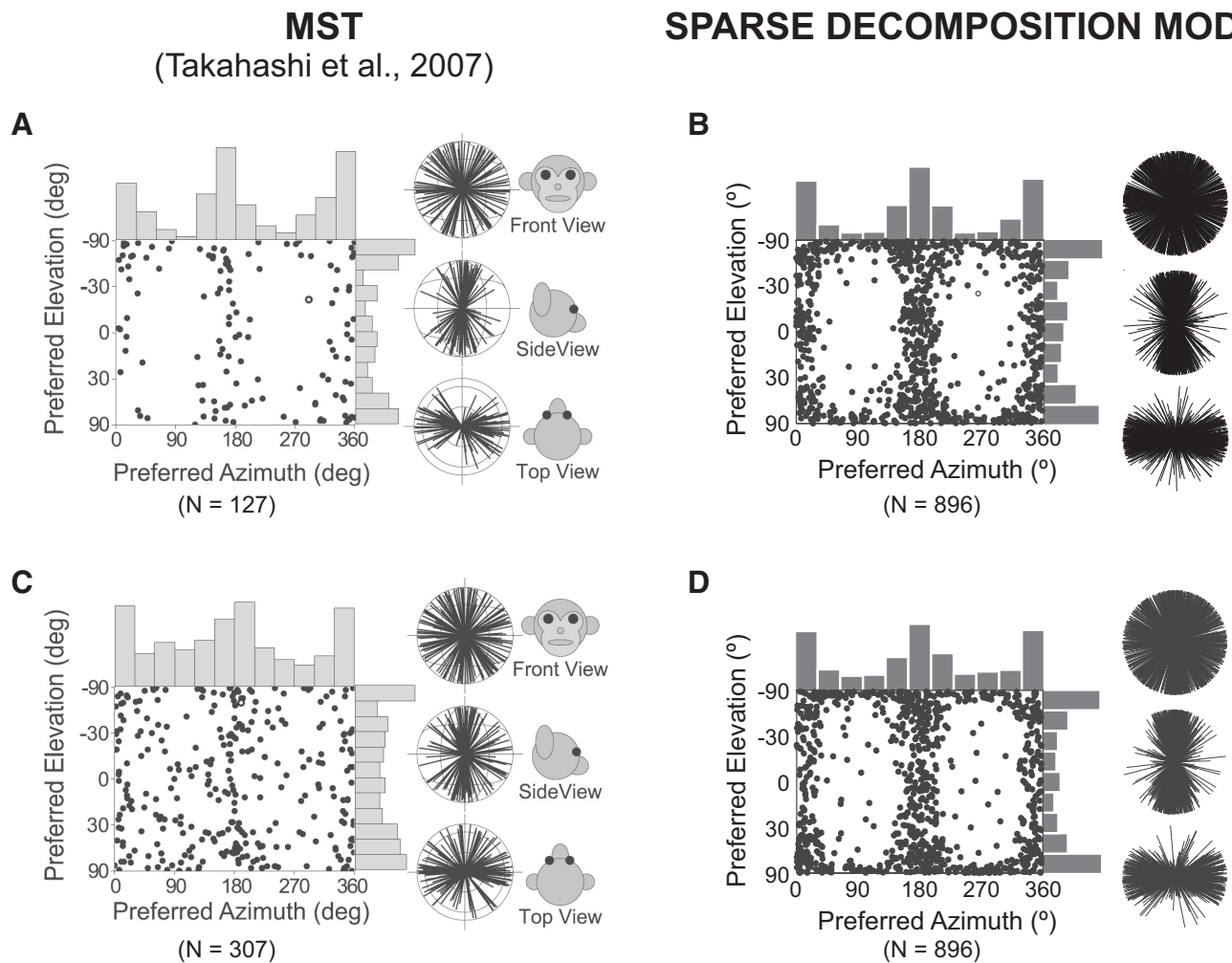
To calculate the tuning curve slope,  $R'_i(x_{\text{ref}})$ , we used a spline function ( $0.01^\circ$  resolution) to interpolate among coarsely sampled data points ( $30^\circ$  spacing). To calculate variance,  $\sigma_i(x_{\text{ref}})$ , Gu et al. (2010) assumed that neurons had independent noise and Poisson spiking statistics. Because the model units described in this article are deterministic, we instead calculated signal variance directly from the response variability of the network when presented with random dot clouds generated from a particular 3D heading (150 repetitions).

## Results

### 3D translation and rotation selectivity

We tested whether individual units in our model of MSTd could signal the 3D direction of self-translation and self-rotation when presented with large-field optic flow stimuli, and compared our results to neurophysiological data from Gu et al. (2006) and Takahashi et al. (2007). Visual stimuli depicted translations and rotations of the observer through a 3D cloud of dots that occupied a space 40 cm deep and subtended  $90^\circ \times 90^\circ$  of visual angle (see 3D translation/rotation protocol). In the translation protocol, movement trajectories were along 26 directions or headings (Fig. 4A), corresponding to all combinations of azimuth and elevation angles in increments of  $45^\circ$  (Fig. 4B, C, straight arrows). In the rotation protocol, these 26 directions served as rotation axes instead (Fig. 4B, C, curved arrows). Linear velocity was 1 m/s, and angular velocity was  $20^\circ/\text{s}$  (Takahashi et al., 2007).

Figure 5 shows the 3D translation and rotation tuning of a particular neuron in macaque MSTd (Fig. 5A, C; Takahashi et al., 2007) and of a similarly tuned MSTd-like model unit (Fig. 5B, D). The 3D directional tuning profile is shown as a contour map in which activity (mean firing rate for neuron in macaque MSTd; unit activation for MSTd-like model unit, see Eq. 8), here represented by color, is plotted as a function of azimuth (abscissa) and elevation (ordinate). Each contour map shows the Lambert cylindrical equal area projection of the original spherical data, where the abscissa corresponds to the azimuth angle and the



**Figure 6.** *A–D*, Distribution of 3D direction preferences of MSTd neurons (rotation, *A*; translation, *C*), reprinted with permission from Takahashi et al. (2007), their Figure 6, and the population of MSTd-like model units (rotation, *B*; translation, *D*). Each data point in the scatter plot corresponds to the preferred azimuth (abscissa) and elevation (ordinate) of a single neuron or model unit. Histograms along the top and right sides of each scatter plot show the marginal distributions. Also shown are 2D projections (front view, side view, and top view) of unit-length 3D preferred direction vectors (each radial line represents one neuron or model unit). The neuron in Figure 5 is represented as open circles in each panel.

ordinate is a sinusoidally transformed version of elevation angle (Snyder, 1987). The peak response of this particular MSTd neuron occurred at 291° azimuth and  $-18^\circ$  elevation in the case of rotation (corresponding approximately to a left ear-down roll rotation, with small components of pitch and yaw rotation), and at 190° azimuth and  $-50^\circ$  elevation in the case of translation (corresponding to backward motion; Takahashi et al., 2007). The corresponding MSTd-like model unit (Fig. 5*B, D*) was selected by finding the unit whose location of the peak response (azimuth and elevation) was closest to the neuron mentioned above (Fig. 5*A, C*). This unit had its peak response at 268° azimuth,  $-21^\circ$  elevation, 176° azimuth, and  $-41^\circ$  elevation for rotation and translation, respectively. The shape of the tuning curve was typical among all MSTd-like model units, which was generally in good agreement with neurophysiological data (Gu et al., 2006; Takahashi et al., 2007). Because similar 2D flow patterns evoke similar unit responses (see Eq. 8), peak responses for translations and rotations occurred at stimulus directions approximately 90° apart on the sphere.

Figure 6 shows the distribution of direction preferences of MSTd cells (Fig. 6*A, C*; Gu et al., 2006; Takahashi et al., 2007) and MSTd-like model units (Fig. 6*B, D*) for visual translation and rotation. Each data point in these scatter plots specifies the pre-

**Table 1. Percentage of neurons and model units with preferred rotation directions within  $\pm 30^\circ$  of each of the cardinal axes**

Visual rotation	Yaw	Pitch	Roll
Takahashi et al., 2007	36/127 (28%)	27/127 (21%)	1/127 (1%)
This article	216/896 (24%)	330/896 (37%)	4/896 (1%)

ferred 3D direction of a single neuron or model unit. Histograms along the boundaries show the marginal distributions of azimuth and elevation preferences. The direction preference (for rotation and translation, separately) was defined by the azimuth and elevation of the vector average of the neural responses (see 3D translation/rotation protocol). The distributions of azimuth and elevation preference were significantly nonuniform for both rotation and translation conditions ( $p < 0.05$ ; see Uniformity test), tightly clustered around 0 and 180° azimuth as well as  $-90^\circ$  and  $+90^\circ$  elevation. In agreement with macaque MSTd, the majority of all 896 MSTd-like model units had rotation preferences within 30° of the yaw or pitch axes, and only a few model units had their rotation preference within 30° of the roll axis (Table 1). For translation, the majority of direction preferences were within 30° of the lateral or vertical axes, with only a handful of fore–aft direction preferences (Table 2).

**Table 2. Percentage of neurons and model units with preferred translation directions within  $\pm 30^\circ$  of each of the cardinal axes**

Visual translation	Lateral	Fore–aft	Vertical
Takahashi et al., 2007	57/307 (19%)	20/307 (7%)	76/307 (25%)
This paper	245/896 (27%)	5/896 (1%)	192/896 (21%)

We also quantified the strength of heading tuning in model MSTd using an HTI (Gu et al., 2006), which ranged from 0 to 1, indicating poor and strong tuning, respectively (see Heading tuning index). For reference, a model unit with idealized cosine tuning would have an HTI value of 0.31, whereas an HTI value of 1 would be reached when the firing rate is zero for all but a single stimulus direction. In the translation condition, Gu et al. (2006) reported HTI values averaging  $0.48 \pm 0.16$  SD for their sample of 251 MSTd cells. Model MSTd achieved very similar HTI values, averaging  $0.43 \pm 0.11$  SD and  $0.47 \pm 0.11$  SD, respectively, in the translation and rotation conditions.

