

Emulating the Visual Receptive-Field Properties of MST Neurons with a Template Model of Heading Estimation

John A. Perrone¹ and Leland S. Stone²

¹Department of Psychology, University of Waikato, Hamilton, New Zealand, and ²Human Information Processing Research Branch, NASA Ames Research Center, Moffett Field, California

We have proposed previously a computational neural-network model by which the complex patterns of retinal image motion generated during locomotion (optic flow) can be processed by specialized detectors acting as templates for specific instances of self-motion. The detectors in this template model respond to global optic flow by sampling image motion over a large portion of the visual field through networks of local motion sensors with properties similar to those of neurons found in the middle temporal (MT) area of primate extrastriate visual cortex. These detectors, arranged within cortical-like maps, were designed to extract self-translation (heading) and self-rotation, as well as the scene layout (relative distances) ahead of a moving observer. We then postulated that heading from optic flow is directly encoded by individual neurons acting as heading detectors within the medial superior temporal (MST) area. Others

have questioned whether individual MST neurons can perform this function because some of their receptive-field properties seem inconsistent with this role. To resolve this issue, we systematically compared MST responses with those of detectors from two different configurations of the model under matched stimulus conditions. We found that the characteristic physiological properties of MST neurons can be explained by the template model. We conclude that MST neurons are well suited to support self-motion estimation via a direct encoding of heading and that the template model provides an explicit set of testable hypotheses that can guide future exploration of MST and adjacent areas within the superior temporal sulcus.

Key words: self-motion perception; navigation; optic flow; gaze stabilization; monkey; vision

Self-motion through the environment generates image motion across the retina often called optic flow. During pure translation, retinal motion radiates out symmetrically from a single point, the focus of expansion (FOE), from which heading (instantaneous direction of translation) can be inferred (Gibson, 1950). Rotation caused by eye and head movements or self-motion along a curved path complicates this simple picture because the radial pattern is replaced by more complex patterns (Fig. 1). Nonetheless, theoretical analyses indicate that heading can be recovered from combined translational and rotational optic flow (e.g., Koenderink and van Doorn, 1975; Longuet-Higgins and Prazdny, 1980), and psychophysical studies have shown that humans are able to do so (e.g., Rieger and Toet, 1985; Cutting, 1986; Stone and Perrone, 1997a).

Ascension through the cortical motion pathways from primary visual cortex (V1) through the middle temporal area (MT) to the medial superior temporal area (MST) is characterized by a systematic increase in receptive-field size and complexity (Maunsell and Newsome, 1987). The sensitivity to large-field motion patterns resembling optic flow in the dorsal portion of MST (MSTd) supports the view that MST is involved in self-motion perception (Saito et al., 1986; Tanaka et al., 1986, 1989; Ungerleider and Desimone, 1986; Komatsu and Wurtz, 1988; Duffy and Wurtz,

1991a,b, 1995; Orban et al., 1992; Lagae et al., 1994; Bradley et al., 1996; Lappe et al., 1996). Neurons that respond preferentially to expansion could convey information about forward translation (Saito et al., 1986; Tanaka et al., 1986, 1989; Perrone, 1987, 1990; Glünder, 1990; Hatsopoulos and Warren, 1991), and this principle can be generalized to combined translation and rotation (Perrone, 1992; Perrone and Stone, 1994). However, because many MST neurons show a form of “position invariance,” i.e., they prefer a specific type of motion (e.g., counterclockwise rotation) regardless of where in their receptive field that motion is presented (Duffy and Wurtz, 1991b; Orban et al., 1992; Graziano et al., 1994; Lagae et al., 1994), MST seemed ill-suited to support navigation. In particular, Graziano et al. (1994) stated that “(t)he position invariant responses described in the present article cannot encode the center of expansion in any straightforward way,” that “any simple formulation of the navigation hypothesis must be rejected,” and that “(t)he only way this navigational information could be accurately derived from MSTd is through the use of a coarse, population encoding.” Lappe and Rauschecker (1993) proposed a population-code model of heading estimation in which individual units do not encode heading but must combine their responses to derive it. We took a different view and proposed a model whose individual units directly code for putative headings (Perrone, 1992; Perrone and Stone, 1994) as an early step in the cascade of processing necessary for self-motion perception and navigation (Stone and Perrone, 1997a). The primary purpose of this study is to determine whether the visual receptive-field properties of MST neurons are consistent with an ability to encode heading directly. To this end, we have taken the approach of simulating the template model and of comparing the properties of model output units and those of MST neurons.

Received Dec. 15, 1997; revised May 1, 1998; accepted May 14, 1998.

This work was supported by National Aeronautics and Space Administration RTOPs 199-16-12-37 and 548-50-12 and Grants NAGW 4127 and NAG 2-1168. We thank Drs. Barbara Chapman, Brent Beutter, and Jeff McCandless for their helpful comments on earlier drafts.

Correspondence should be addressed to Dr. John A. Perrone, Department of Psychology, University of Waikato, Private Bag 3105, Hamilton, New Zealand. E-mail address: jpnz@waikato.ac.nz

Copyright © 1998 Society for Neuroscience 0270-6474/98/185958-18\$05.00/0

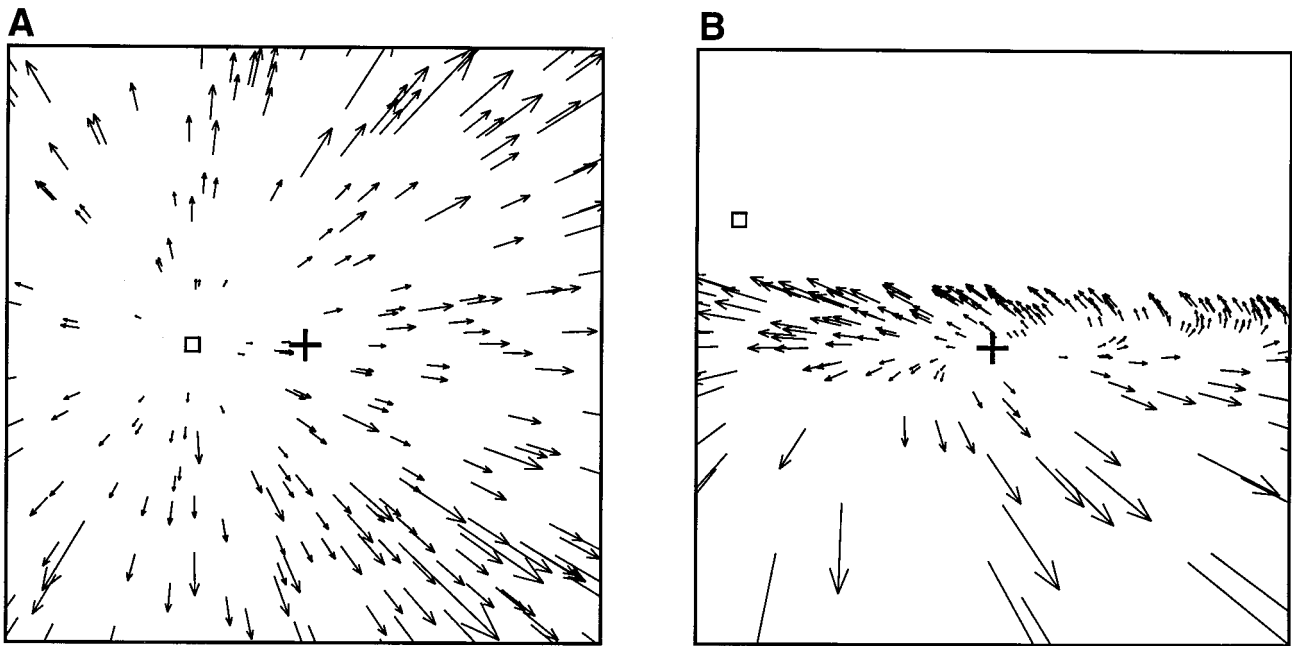


Figure 1. Retinal image motion (optic flow) generated during self-motion. The two velocity vector fields (flow fields) indicate the local velocity (displacement over 1 sec). *A*, This flow field corresponds to observer pure translation with heading direction (open square) to the left of fixation (cross) toward a cloud of random points. In this case, heading coincides with the focus of expansion (FOE). *B*, This flow field results from forward translation over a ground plane combined with an eye rotation caused by gaze stabilization of a ground point below and to the right of heading. Heading is no longer indicated by an FOE, but a pseudo FOE is found at the fixation point. Note the spiral nature of the flow pattern.

Parts of this paper have been published previously (Stone and Perrone, 1994, 1997b).

MATERIALS AND METHODS

A detailed description, rationale, derivation, and demonstration of the performance of the template model can be found elsewhere (Perrone and Stone, 1994). Briefly, the model consists of a two-stage neural network (Fig. 2*A*). It uses MT-like input units (sensors) connected to output units (detectors) that are designed to respond optimally to a specific combination of heading and rotation. Heading is estimated by finding the most active detector within cortical-like maps.

The input units or sensors (Fig. 2*B*) were designed as idealized MT neurons (Maunsell and Van Essen, 1983a; Albright, 1984); they are broadly direction- and speed-tuned motion sensors (30° and 1 octave bandwidths, respectively) whose final output is defined as the product of these two separable factors. The direction output O_d is:

$$O_d = 1.01123 \exp \left[-0.5 \left(\frac{d - d_o}{30} \right)^2 \right] - 0.01123, \quad |d - d_o| < 90^\circ$$

$$= 15 \left[1.01123 \exp \left[-0.5 \left(\frac{d - d_o}{30} \right)^2 \right] - 0.01123 \right], \quad |d - d_o| \geq 90^\circ, \quad (1)$$

with d , the direction of the local image motion, and d_o , the preferred direction of the sensor. The antipreferred inhibition is scaled up to ~15% of the amplitude of the peak-preferred response. The speed output O_s is:

$$O_s = \exp[-0.5(\log_2 s - \log_2 s_o)^2], \quad (2)$$

with s , the speed of the local image motion, and s_o , the preferred speed of the sensor.