In addition, Takahashi et al. (2007) tested a subset of MSTd cells with both the rotation and translation protocols to directly compare the difference in 3D direction preference ( $|\Delta$  preferred direction  $|\Delta$ ) between rotation and translation (Fig. 7A). The distribution was strongly nonuniform with a mode near  $90^\circ$ . The simulated MSTd units appropriately captured this distribution (Fig. 7B). However, because  $|\Delta$  preferred direction  $|\Delta$  is computed as the smallest angle between a pair of preferred direction vectors in three dimensions, it is only defined between 0 and  $180^\circ$ . Therefore, the observed peak near  $90^\circ$  in Figure 7, A and B, could be derived from a single mode at  $90^\circ$  or from two modes at  $\pm 90^\circ$ . Only the former, but not the latter, would indicate that the visual preferences for translation and rotation are linked via the 2D visual motion selectivity of the cell or model unit (Takahashi et al., 2007). Therefore, we also illustrate the differences in 2D direction preferences by projecting each 3D preference vector onto the following three cardinal planes:  $x$ – $y$  (front view),  $y$ – $z$  (side view), and  $x$ – $z$  (top view) for both MSTd (Fig. 7C) and model MSTd (Fig. 7D). Since some planar (P) projections might be small in amplitude, we also calculated the ratio of the lengths of each difference vector in two dimensions and three dimensions, and plotted them against the 2D preferred direction difference (Fig. 7E,F). Consistent with macaque MSTd, the projections of the visual difference vector onto the  $x$ – $z$ - and  $y$ – $z$ -planes were relatively small (Fig. 7E,F, green and blue data points). In contrast, projections onto the  $x$ – $y$ -plane were notably larger (Fig. 7E,F, red data points) and tightly centered at  $-90^\circ$  (Fig. 7C,D), with no cells or model units having direction differences of  $+90^\circ$  between visual translation and rotation. Thus, these simulated data are consistent with the idea that preferred directions for translation and rotation are related through the 2D visual motion selectivity of MSTd neurons, implying that neither model nor macaque MSTd are well equipped to determine the underlying cause of an optic flow pattern (i.e., observer translation vs rotation) on a single-unit level.

### Efficient encoding of multiple perceptual variables

MSTd activity plays a causal role in the perception of heading from optic flow (Gu et al., 2010, 2012). During forward movement, retinal flow radiates out symmetrically from a single point, the FOE, from which heading can be inferred (Gibson, 1950). Page and Duffy (1999) found that the location of the FOE could be decoded to a very high degree of precision (i.e., within  $\pm 10^\circ$ ) from the trial-averaged population response of neurons in MSTd. Ben Hamed et al. (2003) then went on to show that the

FOE in both eye-centered and head-centered coordinates could be decoded from MSTd population activity with an optimal linear estimator even on a single-trial basis, with an error of  $0.5$ – $1.5^\circ$  and an SD of  $2.4$ – $3^\circ$ . Interestingly, they found that other perceptual variables, such as eye position and the direction of ocular pursuit, could be decoded with similar error rates, and that most MSTd neurons were involved in the simultaneous encoding of more than one of these variables (Ben Hamed et al., 2003).

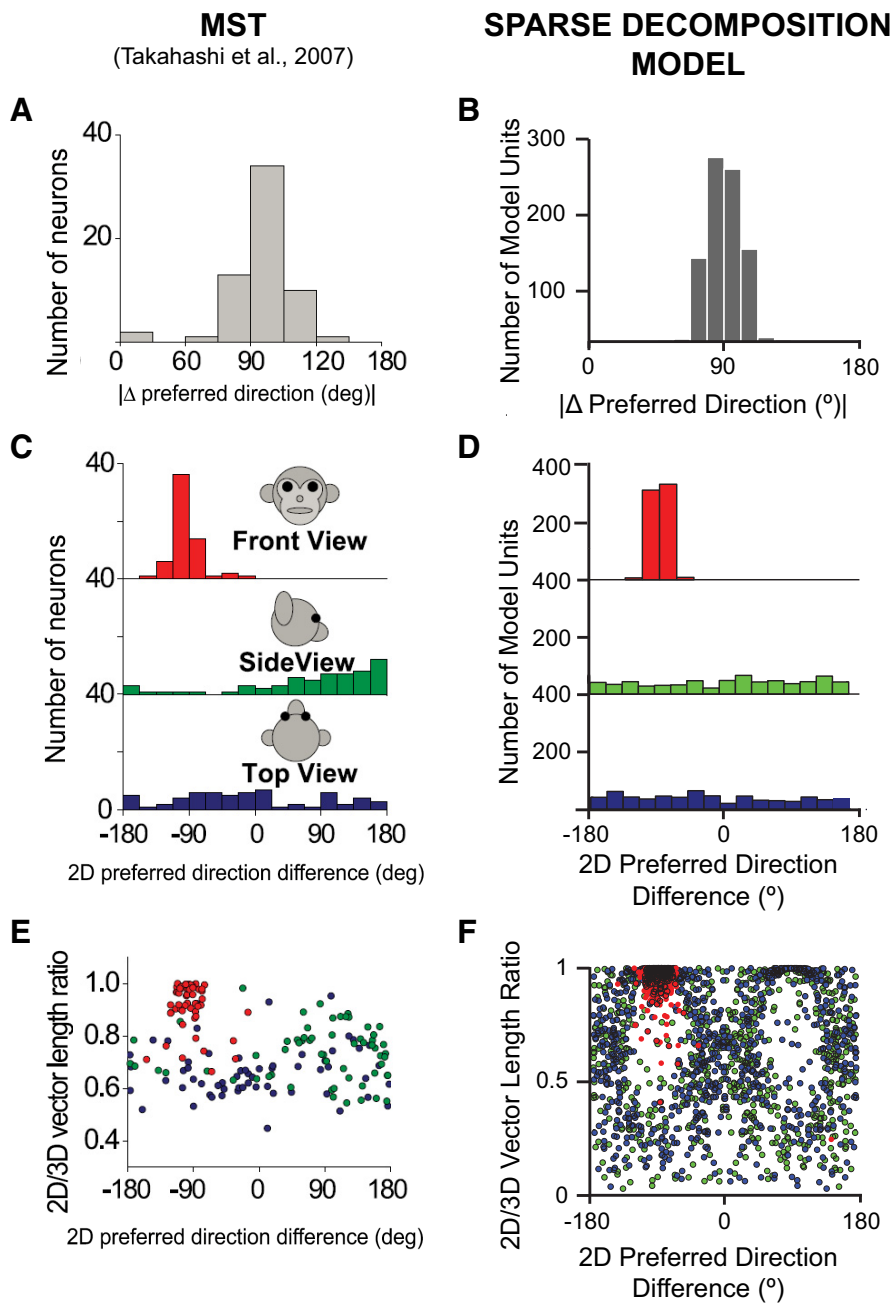
We therefore wondered whether the 2D coordinates ( $x_{\text{FOE}}$ ,  $y_{\text{FOE}}$ ) of the FOE of arbitrary expansive flow fields could be decoded from a population of  $N$  MSTd-like model units with similar precision. To answer this question, we assembled a dataset of  $10^4$  flow fields with randomly selected headings (azimuth between  $45^\circ$  and  $135^\circ$ , elevation between  $-45^\circ$  and  $+45^\circ$ ), depicting linear observer movements toward a back plane located at various distances,  $d = \{1, 2, 4, 8\}$ , in front of the observer. Using simple linear regression in a cross-validation procedure (see Decoding population activity), we obtained a set of  $N \times 2$  linear weights used to decode MSTd population activity in response to samples from the training set. We then tried to predict heading in flow samples from the test set (i.e., headings the network had not seen before) using the learned weights, and assessed prediction error rates. The same procedure was repeated for a set of rotational flow fields designed to mimic visual consequences of slow, pursuit-like eye movements (i.e., restricted to the frontoparallel plane,  $\omega_y = 0$ ). Here, the goal of the network was to predict the 2D Cartesian components of angular velocity ( $\omega_x$ ,  $\omega_z$ ).

The results are summarized in Table 3. Both heading and eye velocity could be inferred from the population activity of model MSTd ( $N = 144$ ) with error rates on the order of  $0.1^\circ$ – $1^\circ$ , consistent with neurophysiological data (Ben Hamed et al., 2003; Page and Duffy, 2003). Note that these numbers were achieved on the test set; that is, on flow fields the model had never seen before. Heading prediction error on the test set thus corresponded to the ability of the model to generalize; that is, to deduce the appropriate response to a novel stimulus using what it had learned from other previously encountered stimuli (Hastie et al., 2009; Spanne and Jörnstell, 2015).

What might be the nature of the population code in MSTd that underlies the encoding of heading and eye velocity? One possibility would be that MSTd contains a set of distinct subpopulations, each specialized to encode a particular perceptual variable. Instead, Ben Hamed et al. (2003) found that neurons in MSTd acted more like basis functions, where a majority of cells was involved in the simultaneous encoding of multiple perceptual variables. A similar picture emerged when we investigated the involvement of MSTd-like model units in the encoding of both heading and simulated eye rotation velocity (Fig. 8A). A model unit was said to contribute to the encoding of a perceptual variable if both of its linear decoding weights exceeded a certain threshold, set at 1% of the maximum weight magnitude across all  $N \times 2$  weight values. If this was true for both heading (i.e., FOE) and eye velocity [i.e., pursuit (P)], the unit was said to code for both (Fig. 8A, FOE and P), which was the case for 57% of all units. Analogously, if the weights exceeded the threshold for only one variable or the other, the unit was labeled either FOE or P. Only a few units did not contribute at all (labeled “none”).