The output units or detectors (Fig. 3) combine the responses of particular sets of sensors in such a way as to respond maximally to the optic flow resulting from the combination of a particular heading (α_H , heading azimuth; β_H , heading elevation) and rotation rate (ω_o). Because, in primates, optic flow will generally be experienced under conditions of gaze stabilization that set the rotation axis and eliminate ocular roll (Perrone and Stone, 1994), the five-dimensional self-motion estimation

problem is reduced to only three dimensions: α_H , β_H , and ω_o . The performance of the model is robust to small deviations from the gaze-stabilization assumptions [Perrone and Stone (1994), their Figs. 10, 13].

The receptive-field structure of model detectors is designed using the standard optic-flow equation. Specifically, to construct a heading detector, we connect MT-like input sensors at image location (α_p , β_p) to the detector tuned to α_H , β_H , and ω_o , with their preferred speed and direction chosen using the following equation (for a derivation, see Perrone and Stone, 1994):

$$\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} = \frac{V}{z} \begin{pmatrix} -1 & 0 & \tan \alpha_p \\ 0 & -1 & \frac{\tan \beta_p}{\cos \alpha_p} \end{pmatrix} \begin{pmatrix} \sin \alpha_H \cos \beta_H \\ \sin \beta_H \\ \cos \alpha_H \cos \beta_H \end{pmatrix} + \frac{\omega_o}{\cos^2 \alpha_p \sqrt{\sin^2 \beta_H + \sin^2 \alpha_H \cos^2 \beta_H}}$$

$$\begin{pmatrix} \sin \alpha_p \tan \beta_p & -1 & 0 \\ \tan^2 \beta_p + \cos^2 \alpha_p & -\sin \alpha_p \tan \beta_p & 0 \end{pmatrix} \begin{pmatrix} \sin \beta_H \\ -\sin \alpha_H \cos \beta_H \\ 0 \end{pmatrix}. \quad (3)$$

Equation 3 ensures that the preferred velocity (\dot{X} , \dot{Y}) of each input sensor coincides with the expected optic flow of a point at depth z for an observer traveling at speed V .

In this paper, we examined two configurations of the model. The “frontoparallel” configuration (Perrone and Stone, 1994) samples depth at five logarithmically spaced frontoparallel reference planes ($z = 2, 4, 8, 16, \text{ and } 32 \text{ m}$) such that there are five sensor inputs to each detector at each location with preferred velocities determined using Equation 3 and the five values of z stated above (Fig. 3*A*). Because translational flow falls off quickly with distance, the 32 m upper limit was shown to be adequate. Furthermore, although the sampling is based on frontoparallel planes and an observer speed of 1 m/sec, the interaction between the sensors associated with the different reference planes allows the model to respond well to arbitrary scene geometries and a range of observer speeds.

Although the original depth sampling was didactically convenient, it is an inefficient strategy for real-world layouts because it is completely unbiased (i.e., designed to handle arbitrary and even discontinuous layouts such as clouds of points). Frontoparallel reference planes do not optimally sample the range of depths encountered as a primate walks or

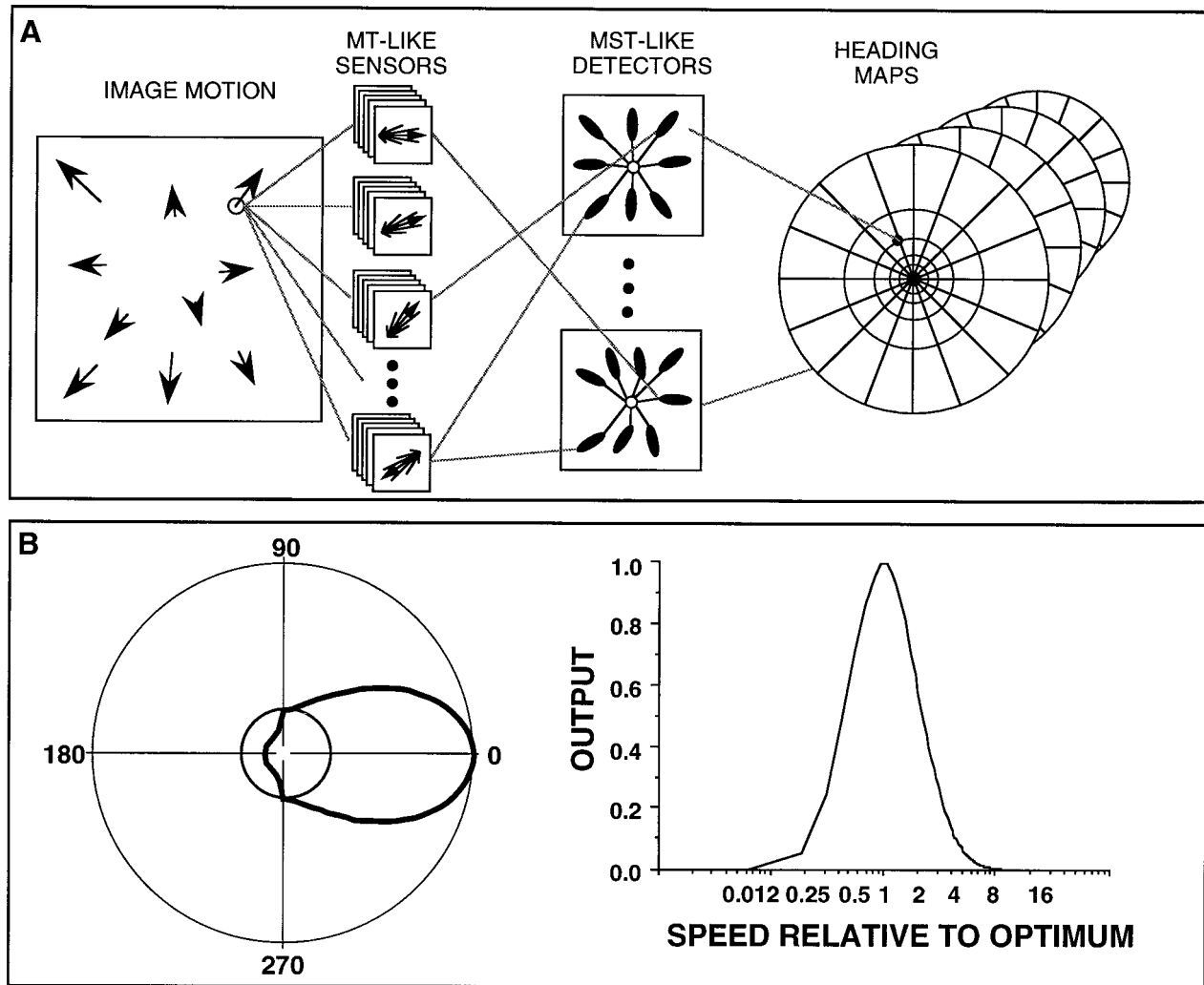


Figure 2. The template model. *A*, The overall structure is shown. Image motion is analyzed using sets of speed- and direction-tuned MT-like motion sensors tiling the entire visual field. *B*, If we assume separability, the output of each sensor is given by the product of the direction (*left*) and speed (*right*) responses. Note that the direction response can be negative and that zero is indicated by the *inner black circle*. The output of specific sets of these sensors are then summed over a wide portion of the field by MST-like detectors (see illustration in Fig. 3). Because of the specificity of the sensor to detector connectivity defined by Equation 3, the detectors are each “tuned” for a particular heading. Heading maps containing arrays of detectors are used to sample heading space as shown in *A*. The detector with the largest output within all of the maps identifies heading. For clarity, only a small subset of the connections is shown in *A*.

runs along the ground. Rather than systematically sampling the whole range of depths at all locations, a different approach is to sample a different restricted range of depths at each location according to the reasonable expectation of depth variation with position. Although this approach makes assumptions about the layout and therefore loses some generality, this loss should be primarily inconsequential because primates do not generally navigate in clouds of random points. They typically encounter environments with systematic statistical covariation of depth with location in the visual field; points are generally closer directly below and farther away in front of the observer.

We therefore designed an alternative parametric configuration of the model that samples depth values that coincide with a ground plane (Fig. 3*B*). Because the inputs to a detector (its receptive field) must be defined in retinocentric coordinates, for simplicity the frontoparallel reference planes were fixed in retinocentric coordinates. However, the ground is by definition fixed in exocentric coordinates, and therefore ground-plane sampling only makes sense in exocentric coordinates. Fortunately, if we assume that the observer’s path through the world is parallel to the ground (i.e., the observer is neither flying nor falling), the necessary exocentric-to-retinocentric coordinate transformation of the layout becomes straightforward. A “ceiling” plane is included to allow for points that lie above the horizontal meridian. As a first cut, we only used one

plane above and one below the line of sight (at ± 1 m), but the number of sensors at each location feeding each detector does not seem critical for the properties tested here and could easily be increased. We must emphasize that the “ground” configuration is merely an attempt to sample the environment in a more ecologically relevant manner and to examine the possible consequences. Although it is more ecologically defensible than is the frontoparallel configuration, it is an extreme and rigid instance of this approach. The actual depth sampling in MSTd would more likely be a compromise between the two configurations, tailored to the actual depths most often encountered during self-motion in the real world and set up via learning through experience.

The rotation rate (ω_s) is naturally logarithmically compressed under normal viewing conditions (reasonably distant and central gaze). In the frontoparallel configuration (set in 1994) only four levels (0, 1, 2, and 4°/sec) are used, corresponding to four heading maps, because this range spans most situations [see Perrone and Stone (1994), their Fig. 4]. We did not include templates tuned to backward observer motion (contraction-tuned detectors). Physiological studies have revealed however that many MSTd neurons respond best to contraction (e.g., see Fig. 4*C*), although the percentage preferring contraction over expansion is $\sim 20\%$ (see Fig. 12*A*). There is also psychophysical evidence that humans do not process expansion and contraction equivalently; object speed and depth relation-

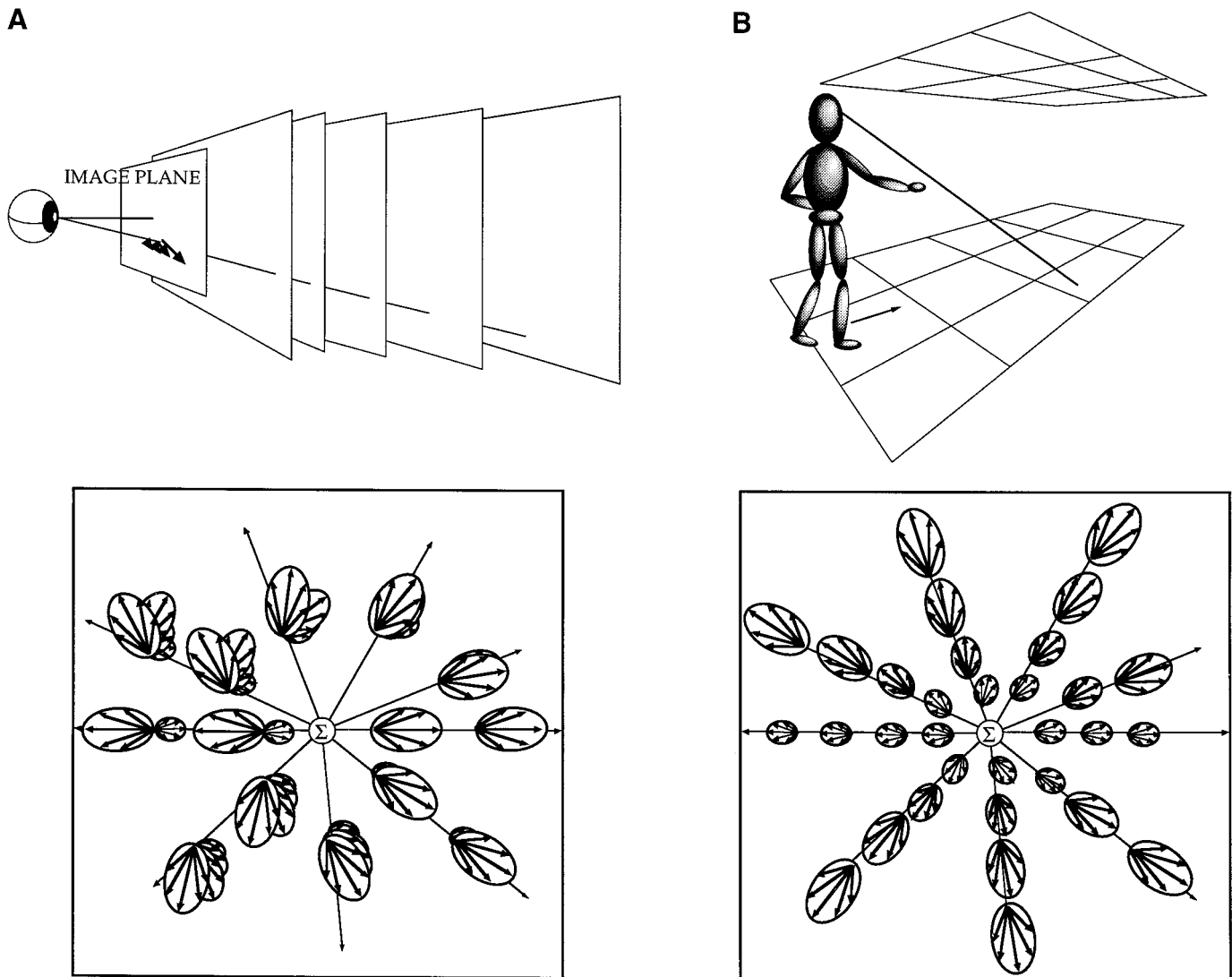


Figure 3. Depth-sampling strategies for the two configurations of the model. *A, Top*, The depth sampling used in the frontoparallel configuration is based on five reference planes, orthogonal to the line of sight. *Bottom*, The corresponding detector architecture is flower-like with five different-sized petals at each location but no spatial gradient. *B, Top*, The depth sampling used by the ground configuration is based on ground and ceiling planes, exocentrically located 1 m above and below eye height and clipped at 32 m. *Bottom*, The corresponding detector architecture is flower-like with a systematic spatial gradient of petal size but only one petal per location.

ships are misperceived during simulated backward self-motion (Perrone, 1986). These physiological and psychophysical data suggest that the “backward” direction may be represented by fewer neurons than the forward direction. In the ground configuration, we have therefore included a single additional map of detectors tuned to backward headings, a set of pure contraction templates opposite in tuning to the pure expansion templates of the original 0°/sec map. We did not include any backward motion templates with rotation.

Every simulation begins with an input velocity vector field (e.g., Fig. 1*A*) that matches as closely as possible the stimulus conditions used in the particular physiological study being examined. To simulate the response of a detector, we assume that MT-like input sensors are located at the position of each of the input flow vectors (i.e., MT is assumed to sample the visual field finely). The output of each detector is derived first by calculating the preferred velocity of each of the sensors using Equation 3 and then by determining the sensor response using Equations 1 and 2. The maximum sensor outputs at each location (winner-take-all) are summed to produce the total detector output. Heading is reported by the most active detector within the heading maps. In many of the figures, we normalized the detector output by dividing the raw output by the maximum possible output, which is simply equal to the number of stimulus points.

It is critical to note that although the model begins with a vector representation of the stimulus, it is the MT-like sensor responses that are used to determine the detector output. Because of the lack of a biologically plausible model of MT responses, Equations 1 and 2 are used as a convenient way to generate the MT-like input signals. Local motion sensors with MT-like responses generated directly from image sequences are currently being developed (e.g., Perrone, 1994). This new front end will obviate the need for vector flow-field inputs. Nonetheless, regardless of how the MT-like responses are generated, the true inputs to the model heading detectors are sensor outputs consistent with MT data, not velocity vectors.

Each map samples heading space at values of 0, 3, 6, 9, 12, 15, 18, 21, 26, 36, 56, and 89.5° in the radial direction (i.e., along the length of spokes) for axial directions (i.e., spoke orientations) ranging from 0 to 360° in 15° steps. In this polar layout, the radial and axial values do not correspond directly to azimuth and elevation; therefore the detectors are rarely tuned for integral values of heading azimuth and elevation. The frontoparallel configuration has a total of 1152 detectors within its four maps. The ground configuration has a total of 1440 within its five maps. Throughout the rest of this paper, we adopt the shorthand notation $\pm(\alpha, \beta, \omega_o)$ to refer to a detector tuned to a heading of azimuth α , elevation β , and rotation rate ω_o (with the negative sign indicating backward

headings). For the simulations, “receptive field” size was set to $100^\circ \times 100^\circ$, and random samples were taken using uniform probability across the full set of detectors.

Although some of the model parameters are based on known physiological properties of primate neurons (e.g., input sensor bandwidths), some of the parametric choices were unconstrained (e.g., five reference planes). We have no attachment to the latter speculative parametric choices. We must emphasize that although the depth-sampling parameters are different for the two configurations, both were constrained to a fixed set of parameters for all of the simulations, and the stimuli used to test the two were identical. Details of the simulations for specific tests are given in the Results.

The template model was designed to use MT-like inputs to solve the self-motion problem and was not explicitly designed to have its output detectors mimic MST neurons. The “physiological” properties of the detectors are therefore truly emergent, and the tests performed below represent an independent evaluation of the inner workings of the model beyond our previous analyses of its overall performance (Perrone, 1992; Perrone and Stone, 1994). It is also interesting to note that much of the neurophysiological data shown here only became available after the template model was developed, so the simulations of the frontoparallel configuration actually represent a priori predictions rather than a posteriori fits to a known database.

RESULTS

Selectivity for optic-flow components

Several groups have examined MST neuronal responses to large flow-field stimuli representing a basic set of possible observer movements: forward and backward translation, rightward/leftward/upward/downward translation, and clockwise and counterclockwise roll around the line of sight. The resulting optic flow patterns are expansion and contraction, left/right/down/up planar motion, and clockwise and counterclockwise circular motion, respectively. Using this “canonical” set of stimuli, Duffy and Wurtz (1991a) found that MSTd neurons typically respond to a range of these “flow components” with some neurons responding selectively to only one and others to two or even three components. Similar data can also be found in other studies (e.g., Tanaka and Saito, 1989; Lagae et al., 1994). The advantage of the Duffy and Wurtz data is that they used the largest stimuli such that one can be reasonably certain that nearly the entire receptive field was stimulated. In studies that use small test patches, any apparent selectivity cannot be dissociated from the effects of suboptimal centering of the stimulus in the receptive field.

Duffy and Wurtz (1991a) used $100^\circ \times 100^\circ$ stimuli containing 300 moving dots. In the planar stimuli, each dot moved at $40^\circ/\text{sec}$, and in the circular and the expansion and contraction stimuli, the average speed was $40^\circ/\text{sec}$. The expansion and contraction stimuli simulated motion toward and away, respectively, from a vertical dot plane 100 cm from the eye. We used input flow fields that matched these instantaneous motion parameters. Because the direction tuning curves of the model sensors incorporate a small amount of inhibition for antipreferred motion (see Fig. 2*B*), the total output of the template can be negative. We set any such negative values to zero and did not attempt to mimic the spontaneous activity levels found in the no-dots control condition. The absolute response levels to the different input flow fields are not especially relevant. It is the pattern of responses to the set of stimuli that is important.

Figure 4 shows the results of template-model simulations along with MST data from Duffy and Wurtz (1991a). Both configurations of the model yielded similar results. Figure 4, *B*, *D*, and *F*, shows responses of ground detectors (for an example response set with a frontoparallel detector, see Perrone and Stone, 1994). Figure 4*A* shows the response of an MSTd neuron (53XL24) that preferred planar motion to the left. The response to all other

patterns of motion was close to or below the spontaneous level. Figure 4*B* plots the outputs of the model detector tuned to $(89.5^\circ, 0^\circ, 0^\circ/\text{sec})$. Like the neural response in Figure 4*A*, the detector output for leftward planar motion is high with little or no output for the other stimuli. Figure 4*C* shows a radial cell (53XL60) that prefers contraction over expansion and does not respond to other flow components. Figure 4*D* illustrates the responses of the detector tuned to $(-2.6^\circ, 1.5^\circ, 0^\circ/\text{sec})$ that also responds nearly exclusively to contraction. Figure 4*E* shows the responses of a planoradial neuron that responded well to both rightward planar motion and radial expansion. Figure 4*F* shows a similar pattern of responses arising from the model detector tuned to $(-36^\circ, 0^\circ, 0^\circ/\text{sec})$. The model detectors can therefore simulate the behavior of three of the response types (planar, radial, and planoradial) identified by Duffy and Wurtz (1991a).