Finally, we asked how the accuracy and efficiency of the encoding would change with the number of basis vectors (i.e., the only open parameter of NMF; see Non-negative matrix factorization). To determine accuracy, we repeated the heading decoding experiment for a number of basis vectors  $B = 2^i$ , where  $i \in \{4, 5, \dots, 8\}$ , and measured the Euclidean distance between





**Figure 7.** *A–F*, Direction differences between rotation and translation, for MSTd neurons (*A, C, E*), reprinted with permission from Takahashi et al. (2007), their Figure 7, and the population of MSTd-like model units (*B, D, F*). *A, B*, Histograms of the absolute differences in 3D preferred direction ( $|\Delta$  preferred direction  $|$ ) between rotation and translation. *C, D*, Distributions of preferred direction differences as projected onto each of the three cardinal planes, corresponding to front view, side view, and top view. *E, F*, The ratio of the lengths of the 2D and 3D preferred direction vectors is plotted as a function of the corresponding 2D projection of the difference in preferred direction (red, green, and blue circles for each of the front view, side view, and top view data, respectively).

**Table 3. Sparse population code for perceptual variables (N = 144)**

	FOE ( $x, y$ )	Eye velocity ( $\omega_x, \omega_y$ )
Ben Hamed et al., 2003	(3.62° ± 6.78°, 3.87° ± 4.96°)	(1.39° ± 3.69°, 1.38° ± 3.02°)
This article	(5.75° ± 5.62°, 6.02° ± 5.51°)	(0.82° ± 0.89°, 0.92° ± 0.99°)

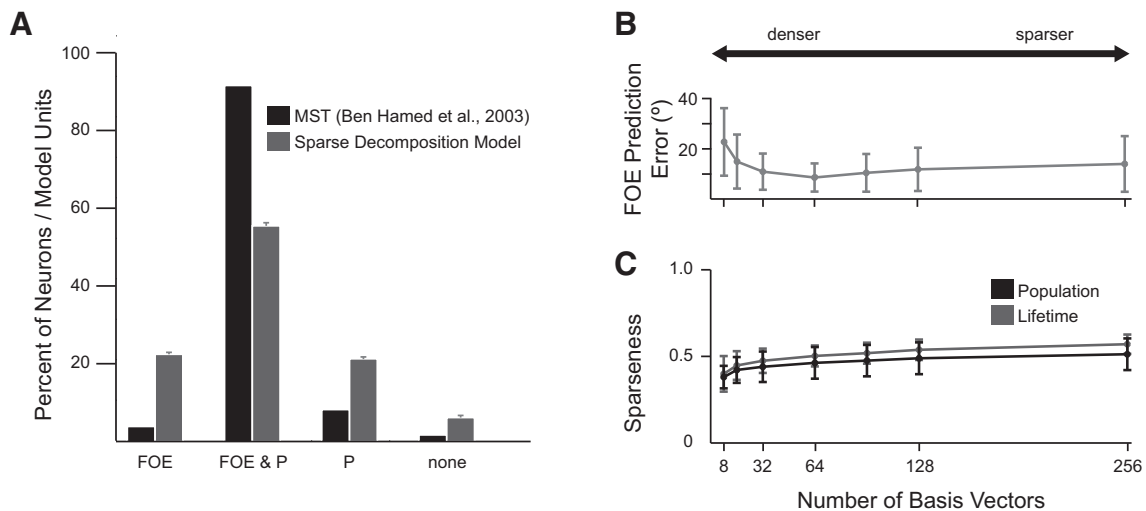
the predicted and actual 2D FOE coordinates. The result is shown in Figure 8*B*. Each data point shows the heading prediction error of the model on the test set for a particular number of basis vectors, averaged over 10 trials (i.e., 10-fold of the cross-

validation procedure). The vertical bars are the SD. The lowest prediction error was achieved with  $B = 64$  (Kolmogorov–Smirnov test,  $p \approx 10^{-12}$ ). To determine the sparseness of the population code, we investigated how many MSTd-like model units were activated by any given stimulus in the dataset (population sparseness) as well as how many stimuli any given model unit responded to (lifetime sparseness; Fig. 8*C*). Sparseness metrics were computed according to the definition by Vinje and Gallant (2000), which can be understood as a measure of both the nonuniformity (“peakiness”) and strength of the population response (see Sparseness). A sparseness value of zero would be indicative of a dense code (where every stimulus would activate every neuron), whereas a sparseness value of one would be indicative of a local code (where every stimulus would activate only one neuron; Spanne and Jörntell, 2015). As expected, the analysis revealed that both population and lifetime sparseness monotonously increased with an increasing number of basis vectors. Sparseness values ranging from  $\sim 0.41$  for  $B = 16$  to 0.65 for  $B = 256$  suggested that the population code was sparse in all cases, as it did not get close to the extreme case of either a local or a dense code.

Overall, these results suggest that our model is compatible with a series of findings demonstrating that macaque MSTd might encode a number of self-motion-related variables using a distributed, versatile population code that is not restricted to a specific subpopulation of neurons within MSTd (Bremmer et al., 1998; Ben Hamed et al., 2003; Gu et al., 2010; Xu et al., 2014). Interestingly, we found that the sparseness regime in which model MSTd achieved the lowest heading prediction error (Fig. 8*C*) was also the regime in which MSTd-like model units reproduced a variety of known MSTd visual response properties (Figs. 5, 6, 7; see also Figs. 10, 11, 12, 13).

### Heading perception during eye movements

We also studied how the accuracy of the population code for heading in model MSTd is affected by eye movements. Eye movements distort the pattern of full-field motion on the retina, causing the FOE of an expanding flow field to no longer indicate heading (see Optic flow stimuli). Under these circumstances, the brain must find a way to discount the motion components caused by eye rotations. The classic viewpoint is that this is achieved with the help of extraretinal signals (Wallach, 1987; but see Kim et al., 2015). Indeed, macaque MSTd might be well equipped to account for eye movement-related signals (Komatsu and Wurtz, 1988; New-



**Figure 8.** Population code underlying the encoding of perceptual variables such as heading (FOE) and eye velocity (pursuit, P). **A**, Distribution of FOE and pursuit selectivities in MSTd (dark gray), adapted with permission from Ben Hamed et al. (2003), their Figure 3, and in the population of MSTd-like model units (light gray). Neurons or model units were involved in encoding heading (FOE), eye velocity (P), both (EYE and P), or neither (none). **B**, Heading prediction (generalization) error as a function of the number of basis vectors using 10-fold cross-validation. Vertical bars are the SD. **C**, Population and lifetime sparseness as a function of the number of basis vectors. Operating the sparse decomposition model with  $B = 64$  basis vectors co-optimizes for both accuracy and efficiency of the encoding, and leads to basis vectors that resemble MSTd receptive fields.

some et al., 1988; Bradley et al., 1996; Page and Duffy, 1999; Morris et al., 2012).

To investigate the effect of eye movements on heading perception, we asked whether model MSTd could predict heading for a number of simulated scene layouts using the linear decoding weights described in the previous section, and compared the results to the psychophysical study of Royden et al. (1994). In these experiments, observers viewed displays of randomly placed dots whose motions simulated translation in the absence and presence of rotations toward a back plane (Fig. 9A), a ground plane (Fig. 9D), and a 3D dot cloud (Fig. 9G). Observers were instructed to fixate a cross that either remained stationary in the center of the display (simulated eye movement condition) or moved horizontally across the display on the horizontal midline (real eye movement condition) with speeds of 0,  $\pm 1^\circ/\text{s}$ ,  $\pm 2.5^\circ/\text{s}$ , and  $\pm 5^\circ/\text{s}$ . At the end of the motion sequence, seven equally spaced lines appeared  $4^\circ$  apart, and the observers indicated the line closest to their perceived headings [seven-alternative, forced-choice paradigm (7AFC)]. Behavioral results are shown in Figure 9, B, E, and H, for the simulated eye movement condition (closed symbols) and the real eye movement condition (open symbols), averaged over 20 trials.

We applied the same experimental protocol to model MSTd. Visual stimuli were generated according to the motion parameters provided by Royden et al. (1994), and subtended  $90^\circ \times 90^\circ$  of visual angle. Heading predictions were generated by applying the linear decoding weights described in the previous section to MSTd population activity. These predictions were then rounded to the nearest available heading in the 7AFC task ( $4^\circ$  apart, therefore spanning  $\pm 12^\circ$  around straight-ahead headings). Figure 9, C, F, and I, summarizes heading predictions of model MSTd averaged over 20 trials for reference headings of  $-4^\circ$  (blue circles), 0 (green squares), and  $+4^\circ$  (red triangles) relative to straight-ahead headings. Consistent with human performance in the simulated eye movement condition, the modeled MSTd could discount eye rotations only if they were relatively small (within  $\pm 1^\circ/\text{s}$ ), but the model made systematic prediction errors for larger rotation rates. These error rates increased with increasing eye rotation velocity, and were largest for observer movement