There are three MSTd neuron types found by Duffy and Wurtz (1991a) and others (e.g., Tanaka and Saito, 1989; Orban et al., 1992) whose existence is not explained by either the frontoparallel or ground configurations: circular, planocircular, and planocircular-radial. For example, one neuron [Duffy and Wurtz (1991a), their Fig. 6*B*, 53XL70] responded only to counterclockwise roll motion with little or no responses to other motion types. The two configurations of the template model tested in this paper do not have detectors tuned to pure roll because they both are constrained to handle only self-motion scenarios under gaze stabilization. We argued that, because of the various gaze-stabilization mechanisms, circular flow during self-motion was minimized and a reduction in template numbers could be achieved by not incorporating “roll” detectors. Thus, the lack of detectors with significant circular responses does not result from some fundamental incompatibility with the template approach but rather from the gaze-stabilization constraint. This constraint could be relaxed to allow the inclusion of roll detectors as was true for the unrestricted version of the template model (Perrone, 1992).

Decomposition

One of the main features separating template models from earlier self-motion models is the fact that they do not rely on the decomposition of optic flow into translational and rotational fields. If MST neurons behaved like the processors in full decomposition models (e.g., Rieger and Lawton, 1985; Heeger and Jepson, 1992; Hildreth, 1992; Royden, 1997), one would expect, for example, that the vector addition of rotation to an expanding stimulus would have no impact on the output of an expansion-tuned MST neuron. Decomposition models go to great lengths to design heading (radial) responses that are immune to rotation. However, in a direct test of the decomposition hypothesis, Orban et al. (1992) showed that MST neurons, like templates, are not immune to the vector addition of nonpreferred flow.

Orban et al. (1992) tested the responses of a variety of MST cells by systematically adding varying amounts of a nonpreferred flow component to the preferred flow stimulus (their Fig. 2*C*). For example, a neuron that preferred clockwise rotation would be stimulated with combinations of rotation along with a certain proportion of expansion or contraction. These combinations were expressed as ratios of the amplitude of the preferred component to that of the nonpreferred component. We simulated their experiment using 25.5° diameter patches of flow field consisting of 126 vectors with an average speed of $4.4^\circ/\text{sec}$ for the pure expansion stimuli. Roll was added vectorially to the expansion pattern with both centers of motion at the patch center.

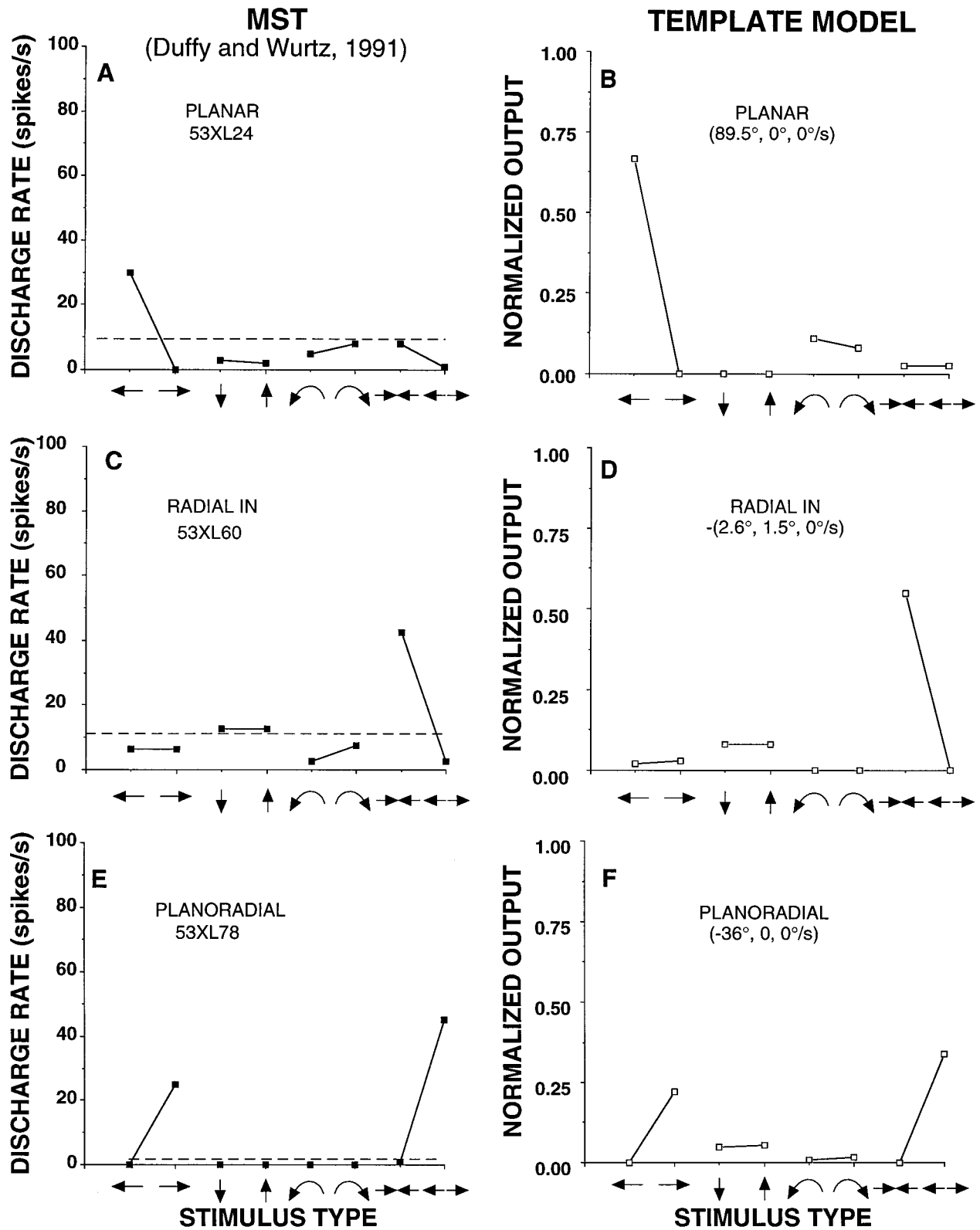


Figure 4. Selectivity test. *A, C, E*, Replots of data from Duffy and Wurtz (1991a, their Fig. 6*A, C, E*). The symbols along the horizontal axis represent the eight canonical stimuli described in the text. The horizontal dashed lines correspond to the mean activity for their no-dots control condition. *B, D, F*, The responses of ground detectors to the same set of stimuli. The detector examples were picked by surveying the population for qualitative matches to the example neurons.

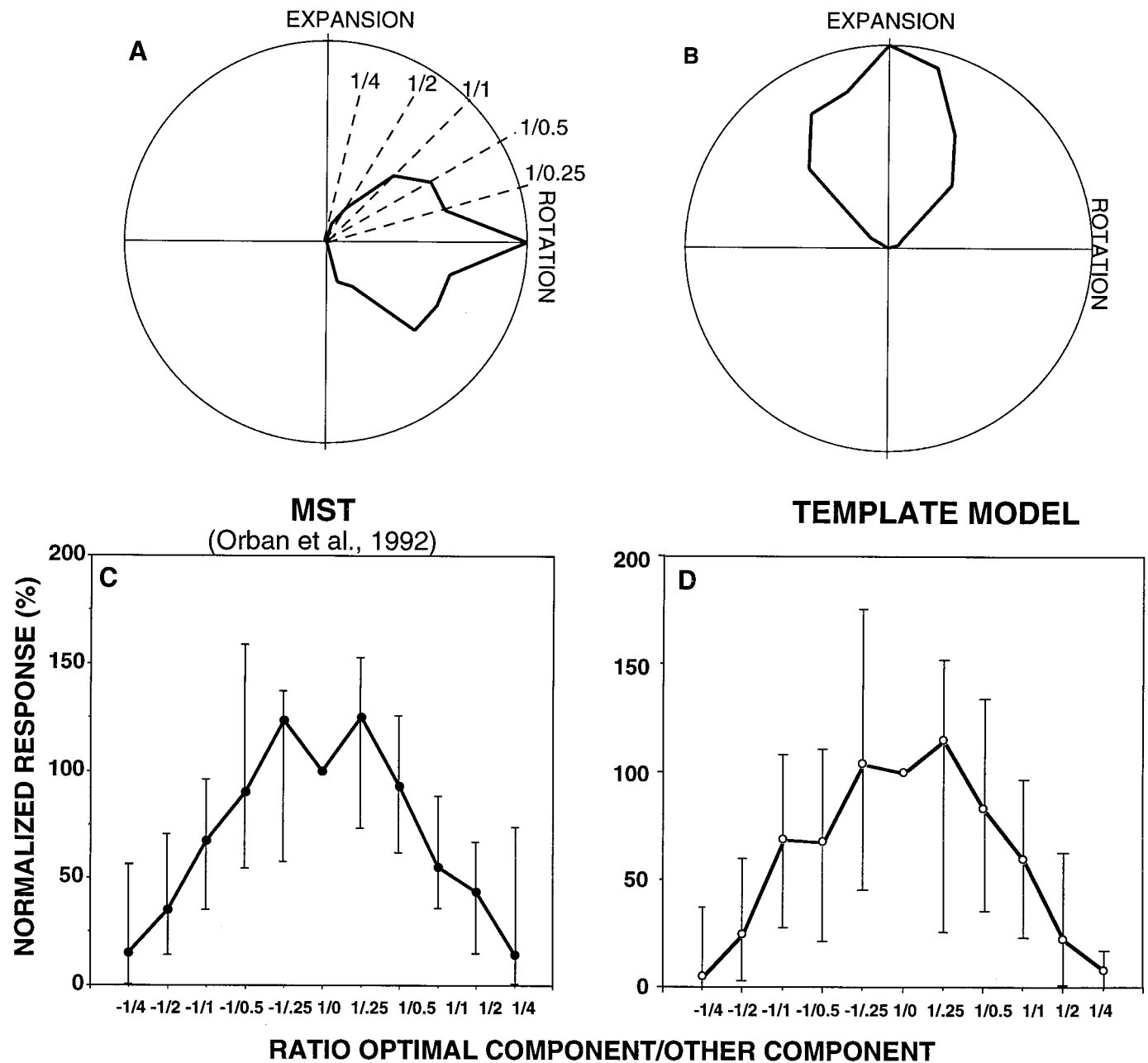


Figure 5. Decomposition test. *A, C*, Data replotted from Orban et al. (1992, their Fig. 2*A,C*) that show responses of MST neurons to combinations of two types of image motion. The *sign* of the amplitude ratio indicates the direction of the nonpreferred component. The amplitude at polar angle 0° represents the response to pure roll, and that at 90° represents the response to pure expansion. Intermediate angles indicate the responses to vector combinations of the two. *B*, The responses of the ground detector tuned to (3°, 0°, 2°/sec). The effect of nonpreferred flow on the response of this arbitrarily chosen detector is however typical. *C, D*, The normalized median responses of seven MST neurons (*C*) and that for 50 detectors (*D*). The error bars in *C* and *D* indicate the quartiles.

Figure 5*A* replots the responses of one of the MST neurons (4207) from Orban et al. (1992). This polar plot has its axial angle corresponding to different ratios of the preferred component (clockwise roll) to the nonpreferred component (expansion) and its radial amplitude corresponding to the normalized response. Note that the progressive addition of nonpreferred flow weakens the response and ultimately drives it to negligible levels. Figure 5*B* illustrates that individual model detectors show the same behavior: nonpreferred flow interferes with the response. This is true regardless of the depth-sampling configuration. It reflects the basic property of a unit in a template model as opposed to one in

a decomposition model. The model data (Fig. 5*B*) were obtained from the ground detector tuned to (3°, 0°, 2°/sec) that prefers expansion, and so it does not exactly match the only raw data example shown by Orban et al. (1992), a clockwise roll-tuned neuron (Fig. 5*A*). However, the preferred component of the detector or neuron is not critical. The point is that MST neurons do not seem able to decompose optic flow. The data in Figure 5 suggest that, regardless of whether MST neurons are involved in heading perception, they act like optic-flow templates and not like flow-decomposition units.

Figure 5*C* replots the median response of the seven MST

neurons in Orban et al. (1992). Although both configurations yielded similar results, Figure 5D shows the median model response from a sample of 50 randomly selected detectors from the ground configuration. The horizontal axis represents how much of the nonpreferred component (as a ratio of the amplitude of the preferred component) was present, and the vertical axis is the normalized output. Although both the neural and model data exhibit high variability, neither MST neurons nor model detectors (regardless of the choice of depth-sampling configuration) are immune to the addition of nonpreferred flow. Sensitivity to non-preferred flow is an explicit property of template models and is therefore not truly emergent; however the *quantitative* nature of the sensitivity is emergent because the detectors were not designed to generate the *curve* shown in Figure 5D.

Complete decomposition models predict that heading units will be immune to nonpreferred flow. The data of Orban et al. (1992) therefore suggest that MST cannot be implementing a complete decomposition of optic flow. Lappe and Rauschecker (1993) proposed a partial decomposition model [based on the Heeger–Jepson (1992) decomposition algorithm] that incorporates units that are immune to the rotational flow generated during gaze stabilization but not to other forms of rotation. The Orban data do not rule out such partial decomposition models. However, testing expansion-preferring MST neurons with expansion plus added roll around the line of sight (rather than around the receptive-field center), or other added rotation inconsistent with gaze stabilization, would resolve this issue.

Spatial integration

One of the basic characteristics of the template model is that two-dimensional (2D) motion information is integrated over a large area of the visual field. The detector that best matches the stimulus determines the heading estimate, and so, generally, the larger the integration area, the higher the signal-to-noise. Because the detectors summate their inputs over space, a change in stimulus size will change their output. We simulated one of the experiments in Tanaka and Saito (1989) in which they compared the response to expanding stimuli displayed in 20, 40, and 80° diameter circular windows. We used 300 randomly distributed vectors in the largest window. The density of vectors was constant across conditions; the largest window had more vectors than did the smallest, consistent with the Tanaka–Saito stimuli.

Figure 6A replots the responses of one of the MSTd neurons (kl215.1) of Tanaka and Saito (1989) for three stimulus sizes. The neural response increases with increasing test-patch size, consistent with the view that the neuron is integrating information over a large part of the field. Figure 6B shows the results for the ground detector tuned to (4.0°, 14.5°, 4°/sec) for the same three stimulus sizes. All detectors from both depth-sampling configurations show similar qualitative behavior; increasing the stimulus size increases the output, consistent with the general finding that larger stimuli generate greater responses in MST cells [Tanaka and Saito (1989), their Fig. 3E1; Duffy and Wurtz (1991b), their Fig. 4A; Lagae et al. (1994), their Fig. 20D].

Although the physiological data generally support the view that MST neurons integrate information over large portions of the visual field, not all MST neurons show a monotonic increase in response with stimulus size. MST responses can show saturation [e.g., Duffy and Wurtz (1991b), their Fig. 4B; Lagae et al. (1994), their Fig. 20C] or even a fall-off in output with increased stimulus size [e.g., Duffy and Wurtz (1991b), their Fig. 4C]. Duffy and Wurtz (1991b) found complex interactions in neuronal responses

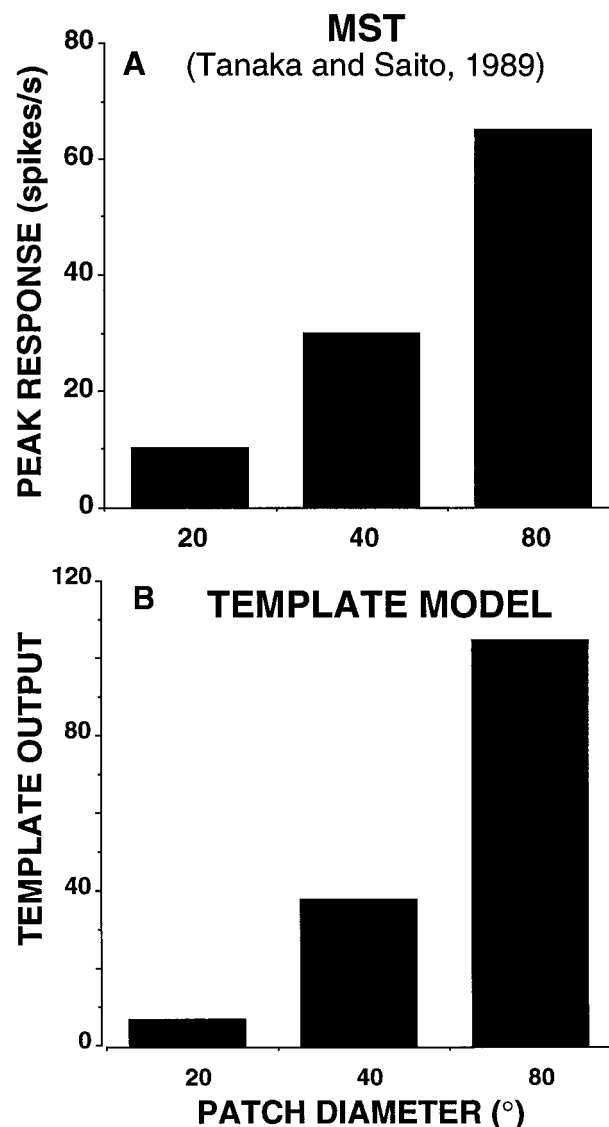


Figure 6. Spatial integration. *A*, The response of MST neuron kl215.1 [Tanaka and Saito (1989), their Fig. 3E1] as a function of stimulus size. *B*, The response of the ground detector tuned to (4.0°, 14.5°, 4°/sec). The effect of stimulus size on the response of this arbitrarily chosen detector is typical of the population.

as they decreased the size of the stimulus patch and changed its location in the receptive field. Many cells seem to exhibit “non-homogenous” response profiles suggesting the existence of inhibitory subregions. Lagae et al. (1994) also found evidence of response selectivity that could not be explained by simple summation.

Although integration over large areas generally offers advantages in terms of increased signal-to-noise, there is a point where the extra information gained is small relative to the additional “noise” generated by low-signal regions of the stimulus. For detectors tuned to expansion, the 2D motion sensors located far from the FOE tend to be tuned to the same direction over large portions of the visual field (see peripheral regions of Fig. 1A). For example, the peripheral sensors feeding the detector tuned to (5°, 0°, 0°/sec) are virtually identical to those feeding the detector tuned to (10°, 0°, 0°/sec), so they provide little information distinguishing these two possible headings. Koenderink and van

Doorn (1987) have presented a mathematical derivation of the fall-off in information with distance from the FOE. Warren and Kurtz (1992) and Crowell and Banks (1993) have verified that this phenomenon applies to human self-motion judgments. Crowell and Banks (1996) have also used an ideal observer model to determine regions of the visual field that contain the most information for self-motion estimation. Such an analysis could be used to optimize the region of visual field feeding into the detectors. For simplicity, our model treats the input field homogeneously. For some detectors, it would be beneficial to restrict the receptive field to specific subregions of the visual field to optimize the signal-to-noise. Furthermore, if the detector receptive fields are not of equal size because this optimization caused different amounts of the visual field to be processed by different detectors, then some form of gain control would be required to keep their relative activity meaningful. Such gain control may contribute to the saturation or even reduction of the response as a function of stimulus size observed in some MST neurons.

Center-of-motion tuning

In a direct test of the heading-detector hypothesis, Duffy and Wurtz (1995) recently examined the effect of the location of the center-of-motion (COM) of optic-flow stimuli on MSTd responses. In the case of expansion, this amounts to moving the FOE while subtending the same portion of the visual field. In our model, a detector tuned to $(\alpha, \beta, 0^\circ/\text{sec})$ will respond maximally for expansion with its COM in the (α, β) direction. If the COM is shifted away from this direction, the output of the detector will fall. MSTd neurons must express this behavior if individual neurons encode heading directly as proposed in the template model.