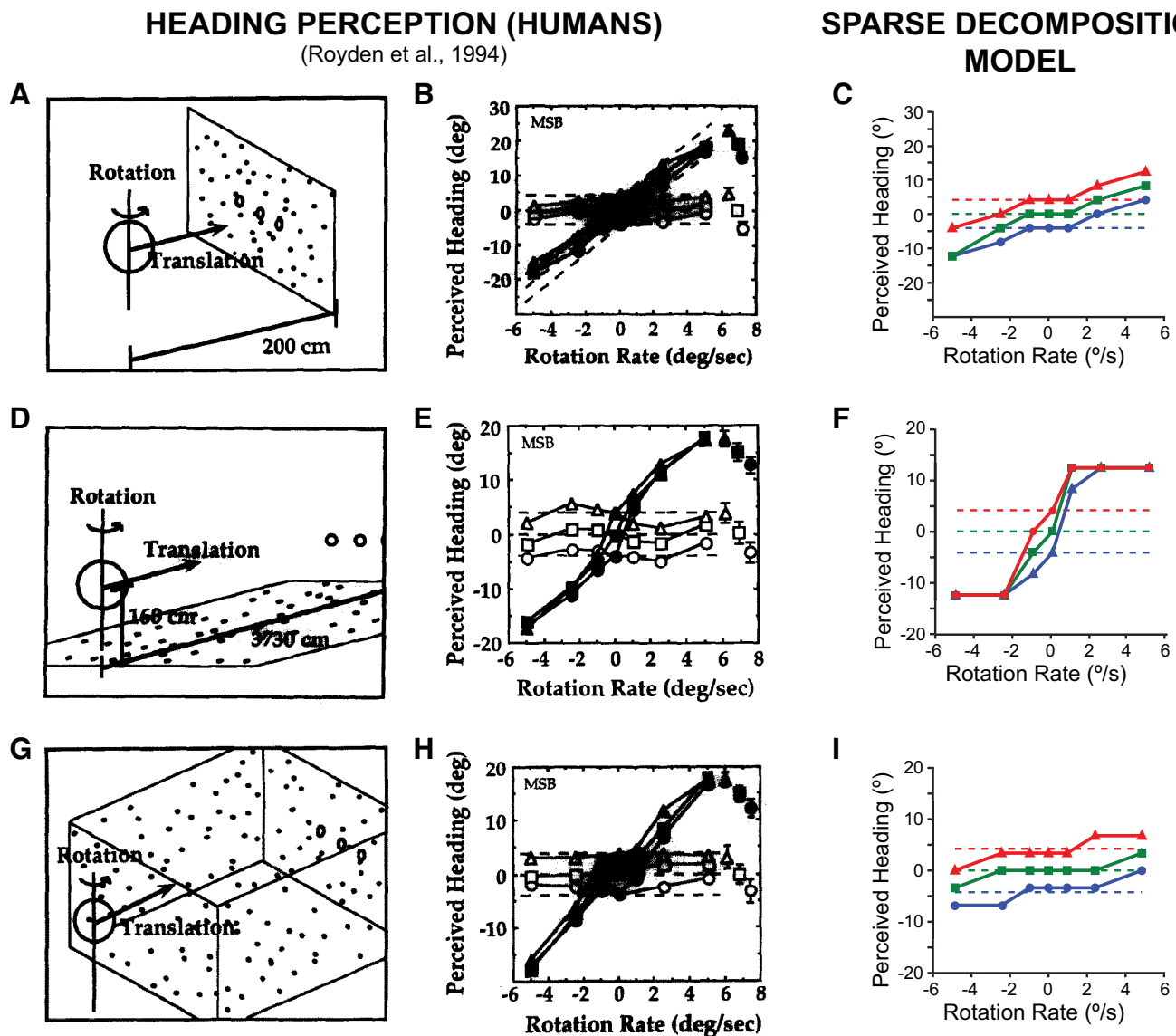
toward a back plane (Fig. 9C) and over a ground plane (Fig. 9F). In all three conditions, the model performed slightly better than humans, which might be due to the model not having to deal with the sources of noise that are commonly present in biological brains (e.g., photoreceptor noise, stochastic signal processing, nonidealized neuronal tuning curves). It is interesting to note that model MSTd could discount eye rotations in the dot cloud condition as long as the rotation rates were within  $\pm 2.5^\circ/\text{s}$  (Fig. 9I). Since dot clouds provide the most depth cues among all three conditions, it is possible that the absence of noise simplified the task of accounting for eye rotations, because translational flow components depend on depth, but rotational components do not (see Eq. 3).

Overall, these simulations suggest that a linear readout mechanism of MSTd-like population activity can account for the behavioral performance of humans in simple heading perception tasks. Both humans and models made systematic prediction errors, which are consistent with a heading perception mechanism based on the idea that heading can be inferred from the FOE location of an optic flow field, which is accurate only in the absence of rotations (see Optic flow stimuli).

#### Population code underlying heading discrimination

Another interesting behavioral effect that might be due to MSTd population coding is the fact that heading discrimination is most precise when subjects have to discriminate small variations around straight-ahead headings (Crowell and Banks, 1993; Gu et al., 2010). It was long believed that this behavioral effect could be explained by an abundance of MSTd neurons preferring straight-ahead headings with sharp tuning curves (Duffy and Wurtz, 1995). However, Gu et al. (2010) were able to demonstrate that this behavior might instead be due to an abundance of MSTd neurons preferring lateral headings with broad, cosine-like tuning curves (Lappe et al., 1996; Gu et al., 2006), causing their peak discriminability (steepest tuning-curve slopes) to lie near straight-ahead headings.

We tested whether simulated population activity in model MSTd could account for these behavioral effects by computing peak discriminability and Fisher information from heading



**Figure 9.** *A–I*, Heading perception during observer translation in the presence of eye movements. Shown are three different scene geometries (back plane, *A–C*; ground plane, *D–F*; dot cloud, *G–I*), reprinted with permission from Royden et al. (1994), their Figures 6, 8, 9, 12, and 13. Observer translation was parallel to the ground and was in one of three directions (open circles), coinciding with the fixation point. Real and simulated eye movements were presented with rates of 0,  $\pm 1^\circ/\text{s}$ ,  $\pm 2.5^\circ/\text{s}$ , or  $\pm 5^\circ/\text{s}$ . *B, E, H*, Perceived heading reported by human subjects for real and simulated eye movements (open and closed symbols, respectively). *C, F, I*, Behavioral performance of model MSTd for simulated eye movements. Horizontal dotted lines indicate the actual headings of  $-4^\circ$  (blue triangles),  $0^\circ$  (green squares), and  $+4^\circ$  (red circles) relative to straight-ahead headings.

tuning curves, following the experimental protocol of Gu et al. (2010). Peak discriminability occurs for motion directions where the slope of a neuronal tuning curve is steepest (Seung and Sompolinsky, 1993; Purushothaman and Bradley, 2005; Jazayeri and Movshon, 2006), and Fisher information puts an upper bound on the precision with which population activity can be decoded with an unbiased estimator (Seung and Sompolinsky, 1993; Pouget et al., 1998).

The heading tuning curve of every model unit was measured by presenting 24 directions of self-translation in the horizontal plane ( $0, \pm 15^\circ, \pm 30^\circ, \dots, \pm 165^\circ$  and  $180^\circ$  relative to straight-ahead headings, while elevation was fixed at  $0^\circ$ ) through a 3D cloud of dots that occupied a space that was 0.75 m deep, subtended  $90^\circ \times 90^\circ$  of visual angle, and drifted at 0.3 m/s (see 3D translation/rotation protocol). Here we adapted the coordinate system to coincide with the one used by Gu et al. (2010), so that azimuth angles were expressed relative to straight-ahead head-

ings. The slope of the tuning curve was computed by interpolating the coarsely sampled data using a spline function (resolution,  $0.01^\circ$ ) and then taking the spatial derivative of the fitted curve at each possible reference heading. Peak discriminability was achieved where the slope of the tuning curve was steepest.

Figure 10 shows the distribution of reference headings at which neurons in MSTd exhibit their minimum neuronal threshold (Fig. 10*A*; Gu et al., 2010), compared with peak discriminability computed from simulated activity of MSTd-like model units (Fig. 10*B*). Both neurophysiologically measured and simulated distributions had clear peaks around forward ( $0^\circ$ ) and backward ( $180^\circ$ ) headings. To further illustrate the relationship between peak discriminability and peak firing rate, the tuning-width at half-maximum is plotted versus heading preference (location of peak firing rate) for neurons in macaque MSTd (Fig. 10*C*) and MSTd-like model units (Fig. 10*D*). Consistent with neurons in

macaque MSTd, the distribution of preferred headings in model MSTd was significantly nonuniform ( $p < 0.05$ ; see Uniformity test) with peaks at  $\pm 90^\circ$  azimuth relative to straight-ahead headings (i.e., lateral headings), and most model units had broad tuning widths ranging between  $90^\circ$  and  $180^\circ$ . Surprisingly, our model was also able to recover what appears to be a distinct subpopulation of narrowly tuned units that prefer forward headings (Fig. 10C,D, open circles).

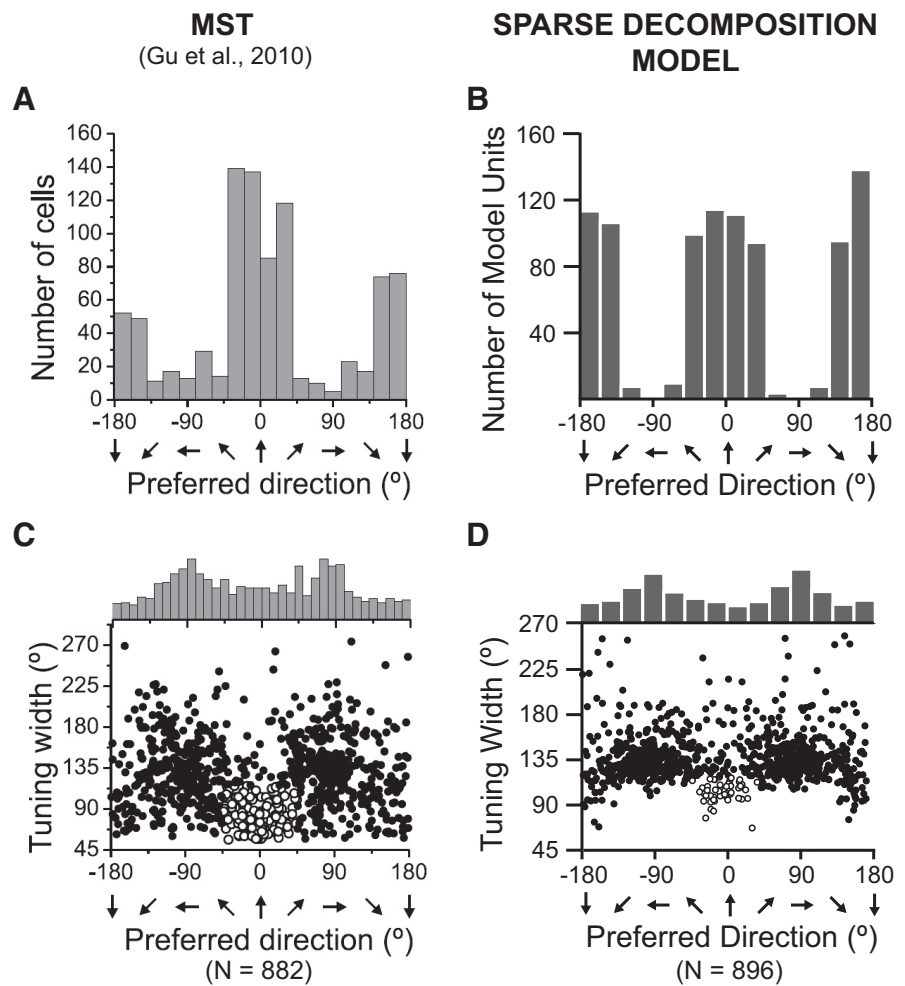
Figure 11 shows population Fisher information derived from interpolated tuning curves for 882 neurons in macaque MSTd (Fig. 11A, red) and 896 MSTd-like model units (Fig. 11B). According to Equation 1, the contribution of each neuron (model unit) to Fisher information is the square of its tuning curve slope at a particular reference heading divided by the corresponding mean firing rate (unit activation). Note that even though MSTd-like model units are deterministic, they exhibit response variability when repeatedly presented with 3D clouds made from dots with random depth (150 repetitions, see Fisher information analysis). Consistent with data from macaque MSTd, there is a clear dependence of Fisher information on the reference headings for our population of MSTd-like model units, with a maximum occurring for headings near  $0^\circ$  (i.e., straight-ahead headings) and a minimum occurring for headings near  $\pm 90^\circ$  (i.e., lateral headings).