The stimuli were designed to mimic those used by Duffy and Wurtz (1995). Eight of these stimuli were planar motion in eight possible directions (which is equivalent to expansion with a COM 90° from fixation). Eight were pure expansion with a COM at 45° eccentricity along the primary oblique axes. Eight more stimuli were pure expansion with their COM at 22.5° eccentricity along the primary oblique axes. The final stimulus had its COM at the center of the field. Duffy and Wurtz (1995) referred to the 22.5° eccentricity stimuli as “pericentric,” the 45° stimuli as “eccentric,” and the 90° stimuli as “peripheral.” There were 360 randomly placed vectors in each stimulus, representing motion at 3.6 m/sec toward a single plane of points located 4 m from the eye and producing an average speed close to the $40^\circ/\text{sec}$ used by Duffy and Wurtz (1995).

Figure 7A replots the data for an MSTd neuron (26KR43). Figure 7B is a radial slice through its preferred axial direction ($\sim 180^\circ$). We fitted a circularly symmetric 2D Gaussian to the data to determine the preferred COM location (x and y shifts), the SD ($\sigma = \sigma_x = \sigma_y$), and the measure of goodness-of-fit (r). For this neuron, the preferred COM location (focus of contraction) was estimated at $(-36^\circ, 6^\circ)$. The SD (bandwidth) of the fitted Gaussian was found to be 31° ($r = 0.96$). Figure 7, C and D, shows the normalized outputs from the frontoparallel detector tuned to $(-33^\circ, 6^\circ, 0^\circ/\text{sec})$. No detector heading within our set exactly matched the preferred COM of the neuron, so we selected the nearest one that produced the best fit to the data. As discussed above, the frontoparallel configuration does not have detectors tuned to contraction, so we tested its COM tuning with the equivalent expansion stimuli. The fitted preferred COM location of the detector was $(-31^\circ, 6^\circ)$, and the estimated bandwidth was 33° ($r = 0.98$). Figure 7, E and F, shows the normalized outputs of the ground detector tuned to $(-33^\circ, 6^\circ, 0^\circ/\text{sec})$. The fitted

preferred COM location of this detector was $(-29^\circ, 8^\circ)$, and the bandwidth was 31° ($r = 0.98$). Both the MSTd and model data are well fit by a 2D Gaussian. Thus, Duffy and Wurtz (1995) found MSTd neurons tuned for a particular COM location with tuning properties *quantitatively* consistent with the template model, with little difference between the two depth-sampling configurations. COM tuning for radial stimuli is expected of template-model heading detectors (although the preferred COM exactly coincides with the preferred heading only for the pure-translation detectors). However, the bandwidth and shape of this tuning are emergent and remarkably close to those of MSTd neurons. The similarity of the peaked responses of both the model and neurophysiological responses lends support to the conclusion of Duffy and Wurtz (1995) that MSTd neurons could form a population, with each neuron tuned to a different heading and performing a role similar to template-model detectors. Such COM tuning is a fundamental property of our model and distinguishes it from the units predicted by the Lappe and Rauschecker (1993) model. Their model predicts that MST units will show sigmoidal response tuning only along a specific one-dimensional (1D) axis and no variation along the other axis. Because this feature is a key difference between the two models, we now examine this issue more closely.

Gaussian or bell-shaped tuning is incompatible with the sigmoidal-tuned units proposed by Lappe and Rauschecker (1993). Their sigmoidal model units do not show a response peak for a particular COM location but rather show broad regions over which their response is primarily invariant. Units with sigmoidal tuning cannot produce plots like those shown in Figure 7. To illustrate this point, we stimulated a sigmoidal unit (integral of a Gaussian with $\sigma = 40^\circ$) with the COM stimulus set. If a broad sigmoidal function is used in the Lappe and Rauschecker (1993) model, bell-shaped tuning can occur along the axial direction, but the bandwidth will change systematically with eccentricity (Fig. 8A). Bell-shaped tuning is however not possible along the preferred radial direction (Fig. 8B), as is seen in MST neurons (Fig. 7B).

Duffy and Wurtz (1995) found that most of the neurons in their sample (55% of $n = 142$) were tuned to either eccentric or central COMs and therefore showed a clear peak in their response profiles. More recently, Lappe et al. (1996) claimed that peaked tuning is rare in MST (8% of $n = 134$). They argued that the majority of neurons have sigmoidal tuning in accord with the basic mechanism of their model and provided some supporting physiological evidence for their view. However, they tested their neurons with COM locations only out to 40° eccentricity and truncated the individual data plots to $\pm 30^\circ$. Examination of Figure 7B reveals that within this limited range of eccentricities (dashed vertical lines), the Duffy and Wurtz (1995) data would be mistaken for sigmoidal. If tested over a sufficiently wide range of eccentricities, more of the neurons of Lappe et al. (1996) may have revealed bell-shaped tuning, and this would have brought the relative proportions more in line with those of Duffy and Wurtz (1995). In addition, the coarse averaging over neurons that Lappe et al. (1996) performed, after aligning the preferred axes to within $\pm 22.5^\circ$, will blur the 2D structure of the receptive fields and tend to make the average look sigmoidal even if the individual neurons were peaked. This possibility is further supported by the fact that the only raw data example of a “sigmoidally tuned” expansion neuron shown (Lappe et al., 1996, their Fig. 7) shows a dip at the edge of their plot (at 30° eccentricity), suggesting that it may actually have had peaked tuning.

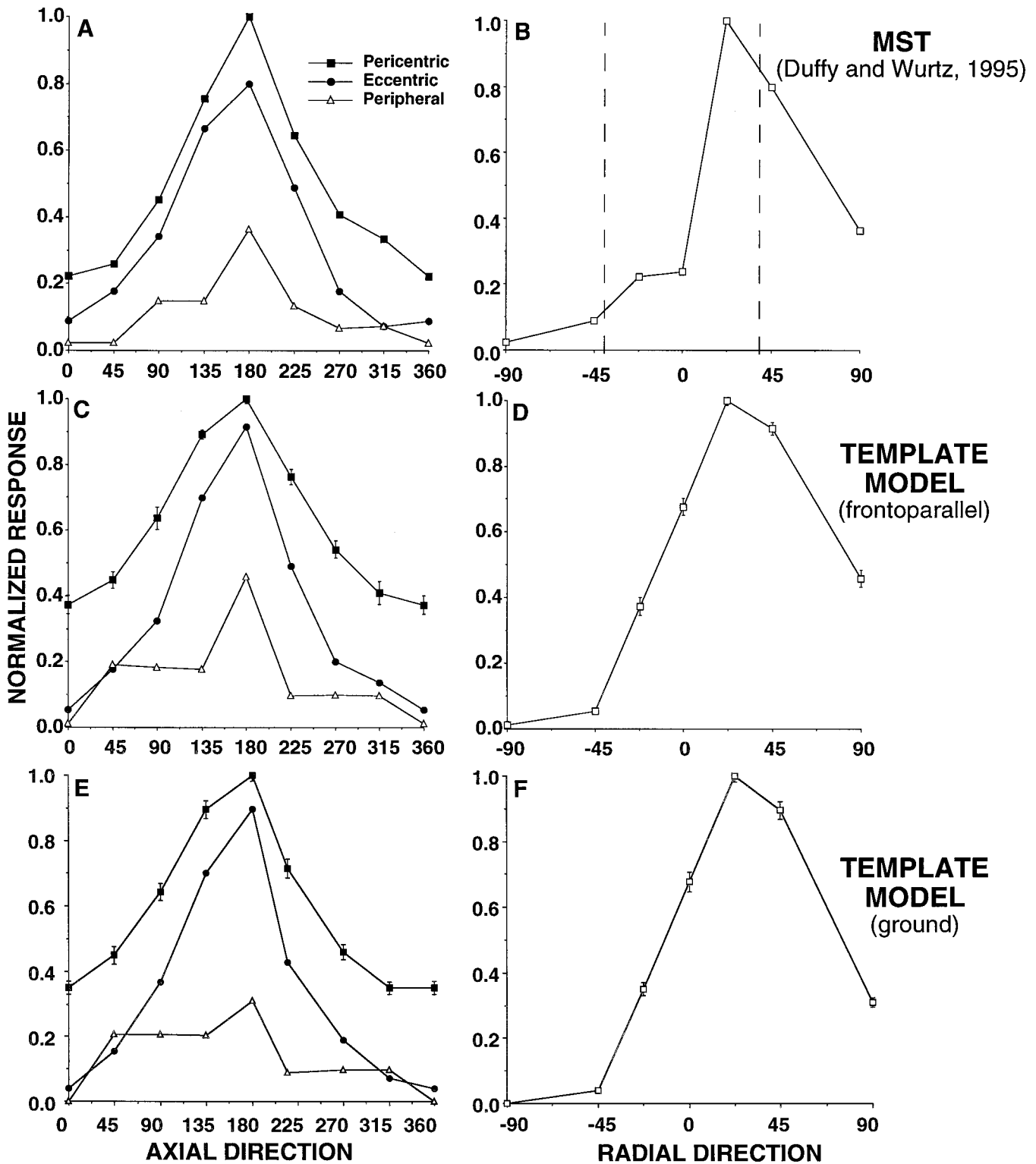


Figure 7. Center-of-motion (COM) tuning test. *A*, Replotted normalized responses of an MSTd neuron from Duffy and Wurtz (1995, their Fig. 12*B*). Pericentric stimuli (solid squares) had their COM 22.5° out from fixation, and eccentric stimuli (solid circles) had their COM 45° out from fixation. Peripheral stimuli (open triangles) were planar stimuli with extrapolated COMs at 90°. *B*, A plot of the response of the neuron along its preferred axis (~180°). Vertical dashed lines indicate the range of test conditions used by Lappe et al. (1996). *C*, A plot of the normalized mean responses of the frontoparallel detector tuned to (-33°, 6°, 0°/sec) an equivalent set of stimuli. *D*, A plot of the response of the same detector shown in *C* along its preferred axis (~180°). *E*, A plot of the normalized mean responses of the ground detector tuned to (-33°, 6°, 0°/sec) to the same set of stimuli as in *A* and *C*. *F*, A plot of the response of the same detector shown in *E* along its preferred axis (~180°). Error bars on the pericentric data in *C* and *E* and on the data in *D* and *F* represent the SD across 12 simulation runs (with different randomly located input flow vectors). The detectors in *C*-*F* were chosen because they provided the best fit from among those in the 0°/sec map whose preferred heading was close to the preferred COM of the neuron shown in *A* and *B*.