Overall, these results are in good agreement with population statistics reported by Gu et al. (2010) and provide further computational support that behavioral dependence on reference heading can be largely explained by the precision of an MSTd-like population code.

### Gaussian tuning in spiral space

The finding that lateral headings are overrepresented in MSTd (Gu et al., 2006, 2010; Takahashi et al., 2007) then raises the question as to how these recent data might be reconciled with earlier studies reporting an abundance of MSTd cells preferring forward motions (i.e., expanding flow stimuli; Duffy and Wurtz (1995)). A possible explanation was offered by Gu et al. (2010), who hypothesized that observed differences in population statistics might be (at least partially) due to a selection bias: to locate visually responsive cells in MSTd, the authors of earlier studies tended to screen for cells that discharge in response to expansion stimuli (Duffy and Wurtz, 1991a,b, 1995; Graziano et al., 1994). In contrast, later studies recorded from any MSTd neuron that was spontaneously active or responded to a large-field, flickering random-dot stimulus (Gu et al., 2006, 2010; Takahashi et al., 2007).

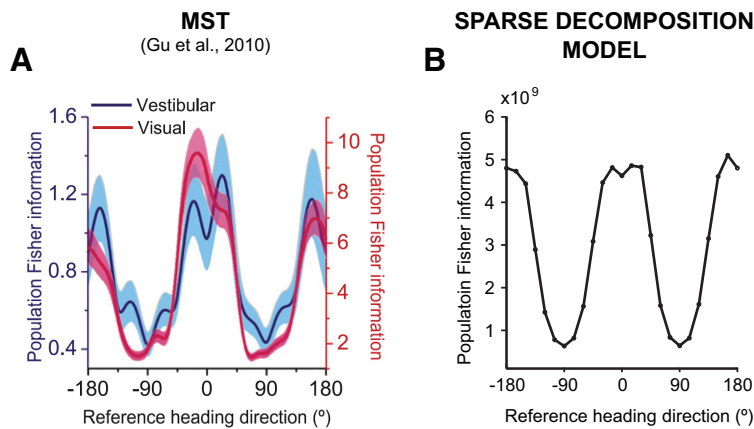
To test their hypothesis, we applied the same selection bias to our sample of MSTd-like model units, and restricted further analysis to



**Figure 10.** *A–D*, Heading discriminability based on population activity in macaque MSTd (*A, C*), reprinted with permission from Gu et al. (2010), their Figure 4, and in model MSTd (*B, D*). For the sake of clarity, we abandon our previously used coordinate system in favor of the one used by Gu et al. (2010), where lateral headings correspond to  $\pm 90^\circ$ , and  $0^\circ$  corresponds to straight-ahead headings. *A, B*, Distribution of the direction of maximal discriminability, showing a bimodal distribution with peaks around the forward ( $0^\circ$ ) and backward ( $\pm 180^\circ$ ) directions. *C, D*, Scatter plot of the tuning-width at half-maximum vs the preferred direction of each neuron or model unit. The top histogram shows the marginal distribution of heading preferences. Also highlighted is a subpopulation of neurons or model units with direction preferences within  $45^\circ$  of straight-ahead headings and tuning width  $< 115^\circ$  (open symbols).

the selected subpopulation. The selection process was mimicked simply by making it more likely for a model unit to be selected the stronger it responded to an expansion stimulus. Of a total of 896 MSTd-like model units, 188 were selected for further processing.

We then proceeded with the experimental protocol described by Graziano et al. (1994) to establish a visual tuning profile for the remaining 188 model units. Model units were presented with eight optic flow stimuli that spanned what they termed “spiral space”: expansion, contraction, clockwise rotation, counter-clockwise rotation, and four intermediate spiral patterns. These stimuli differed from previously described optic flow stimuli in that they did not accurately depict observer motion in three dimensions. Instead, stimulus diameter was limited to  $30^\circ$ , and angular velocity simply increased with distance to the center of the stimulus, with a maximum speed of  $17.2^\circ/s$ . As in the study by Graziano et al. (1994), we fit the resulting tuning curves with a Gaussian function to find the peak (the mean of the Gaussian) that corresponded to the preferred direction in spiral space, and to provide a measure of bandwidth ( $\sigma$ , the SD of the Gaussian) and goodness-of-fit ( $r$ , the correlation coefficient).



**Figure 11.** *A, B*, Population Fisher information of macaque MSTd (*A*), reprinted with permission from Gu et al. (2010), their Figure 5, and of model MSTd (*B*). Error bands in *A* represent 95% confidence intervals derived from a bootstrap procedure.

The results are shown in Figure 12. When looking at the preferred motion direction of MSTd cells (Fig. 12*A*), Graziano et al. (1994) observed Gaussian tuning across the full range of rotational flow fields. Here, 0° corresponded to clockwise rotation, 90° to expansion, 180° to counterclockwise rotation, 270° to contraction, and the oblique directions (45°, 135°, 225° and 315°) corresponded to four intermediate spiral stimuli. Each arrow indicated the preferred direction or peak response (the mean of the Gaussian) of each neuron ( $N = 57$ ) in spiral space. Consistent with these data from macaque MSTd, the majority of preferred spiral directions in our subpopulation of MSTd-like model units ( $N = 188$ ) clustered near expansion, with only few units preferring contraction, and ~51% of units were spiral tuned (vs 35% in their sample; Fig. 12*B, E*). Similar to 86% of isolated MSTd cells having smooth tuning curves and good Gaussian fits (i.e., a correlation coefficient of  $r \geq 0.9$ , with an average  $r$  of 0.97), 112 of 188 MSTd-like model units (60%) had  $r \geq 0.9$ , with an average  $r$  of 0.924. Compared with real MSTd neurons, the model units had comparable, although slightly broader Gaussian widths (an average  $\sigma$  of 111° and an SE of 24° in our case versus an average  $\sigma$  of 61° and an SE of 5.9° in their case).

If the predominance of expansion cells in Figure 12*D* is truly due to a preselection procedure, then any bias in the data should be removed when the above experimental procedure is applied to all cells in MSTd. Indeed, extending the analysis to the entire population of MSTd-like model units revealed a more uniform sampling of direction selectivity across spiral space (Fig. 12*C*), which flattened the distribution of preferred motion directions (Fig. 12*F*). A total of 677 of 896 MSTd-like model units (76%) had  $r \geq 0.9$ , with an average  $r$  of 0.929, an average  $\sigma$  of 108°, and an SE of 9.1°.

### Continuum of 2D response selectivity

Interestingly, the same narrative can be applied to other early studies of MSTd in which cells were supposedly selected depending on how well they responded to radial (R) motion. For example, using a “canonical” set of 12 flow stimuli (eight directions of planar motion; expansion and contraction; clockwise and counterclockwise rotation), Duffy and Wurtz (1995) showed that most neurons in MSTd were sensitive to multiple flow components, with only few neurons responding exclusively to planar, circular (C), or radial motion. In a sample of 268 MSTd cells, Duffy and Wurtz (1995) found that 18% of cells primarily responded to one compo-

nent of motion (planar, circular, or radial), that 29% responded to two components [planocircular (PC) or planoradial (PR), but rarely circularradial (CR)], and that 39% responded to all three components (Fig. 13*A*).

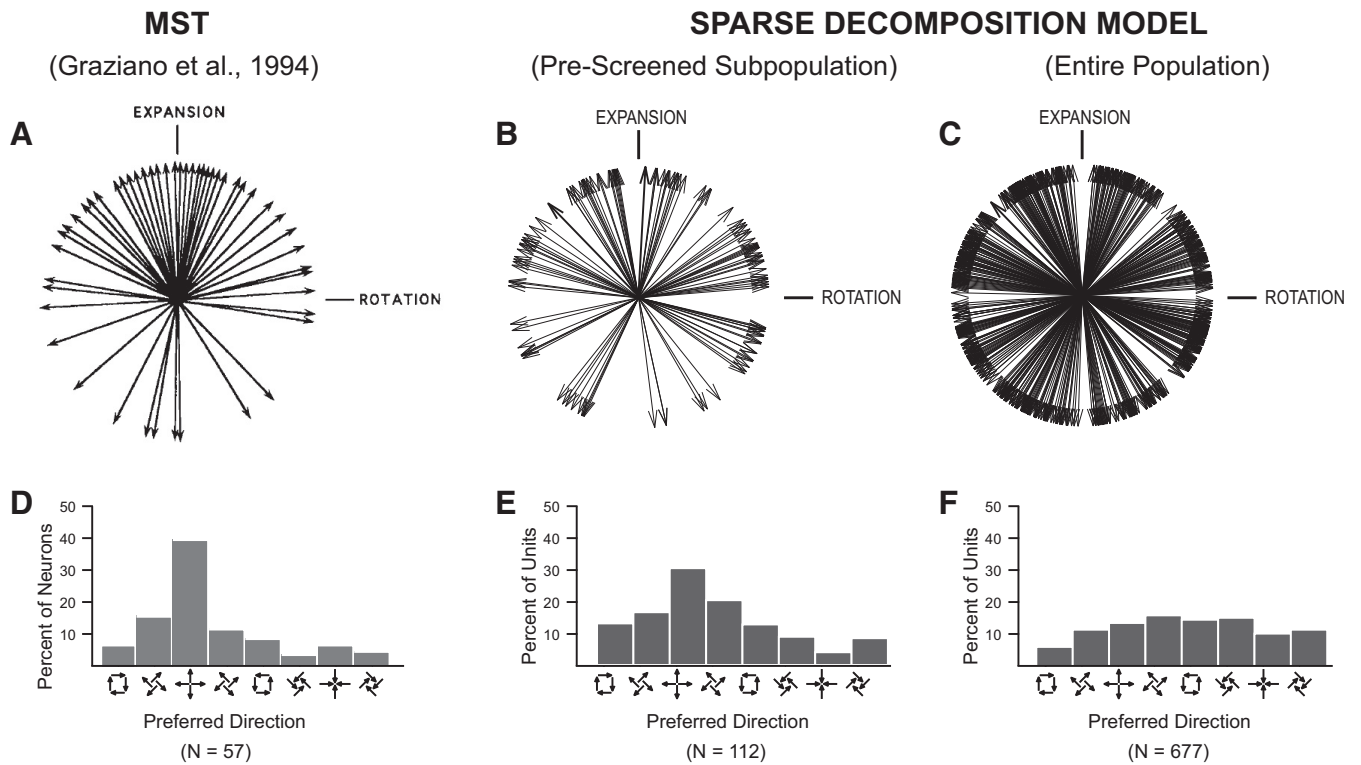
We simulated their experiments by presenting the same 12 stimuli to the subpopulation of 188 MSTd-like model units described above, and classified their responses. Duffy and Wurtz (1995) classified neurons according to the statistical significance of their responses, which was difficult to simulate since MSTd-like model units did not have a defined noise level or baseline output. Instead, we mimicked their selection criterion by following a procedure from Perrone and Stone (1998), where the response of a model unit was deemed “significant” if it exceeded 12% of the largest response that was observed for any model unit in response to any of the tested stimuli.

Without further adjustments, our model recovered a distribution of response selectivities very similar to those reported by Duffy and Wurtz (1995) (Fig. 13*B*). The largest group was formed by triple-component or planocircularradial (PCR) MSTd-like model units with selective responses to planar, circular, and radial stimuli (white). Double-component units were mainly PR with responses to a planar or radial stimulus (magenta), with few PC units (yellow) and only a handful of CR units (cyan). Single-component cells were mainly P units (red), with only few R units and C units (blue). We did not find any nonselective excitatory (NSE) or nonselective inhibitory (NSI) units (because our model did not include inhibitory units).

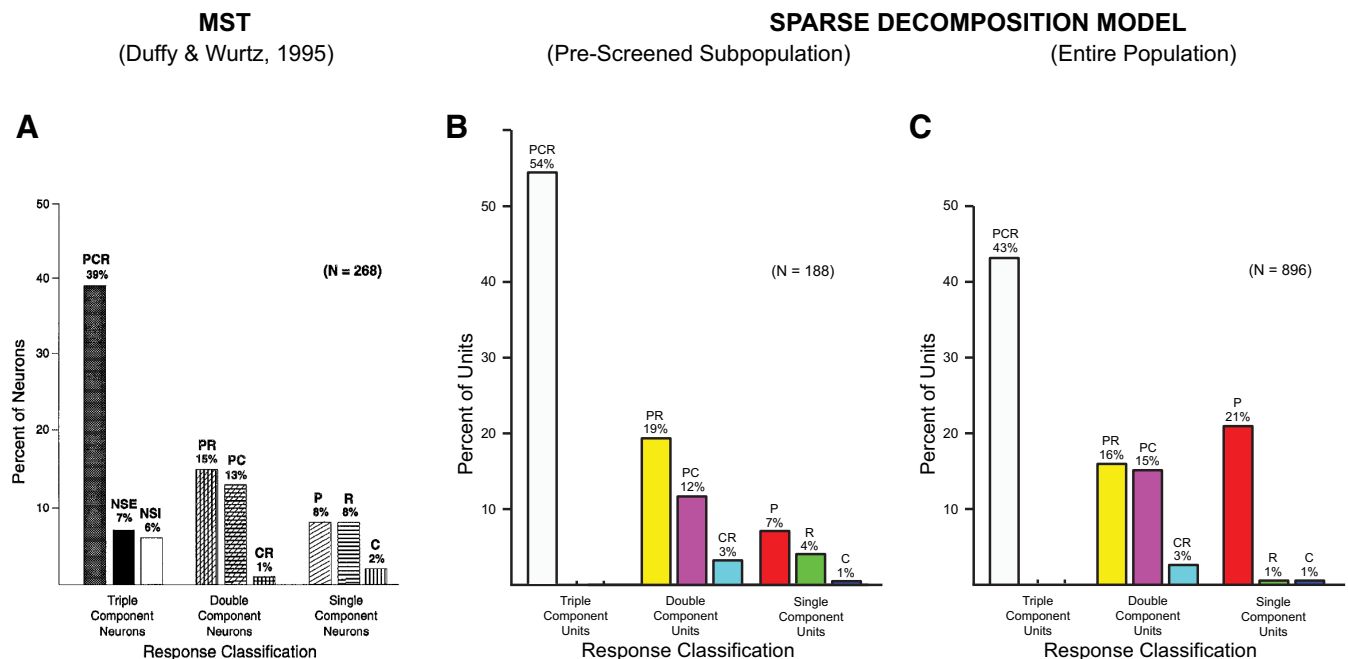
Only 1% of the cells observed by Duffy and Wurtz (1995) were CR cells, which they suggested were equivalent to the spiral-selective neurons reported by Graziano et al. (1994). Thus, they concluded that spiral tuning was rare in MSTd. Interestingly, our model recovers distributions that are comparable to both empirical studies; that is, an abundance of spiral-tuned cells in Figure 12, and an abundance of PCR cells in Figure 13. Our results thus offer an alternative explanation to these seemingly contradictory findings, by considering that most spiral-tuned cells, as identified by Graziano et al. (1994), might significantly respond to planar stimuli when analyzed under the experimental protocol of Duffy and Wurtz (1995), effectively making most spiral-tuned cells part of the PCR class of cells, as opposed to the CR class of cells.

What might these data look like in the absence of any preselection procedure? To answer this question, we extended the response classification to the entire population of 896 MSTd-like model units. As is evident in Figure 13*C*, this analysis revealed a large fraction of P units that had not previously been included. In addition, the relative frequency of units responding to radial motion (i.e., R, PR, and PCR units) decreased noticeably.

Overall, these results suggest that our model is in agreement with a wide variety of MSTd data collected under different experimental protocols, and offers an explanation of how these data might be brought into agreement when differences in the sampling procedure are accounted for.



**Figure 12.** Gaussian tuning in spiral space. **A**, Gaussian tuning of a sample of 57 neurons across the full range of rotational flow fields, reprinted with permission from Graziano et al. (1994), their Figure 9. Each arrow indicates the peak response (the mean of the Gaussian fit) of each neuron in spiral space. **B**, The distribution of preferred spiral directions of a sample of 112 MSTd-like model units whose tuning curves were well fit with a Gaussian. Model units were more likely to be included in the sample the stronger they responded to an expansion stimulus. **C**, The distribution of preferred spiral directions applied to the entire population of 896 MSTd-like model units, of which 677 had smooth Gaussian fits. **D**, Bar plot of the data in **A**, for better comparison, adapted with permission from Clifford et al. (1999), their Figure 5. **E**, Bar plot of the data in **B**. **F**, Bar plot of the data in **C**.



**Figure 13.** Continuum of response selectivity. **A**, Response classification of a sample of 268 MSTd neurons, reprinted with permission from Duffy and Wurtz (1995), their Figure 3. **B**, Response classification of a sample of 188 MSTd-like model units. Model units were more likely to be included in the sample the stronger they responded to an expansion stimulus. **C**, Response classification of the entire population of 896 MSTd-like model units. Triple-component cells were as follows: PCR, NSE, and NSI. Double-component cells were as follows: PR, PC, and CR. Single-component cells were as follows: P, R, and C. Eighty-one percent of neurons in **A**, 88% of model units in **B**, and 77% of model units in **C** responded to more than one type of motion.

## Discussion

We found that applying NMF (Paatero and Tapper, 1994; Lee and Seung, 1999, 2001) to MT-like patterns of activity can account for several essential response properties of MSTd neurons, such as 3D translation and rotation selectivity (Gu et al., 2006; Takahashi et al., 2007); tuning to radial, circular, and spiral motion patterns (Graziano et al., 1994; Lagae et al., 1994; Duffy and Wurtz, 1995); as well as heading selectivity (Page and Duffy, 1999, 2003; Ben Hamed et al., 2003). This finding suggests that these properties might emerge from MSTd neurons performing a biological equivalent of dimensionality reduction on their inputs. Furthermore, the model accurately captures prevalent statistical properties of visual responses in macaque MSTd, such as an overrepresentation of lateral headings (Lappe et al., 1996; Gu et al., 2006, 2010; Takahashi et al., 2007) that can predict behavioral thresholds of heading discrimination (Gu et al., 2010) and heading perception during eye movements (Royden et al., 1994). At the population level, model MSTd efficiently and accurately predicts a number of perceptual variables (e.g., heading and eye rotation velocity) using a sparse distributed code (Olshausen and Field, 1996, 1997; Ben Hamed et al., 2003), consistent with ideas from the efficient-coding and free-energy principles (Attneave, 1954; Barlow, 1961; Linsker, 1990; Simoncelli and Olshausen, 2001; Friston et al., 2006; Friston, 2010).