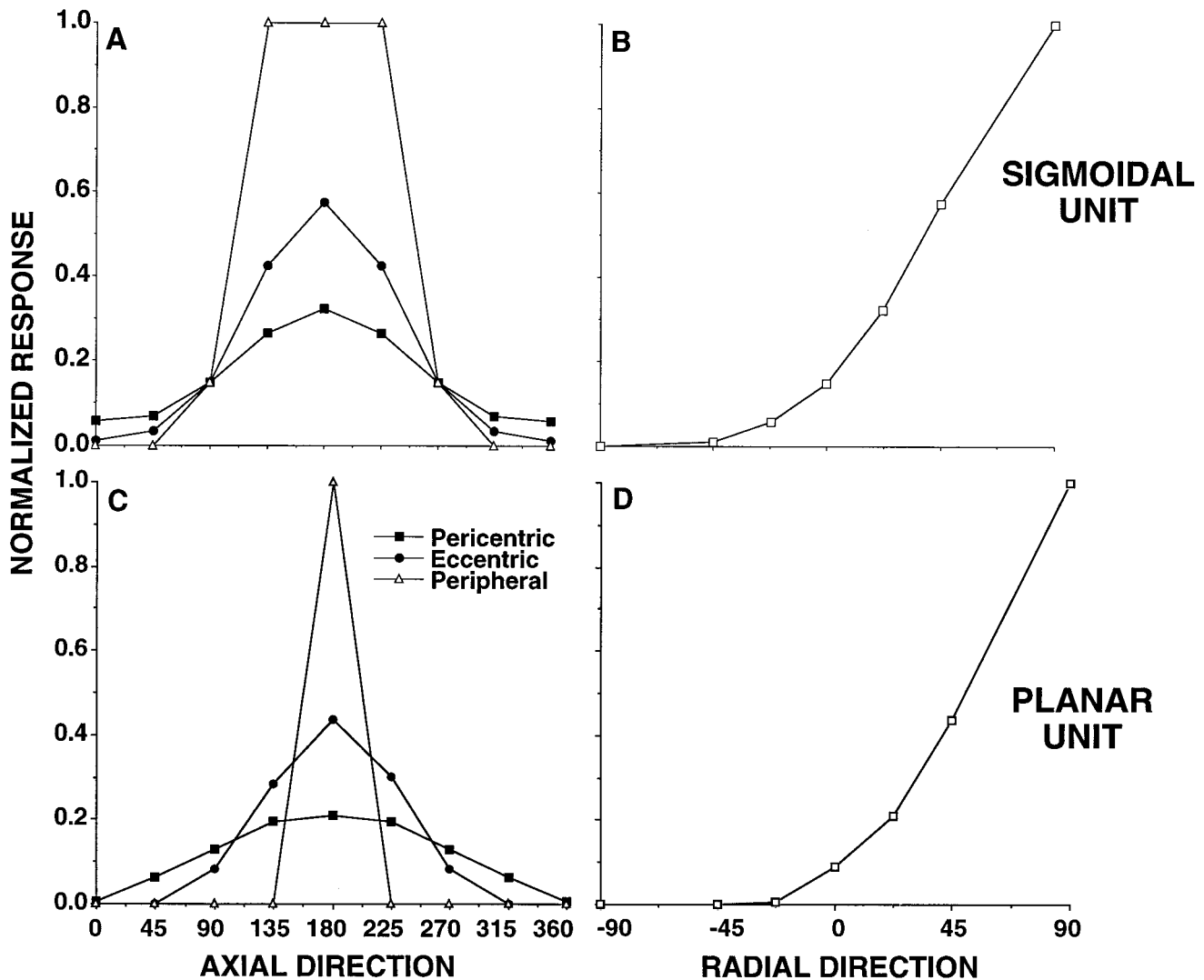


Figure 8. Center-of-motion tuning of a planar and sigmoidal unit. *A, B*, The responses of a unit with a sigmoidally tuned expansion response to the same stimuli described in Figure 7. *C, D*, The responses to the same stimuli of a template-model detector tuned to *rightward* planar motion.

It could be argued that the observed COM or heading tuning of MSTd neurons found by Duffy and Wurtz (1995) is an artifact of their stimulus paradigm. Specifically, the tuning might be trivially explained by a large receptive field tuned to a single planar direction. To test this possibility, we ran the COM simulations using the ground detector tuned to planar (unidirectional) motion (89.5°, 0°, 0°/sec) (Fig. 8*C, D*). Although planar units can show changes in their responses with shifted COMs, they generate a qualitatively different pattern of results from those seen with the MSTd neuron shown in Figure 7, *A* and *B*. The radial tuning appears sigmoidal with the peripheral stimuli generating the greatest output (Fig. 8*D*), and the widths of the axial tuning curves change systematically with eccentricity (Fig. 8*C*). The sigmoidal response in Figure 8*D* relies on the fact that the planar detector, like MSTd planar neurons (Fig. 12*A*) (Duffy and Wurtz, 1991a), is broadly tuned for speed. If the planar units were narrowly speed tuned, it would be possible to generate 1D bell-shaped tuning along the radial direction. However, the axial direction tuning would remain inconsistent with that of the MSTd neuron shown in Figure 7*A*.

The simulations in Figure 8, *C* and *D*, show that planar-tuned

units will produce a different pattern of results from that described by Duffy and Wurtz (1995). This argues that their data replotted in Figure 7, *A* and *B*, are from a heading-tuned neuron and not simply a planar-tuned neuron. Furthermore, because the template-model planar detector produces a sigmoidal expansion response curve, sigmoidal tuning is therefore not a unique signature of the Lappe and Rauschecker (1993) model. The template model can explain both bell-shaped and sigmoidal responses within its population of detectors at approximately the ratio found by Duffy and Wurtz (1995). The Lappe and Rauschecker (1993) model, however, must add a third layer of units (Lappe et al., 1996) to explain bell-shaped MST responses.

Position invariance

In some studies of MST, many neurons retained their selectivity for a particular stimulus even when the stimulus was moved to different locations in the receptive field (Duffy and Wurtz, 1991b; Orban et al., 1992; Graziano et al., 1994; Lagae et al., 1994). This was taken as evidence that individual MST neurons, which respond selectively to expansion patterns, could nonetheless not be used to encode the location of the FOE, and hence heading, in

any straightforward way. The difficulty in reconciling the position invariance of MST neurons with the fact that template-model detectors are individually tuned to a specific heading has been a serious obstacle to the acceptance of the view that MST neurons may directly encode heading and act as heading templates. However, more recent studies (Duffy and Wurtz, 1995; Bradley et al., 1996; Lappe et al., 1996) as well as some earlier results (Duffy and Wurtz, 1991b) show that MST neurons do not show strict invariance and in fact possess properties consistent with individual neurons encoding heading (see previous section). In this section, we test the model under the conditions examined by Graziano et al. (1994) and Duffy and Wurtz (1991b) to see whether its detectors exhibit the limited position invariance observed in MST.

We simulated the experiments of Graziano et al. (1994) using a clover leaf arrangement of five circular test patches, each 10° in diameter, spanning a distance of 20° vertically and horizontally (5° of overlap between patches). Each patch consisted of 126 points moving at a mean speed of $4.4^\circ/\text{sec}$ in an expansion or contraction pattern. Directional selectivity (DS) was defined in the usual way [$DS = 1 - (\text{response to antipreferred stimulus} / \text{response to preferred stimulus})$]. A neuron or detector that is very selective will have a DS close to or >1.0 . The DS was determined at the five different patch locations. As did Graziano et al. (1994), we took the DS at each surrounding position and divided it by the DS at the central position to derive a position invariance index defined as $PI = DS_{\text{surround}} / DS_{\text{center}}$. Four PIs were thus obtained for each detector. Graziano et al. (1994) indicated that all of the MSTd neurons included in their sample responded significantly (t test, $p < 0.05$) and were directionally selective in that the neurons showed a response to the preferred stimulus that was significantly ($p < 0.05$) greater than the response to the antipreferred stimulus. To mimic this, we established selection criteria such that the preferred radial direction needed to be $>12\%$ of the maximum response and the DS index at the central location needed to be >0.25 . Because the model detectors have no defined noise or baseline output, it is difficult to compare quantitatively our selection criteria and theirs.

Graziano et al. (1994, their Fig. 11) found that, for their sample of MSTd neurons, the resulting PIs were tightly clustered around 1.0, which is the value that indicates perfect position invariance (Fig. 9A). No negative values were found, indicating that, for their sample, directional preference never reversed. The PIs for a random sample of frontoparallel detectors that met the above response criteria are shown in Figure 9B. The large majority have PIs near 1.0, consistent with the MSTd data. The ground configuration yielded similar results. The PIs for the same sample of detectors from the ground configuration are shown in Figure 9C. The distribution is again tightly clustered around 1.0, although a few negative values are evident. Thus, like MSTd neurons, detectors from both model configurations exhibit limited position invariance (defined as little change in the directional *selectivity*) when tested with small stimulus patches separated by 5° .