### Sparse decomposition model of MSTd

It is well known that MSTd neurons do not decompose optic flow into its first-order differential invariants (i.e., into components of divergence, curl, and deformation; Koenderink and van Doorn, 1975). Instead, neurons in MSTd act as “templates” (Perrone, 1992) that perform a dot product-type operation to signal how well the apparent motion on the retina matches their preferred flow component or mixture of components (Saito et al., 1986; Tanaka et al., 1989b; Orban et al., 1992). This “template matching” was later demonstrated to be a powerful approach to self-motion analysis, possibly at work at least in mammals (Lappe and Rauschecker, 1994; Perrone and Stone, 1994, 1998) and insects (Krapp and Hengstenberg, 1996; Srinivasan et al., 1996; Krapp, 2000), all without the need to perform a mathematical decomposition of optic flow.

Alternatively, we provide computational evidence that MSTd neurons might decompose optic flow, in the sense of matrix factorization, to reduce the number of templates used to represent the spectrum of retinal flow fields encountered during self-movement in an efficient and parsimonious fashion. Such a representation would be in agreement with the efficient-coding or infomax principle (Attneave, 1954; Barlow, 1961; Linsker, 1990; Simoncelli and Olshausen, 2001), which posits that sensory systems should use knowledge about statistical regularities of their input to maximize information transfer. If information maximization is interpreted as optimizing for both accuracy (prediction error) and efficiency (model complexity), then the efficient coding principle can be understood as a special case of the free-energy principle (Friston et al., 2006; Friston, 2010).

A sparse, parts-based decomposition model of MSTd implements these principles in that it can co-optimize accuracy and efficiency by representing high-dimensional data with a relatively small set of highly informative variables. As such, the model is intimately related to the framework of non-negative sparse coding (Hoyer, 2002; Eggert and Korner, 2004). Efficiency is achieved through sparseness, which is a direct result of the non-negativity constraints of NMF (Lee and Seung, 1999). Accuracy trades off with efficiency, and can be achieved by both minimiz-

ing reconstruction error and controlling for model complexity (i.e., by tuning the number of basis vectors; Fig. 10). Statistical knowledge enters the model via the relative frequencies of observations in the input data, which are controlled by both natural viewing conditions and natural scene statistics. NMF explicitly discourages statistically inefficient representations, because strongly accounting for a rare observation at the expense of ignoring a more common stimulus component would result in an increased reconstruction error.

Interestingly, we found that the sparseness regime in which model MSTd achieved the lowest heading prediction error and thus showed the greatest potential for generalization (Fig. 8*B,C*) was also the regime in which MSTd-like model units reproduced a variety of known MSTd visual response properties (Figs. 5, 6, 7, 10, 11, 12, 13). In contrast to findings about early sensory areas (Olshausen and Field, 1996, 1997; Vinje and Gallant, 2000), this regime does not use an overcomplete dictionary, yet can still be considered a sparse coding regime (Spanne and Jörntell, 2015). Sparse codes are a trade-off between dense codes (where every neuron is involved in every context, leading to great memory capacity but suffering from cross talk among neurons) and local codes (where there is no interference but also no capacity for generalization; Spanne and Jörntell, 2015). We speculate that sparse codes with a relatively small dictionary akin to the one described in this article might be better suited (as opposed to overcomplete basis sets) for areas such as MSTd, because the increased memory capacity of such a code might lead to compact and multifaceted encodings of various perceptual variables (Bremmer et al., 1998; Ben Hamed et al., 2003; Brostek et al., 2015).

### Model limitations

Despite its simplicity, the present model is able to explain a variety of MSTd visual response properties. However, a number of issues remain to be addressed in the future, such as the fact that neurons in MSTd are also driven by vestibular signals (Page and Duffy, 2003; Gu et al., 2006; Takahashi et al., 2007) and eye movement-related signals (Komatsu and Wurtz, 1988; Newsome et al., 1988; Bradley et al., 1996; Page and Duffy, 1999; Morris et al., 2012). In addition, some neurons are selective not just for heading, but also for path and place (Froehler and Duffy, 2002; Page et al., 2015). Also, we have not yet attempted to model the complex, nonlinear interactions found among different subregions of the receptive field (Duffy and Wurtz, 1991b; Lagae et al., 1994; Mineault et al., 2012), but speculate that they might be similar to the ones reported in MT (Majaj et al., 2007; Richert et al., 2013). At a population level, MSTd exhibits a range of tuning behaviors, from pure retinal to head-centered stimulus velocity coding (Ben Hamed et al., 2003; Yu et al., 2010; Brostek et al., 2015), that include intermediate reference frames (Fetsch et al., 2007). From a theoretical standpoint, the sparse decomposition model seems a good candidate to find an efficient, reference frame-agnostic representation of various perceptual variables (Pouget and Sejnowski, 1997; Pouget and Snyder, 2000; Ben Hamed et al., 2003), but future iterations of the model will have to address these issues step by step.

### Model alternatives

Several computational models have tried to account for heading-selective cells in area MSTd. In the heading-template model (Perrone and Stone, 1994, 1998), MSTd forms a “heading map,” with each MSTd-like unit receiving input from a mosaic of MT-like motion sensors that correspond to the flow that would arise from

a particular heading. Heading is then inferred by the preferred FOE of the most active heading template. A complication of this type of model is that it requires an extremely large number of templates to cover the combinatorial explosion of heading parameters, eye rotations, and scene layouts (Perrone and Stone, 1994), even when the input stimulus space is restricted to gaze-stabilized flow fields (Perrone and Stone, 1998). The velocity gain field model (Beintema and van den Berg, 1998, 2000) tries to reduce the number of combinations by using templates that uniquely pick up the rotational component of flow, but recent evidence argues against this type of model (Berezovskii and Born, 2000; Duijnhouwer et al., 2013). In a related approach to the one presented in this article, Zemel and Sejnowski (1998) proposed that neurons in MST encode hidden causes of optic flow, by using a sparse-distributed representation that facilitates image segmentation and object- or self-motion estimation by downstream readout neurons. Unfortunately, they did not quantitatively assess many of the response properties described in this article. In addition, their model showed an overrepresentation of expanding flow stimuli, which is hard to reconcile with recent findings (Gu et al., 2006).

In contrast, our model offers a biologically plausible account of a wide range of visual response properties in MSTd, ranging from single-unit activity to population statistics. Adoption of the sparse decomposition model supports the view that single-unit preferences emerge from the pressure to find an efficient representation of large-field motion stimuli (as opposed to encoding a single variable such as heading), and that these units act in concert to represent perceptual variables both accurately (Gu et al., 2010) and efficiently (Ben Hamed et al., 2003).

## References

- Attneave F (1954) Some informational aspects of visual perception. *Psychol Rev* 61:183–193. [CrossRef Medline](#)
- Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: *Sensory communication* (Rosenblith WA, ed). Chapter 13, pp. 217–234. Cambridge, MA: MIT.
- Beintema JA, van den Berg AV (1998) Heading detection using motion templates and eye velocity gain fields. *Vision Res* 38:2155–2179. [CrossRef Medline](#)
- Beintema JA, van den Berg AV (2000) Perceived heading during simulated torsional eye movements. *Vision Res* 40:549–566. [CrossRef Medline](#)
- Ben Hamed S, Page W, Duffy C, Pouget A (2003) MSTd neuronal basis functions for the population encoding of heading direction. *J Neurophysiol* 90:549–558. [CrossRef Medline](#)
- Berezovskii VK, Born RT (2000) Specificity of projections from wide-field and local motion-processing regions within the middle temporal visual area of the owl monkey. *J Neurosci* 20:1157–1169. [Medline](#)
- Boussaoud D, Ungerleider LG, Desimone R (1990) Pathways for motion analysis: cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *J Comp Neurol* 296:462–495. [CrossRef Medline](#)
- Bradley DC, Maxwell M, Andersen RA, Banks MS, Shenoy KV (1996) Mechanisms of heading perception in primate visual cortex. *Science* 273:1544–1547. [CrossRef Medline](#)
- Bremmer F, Pouget A, Hoffmann KP (1998) Eye position encoding in the macaque posterior parietal cortex. *Eur J Neurosci* 10:153–160. [CrossRef Medline](#)
- Britten KH, Newsome WT (1998) Tuning bandwidths for near-threshold stimuli in area MT. *J Neurophysiol* 80:762–770. [Medline](#)
- Britten KH, Van Wezel RJ (2002) Area MST and heading perception in macaque monkeys. *Cereb Cortex* 12:692–701. [CrossRef Medline](#)
- Brostek L, Büttner U, Mustari MJ, Glasauer S (2015) Eye velocity gain fields in MSTd during optokinetic stimulation. *Cereb Cortex* 25:2181–2190. [CrossRef Medline](#)
- Clifford CW, Beardsley SA, Vaina LM (1999) The perception and discrimination of speed in complex motion. *Vision Res* 39:2213–2227. [CrossRef Medline](#)
- Crowell JA, Banks MS (1993) Perceiving heading with different retinal regions and types of optic flow. *Percept Psychophys* 53:325–337. [CrossRef Medline](#)
- Duffy CJ, Wurtz RH (1991a) Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J Neurophysiol* 65:1329–1345. [Medline](#)
- Duffy CJ, Wurtz RH (1991b) Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *J Neurophysiol* 65:1346–1359. [Medline](#)
- Duffy CJ, Wurtz RH (1995) Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *J Neurosci* 15:5192–5208. [Medline](#)
- Duijnhouwer J, Noest AJ, Lankheet MJ, van den Berg AV, van Wezel RJ (2013) Speed and direction response profiles of neurons in macaque MT and MST show modest constraint line tuning. *Front Behav Neurosci* 7:22. [CrossRef Medline](#)
- Eggert J, Korneer E (2004) Sparse coding and NMF. *Proceedings 2004 IEEE International Joint Conference*, pp 2529–2533. New York: IEEE. [CrossRef](#)
- Fetsch CR, Wang S, Gu Y, Deangelis GC, Angelaki DE (2007) Spatial reference frames of visual, vestibular, and multimodal heading signals in the dorsal subdivision of the medial superior temporal area. *J Neurosci* 27:700–712. [CrossRef Medline](#)
- Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138. [CrossRef Medline](#)
- Friston K, Kilner J, Harrison L (2006) A free energy principle for the brain. *J Physiol Paris* 100:70–87. [CrossRef Medline](#)
- Froehner MT, Duffy CJ (2002) Cortical neurons encoding path and place: where you go is where you are. *Science* 295:2462–2465. [CrossRef Medline](#)
- Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J Neurosci* 2:1527–1537. [Medline](#)
- Gibson JJ (1950) *The perception of the visual world*. Boston: Houghton Mifflin.
- Graziano MS, Andersen RA, Snowden RJ (1994) Tuning of MST neurons to spiral motions. *J Neurosci* 14:54–67. [Medline](#)
- Grossberg S, Mingolla E, Pack C (1999) A neural model of motion processing and visual navigation by cortical area MST. *Cereb Cortex* 9:878–895. [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, 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, Deangelis GC, Angelaki DE (2012) Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. *J Neurosci* 32:2299–2313. [CrossRef Medline](#)
- Hastie T, Tibshirani R, Friedman JH (2009) *The elements of statistical learning: data mining, inference, and prediction*, Ed 2. New York: Springer.
- Hoyer PO (2002) Non-negative sparse coding, pp 557–565. In: *Proceedings of the 2002 12th IEEE Workshop on Neural Networks for Signal Processing*, 2002. New York: IEEE. [CrossRef](#)
- Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci* 9:690–696. [CrossRef Medline](#)
- Kim HR, Angelaki DE, DeAngelis GC (2015) A novel role for visual perspective cues in the neural computation of depth. *Nat Neurosci* 18:129–137. [CrossRef Medline](#)
- Koenderink JJ, van Doorn AJ (1975) Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Opt Acta* 22:773–791. [CrossRef](#)
- Komatsu H, Wurtz RH (1988) Relation of cortical areas MT and MST to pursuit eye movements. III. Interaction with full-field visual stimulation. *J Neurophysiol* 60:621–644. [Medline](#)
- Krapp HG (2000) Neuronal matched filters for optic flow processing in flying insects. *Int Rev Neurobiol* 44:93–120. [CrossRef Medline](#)
- Krapp HG, Hengstenberg R (1996) Estimation of self-motion by optic flow processing in single visual interneurons. *Nature* 384:463–466. [CrossRef Medline](#)
- Lagae L, Maes H, Raiguel S, Xiao DK, Orban GA (1994) Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. *J Neurophysiol* 71:1597–1626. [Medline](#)
- Lappe M, Rauschecker JP (1994) Heading detection from optic flow. *Nature* 369:712–713. [CrossRef Medline](#)



- Lappe M, Bremmer F, Pekel M, Thiele A, Hoffmann KP (1996) Optic flow processing in monkey STS: a theoretical and experimental approach. *J Neurosci* 16:6265–6285. [Medline](#)
- Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401:788–791. [CrossRef Medline](#)
- Lee DD, Seung HS (2001) Algorithms for non-negative matrix factorization. In: *Advances in neural information processing systems 13* (Leen TK, Dietterich TG, Tresp V, eds), pp 556–562. Neural Information Processing Systems Foundation. [CrossRef](#)
- Linsker R (1990) Perceptual neural organization: some approaches based on network models and information theory. *Annu Rev Neurosci* 13:257–281. [CrossRef 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](#)
- Longuet-Higgins HC, Prazdny K (1980) The interpretation of a moving retinal image. *Proc R Soc Lond B Biol Sci* 208:385–397. [CrossRef Medline](#)
- Majaj NJ, Carandini M, Movshon JA (2007) Motion integration by neurons in macaque MT is local, not global. *J Neurosci* 27:366–370. [CrossRef Medline](#)
- Mardia KV (1972) *Statistics of directional data*. London, New York: Academic.
- Mineault PJ, Khawaja FA, Butts DA, Pack CC (2012) Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proc Natl Acad Sci U S A* 109:E972–E980. [CrossRef Medline](#)
- Morris AP, Kubischik M, Hoffmann KP, Krekelberg B, Bremmer F (2012) Dynamics of eye-position signals in the dorsal visual system. *Curr Biol* 22:173–179. [CrossRef Medline](#)
- Newsome WT, Wurtz RH, Komatsu H (1988) Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. *J Neurophysiol* 60:604–620. [Medline](#)
- Nover H, Anderson CH, DeAngelis GC (2005) A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J Neurosci* 25:10049–10060. [CrossRef Medline](#)
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609. [CrossRef Medline](#)
- Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res* 37:3311–3325. [CrossRef Medline](#)
- Orban GA, Lagae L, Verri A, Raiguel S, Xiao D, Maes H, Torre V (1992) First-order analysis of optical flow in monkey brain. *Proc Natl Acad Sci U S A* 89:2595–2599. [CrossRef Medline](#)
- Paatero P, Tapper U (1994) Positive matrix factorization—a nonnegative factor model with optimal utilization of error-estimates of data values. *Environmetrics* 5:111–126. [CrossRef](#)
- Page WK, Duffy CJ (1999) MST neuronal responses to heading direction during pursuit eye movements. *J Neurophysiol* 81:596–610. [Medline](#)
- Page WK, Duffy CJ (2003) Heading representation in MST: sensory interactions and population encoding. *J Neurophysiol* 89:1994–2013. [Medline](#)
- Page WK, Sato N, Froehler MT, Vaughn W, Duffy CJ (2015) Navigational path integration by cortical neurons: origins in higher-order direction selectivity. *J Neurophysiol* 113:1896–1906. [CrossRef Medline](#)
- Perrone JA (1992) Model for the computation of self-motion in biological systems. *J Opt Soc Am A Opt Image Sci Vis* 9:177–194. [CrossRef](#)
- Perrone JA, Stone LS (1994) A model of self-motion estimation within primate extrastriate visual cortex. *Vision Res* 34:2917–2938. [CrossRef Medline](#)
- Perrone JA, Stone LS (1998) Emulating the visual receptive-field properties of MST neurons with a template model of heading estimation. *J Neurosci* 18:5958–5975. [Medline](#)
- Picard RR, Cook RD (1984) Cross-validation of regression-models. *J Am Stat Assoc* 79:575–583. [CrossRef](#)
- Pouget A, Sejnowski TJ (1997) Spatial transformations in the parietal cortex using basis functions. *J Cogn Neurosci* 9:222–237. [CrossRef Medline](#)
- Pouget A, Snyder LH (2000) Computational approaches to sensorimotor transformations. *Nat Neurosci* 3 [Suppl]:1192–1198. [CrossRef Medline](#)
- Pouget A, Zhang K, Deneve S, Latham PE (1998) Statistically efficient estimation using population coding. *Neural Comput* 10:373–401. [CrossRef Medline](#)
- Purushothaman G, Bradley DC (2005) Neural population code for fine perceptual decisions in area MT. *Nat Neurosci* 8:99–106. [CrossRef Medline](#)
- Raiguel S, Van Hulle MM, Xiao DK, Marcar VL, Orban GA (1995) Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *Eur J Neurosci* 7:2064–2082. [CrossRef Medline](#)
- Raudies F (2013) Optic flow. *Scholarpedia* 8:30724. [CrossRef](#)
- Raudies F, Neumann H (2012) A review and evaluation of methods estimating ego-motion. *Comput Vis Image Underst* 116:606–633. [CrossRef](#)
- Richert M, Albright TD, Krekelberg B (2013) The complex structure of receptive fields in the middle temporal area. *Front Syst Neurosci* 7:2. [CrossRef Medline](#)
- Royden CS, Crowell JA, Banks MS (1994) Estimating heading during eye movements. *Vision Res* 34:3197–3214. [CrossRef Medline](#)
- Saito H, Yukie M, Tanaka K, Hikosaka K, Fukada Y, Iwai E (1986) Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J Neurosci* 6:145–157. [Medline](#)
- Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci U S A* 90:10749–10753. [CrossRef Medline](#)
- Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. *Annu Rev Neurosci* 24:1193–1216. [CrossRef Medline](#)
- Snyder JP (1987) *Map projections—a working manual*. Washington, DC: U.S. Government Printing Office.
- Spanne A, Jörnig H (2015) Questioning the role of sparse coding in the brain. *Trends Neurosci* 38:417–427. [CrossRef Medline](#)
- Srinivasan M, Zhang S, Lehrer M, Collett T (1996) Honeybee navigation en route to the goal: visual flight control and odometry. *J Exp Biol* 199:237–244. [Medline](#)
- Swindale NV (1998) Orientation tuning curves: empirical description and estimation of parameters. *Biol Cybern* 78:45–56. [CrossRef Medline](#)
- Takahashi K, Gu Y, May PJ, Newlands SD, DeAngelis GC, Angelaki DE (2007) Multimodal coding of three-dimensional rotation and translation in area MSTd: comparison of visual and vestibular selectivity. *J Neurosci* 27:9742–9756. [CrossRef Medline](#)
- Tanaka K, Saito H (1989a) Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *J Neurophysiol* 62:626–641. [Medline](#)
- Tanaka K, Fukada Y, Saito HA (1989b) Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *J Neurophysiol* 62:642–656. [Medline](#)
- Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287:1273–1276. [CrossRef Medline](#)
- Wallach H (1987) Perceiving a stable environment when one moves. *Annu Rev Psychol* 38:1–27. [CrossRef Medline](#)
- Xu H, Wallisch P, Bradley DC (2014) Spiral motion selective neurons in area MSTd contribute to judgments of heading. *J Neurophysiol* 111:2332–2342. [CrossRef Medline](#)
- Yu CP, Page WK, Gaboriski R, Duffy CJ (2010) Receptive field dynamics underlying MST neuronal optic flow selectivity. *J Neurophysiol* 103:2794–2807. [CrossRef Medline](#)
- Zemel RS, Sejnowski TJ (1998) A model for encoding multiple object motions and self-motion in area MST of primate visual cortex. *J Neurosci* 18:531–547. [Medline](#)
- Zhang KC, Sereno MI, Sereno ME (1993) Emergence of position-independent detectors of sense of rotation and dilation with Hebbian learning—an analysis. *Neural Comput* 5:597–612. [CrossRef](#)