The effect of moving larger test patches over larger distances has also been examined (Duffy and Wurtz, 1991b; Lagae et al., 1994). Although such procedures typically reveal large variations in the response amplitudes with stimulus location [see Duffy and Wurtz (1991b), their Figs. 7, 8], Duffy and Wurtz focused on the variation in the binary directional *preference* along a cardinal axis of motion. The important advantage of this approach is that it is immune to the problem of artifactual amplitude variations caused by a stimulus patch being only partially in the receptive field. The disadvantage is that it de-emphasizes legitimate variations in the

response amplitude with position that may encode important information. Examining the response of MSTd neurons to $33^\circ \times 33^\circ$ patches of optic flow placed in one of nine positions in a 3×3 grid tiling the same $100^\circ \times 100^\circ$ area as their initial probe stimulus, Duffy and Wurtz found that many MSTd neurons retain their directional preference (e.g., continue to prefer expansion over contraction) over their entire receptive field [see Duffy and Wurtz (1991b), their Fig. 7]. Such neurons therefore display, over larger distances, a different form of limited position invariance than that examined by Graziano et al. (1994). When tested under the Duffy and Wurtz (1991b) stimulus conditions, 38% of the total population of model detectors from the frontoparallel configuration and 27% of those from the ground detectors maintain the same preference for one direction of radial motion at all nine locations. Unlike the subset of MSTd neurons and template-model detectors, invariance of directional preference is never found over the whole receptive field for Lappe–Rauschecker (1993) sigmoidal units that always show a systematic reversal of directional preference across a line dividing their receptive field [see Lappe et al. (1996), their Fig. 5].

To quantify this limited position invariance further, Duffy and Wurtz (1991b, their Table 2) performed the following analysis. They compared the nine expansion–contraction pairs of small-patch responses to radial motion with the response pair to large-field radial motion; 77% of the small-patch radial responses of their population of MSTd neurons showed the same directional preference as the corresponding large-field patch response. Furthermore, MSTd responses to roll motion appeared less invariant; only 59% of the small-patch roll responses showed the same form of invariance. The behavior of the entire population of frontoparallel detectors is quite similar; 86% of radial and 53% of roll small-patch responses kept the same directional preference as that of the corresponding large-field response. The behavior of the entire population of ground detectors is also quite similar; 83% of radial and 50% of roll small-patch responses kept the same directional preference as that of the corresponding large-field response. In summary, although patches of motion exploring the whole receptive field (Duffy and Wurtz, 1991b; Lagae et al., 1994) as well as large input fields with different centers of motion (Duffy and Wurtz, 1995; Lappe et al., 1996) reveal large variations in response amplitude, many MSTd neurons and model detectors maintain their radial or roll response directional preferences over large portions of their receptive field. Therefore, although strict position invariance (defined as a response that does not change with stimulus position) is not a property of either the model detectors or MSTd neurons, limited position invariance (defined as a response that maintains its directional preference) can manifest itself for a sizable subset of MSTd neurons and model detectors when test patches are moved over the entire receptive field.

Spiral tuning

Graziano et al. (1994) also found many MSTd cells that seem to respond best to spiral motion (radial plus roll), although there exist conflicting reports as to the predominance of such cells. Lagae et al. (1994) and Duffy and Wurtz (1995) claim that such cells are uncommon, although the methodologies differed considerably across the studies. In this section, we test the model detectors for spiral tuning, and in the next section, we test for spiral invariance. Any emergent spiral tuning in the detectors would indicate that this property is compatible with individual MSTd neurons encoding heading. We simulated the spiral-tuning

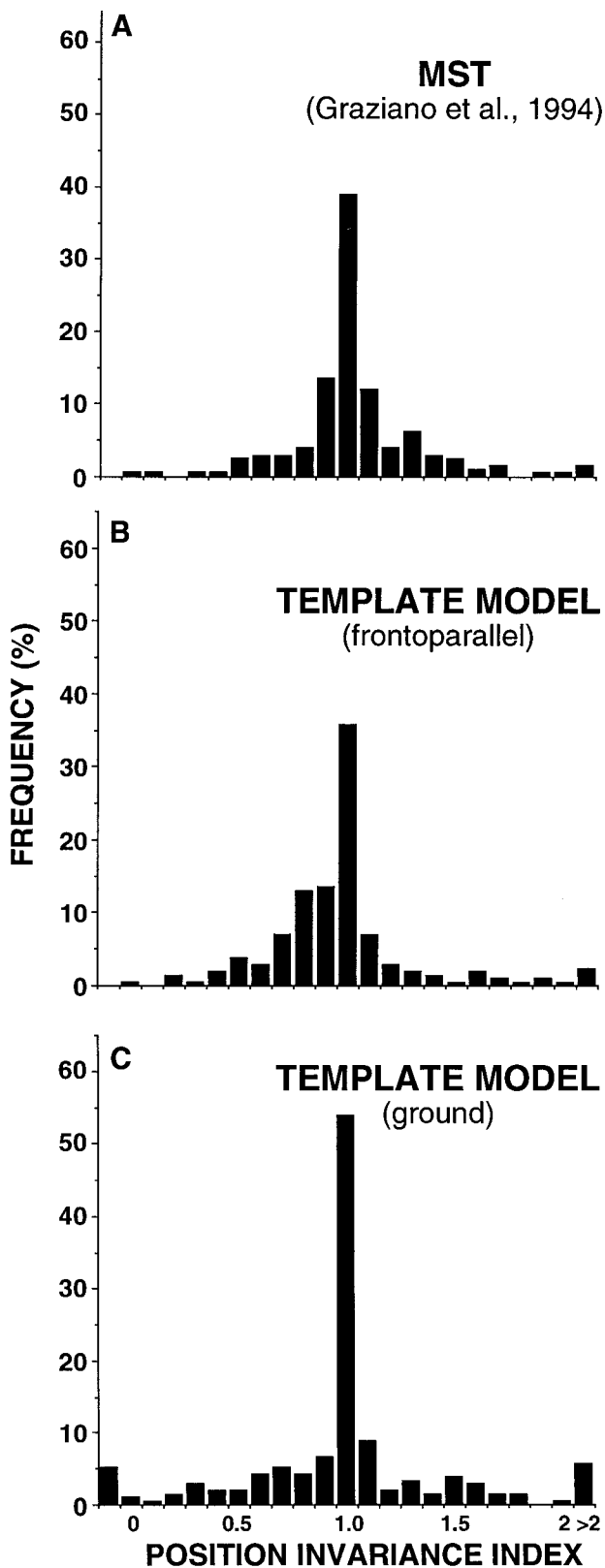


Figure 9. Position-invariance test. *A*, Distribution of the position invariance index for a sample of 52 MST neurons (208 responses) replotted from Graziano et al. (1994, their Fig. 11). The vertical axis shows the percentage of responses for each position invariance index value. *B*, The same distribution for a sample of 54 detectors (216 responses) from the frontoparallel configuration. *C*, The distribution for the same 54 detectors from the ground configuration.

experiments of Graziano et al. (1994) using their set of eight stimuli: expansion, contraction, clockwise rotation (CW), counterclockwise rotation (CCW), and four intermediate spiral patterns (expanding clockwise spiral, expanding counterclockwise spiral, contracting counterclockwise spiral, and contracting clockwise spiral). They represented these stimuli in “spiral space” (Fig. 10*A*). In such plots, 90° corresponds to expansion, 270° to contraction, 0° to clockwise roll, and 180° to counterclockwise roll. The oblique directions (45, 135, 225, and 315°) correspond to the four intermediate spiral stimuli listed above. The stimuli were 20° diameter patches containing 126 dots moving at an average speed of 4.4°/sec, over the center of the receptive field. As in the study of Graziano et al. (1994), the resulting tuning curves (plotted in Cartesian coordinates) were fit with a Gaussian to find the peak (the mean of the Gaussian) that corresponds to the preferred direction in spiral space and to provide a measure of bandwidth (σ , the SD of the Gaussian) and goodness-of-fit (r , the correlation coefficient).

Model detectors show tuning in spiral space similar to that of MST neurons. Figure 10*A* is the spiral space plot for the frontoparallel detector tuned to (−25.2°, −6.5°, 1°/sec) that prefers clockwise outward spiral patterns. Figure 10*B* shows the same tuning curve plotted in Cartesian coordinates along with its fitted Gaussian. The preferred spiral direction for this detector is 66°. The SD of the fitted Gaussian is 84° ($r = 0.98$). This response is, however, not typical of the population. Although the spiral space tuning of frontoparallel detectors is well fit by a Gaussian (72% with $r > 0.9$; mean $\sigma = 65^\circ$), the preferred spiral direction is nearly always $\sim 90^\circ$ (pure expansion).

The ground configuration produces a better match to the spiral tuning of the sample of MSTd neurons from Graziano et al. (1994). Figure 11 shows spiral-tuning curves of two of their neurons and of two ground detectors. An example of an expansion-tuned neuron is shown in Figure 11*A*. It had a preferred direction of 89° and a bandwidth of 33° ($r = 0.99$). Figure 11*B* shows the tuning curve for the detector tuned to (0°, 6°, 1°/sec). The preferred direction was 90°, and the bandwidth was 32° ($r = 0.99$). An example of a spiral-tuned neuron is depicted in Figure 11*C*. It has a preferred direction of 133° and a bandwidth of 57° ($r = 0.99$). Figure 11*D* is the tuning curve for the detector tuned to (20.3°, −5.3°, 1°/sec). Its preferred spiral direction is 134°, and the bandwidth is 42° ($r = 0.99$). The examples in Figure 11, *B* and *D*, are typical of detectors from the ground configuration.

Graziano et al. (1994) found that 20 MSTd neurons ($\sim 35\%$) out of their sample of 57 Gaussian-tuned units ($r > 0.9$) were spiral tuned, i.e., had preferred spiral directions within $\pm 22.5^\circ$ of the oblique axes (Fig. 12*A*). To compare this with the distribution of preferred spiral directions of the frontoparallel detectors, we randomly sampled 100 detectors and plotted their preferred spiral space direction tuning in polar form. Seventy-nine detectors met their goodness-of-fit criterion ($r > 0.9$). Although there are examples of spiral-tuned frontoparallel detectors (e.g., Fig. 10), spiral tuning is much rarer than in the sample of MSTd neurons of Graziano et al. (1994). The frontoparallel detectors cluster around expansion (Fig. 12*B*). Even when the entire population is tested, only $\sim 1\%$ of the Gaussian-tuned detectors prove to be spiral tuned. One obvious difference between the frontoparallel configuration data (Fig. 12*B*) and the MSTd data (Fig. 12*A*) is the lack of contraction detectors resulting from our previous arbitrary choice to ignore backward headings. Another possible contribu-

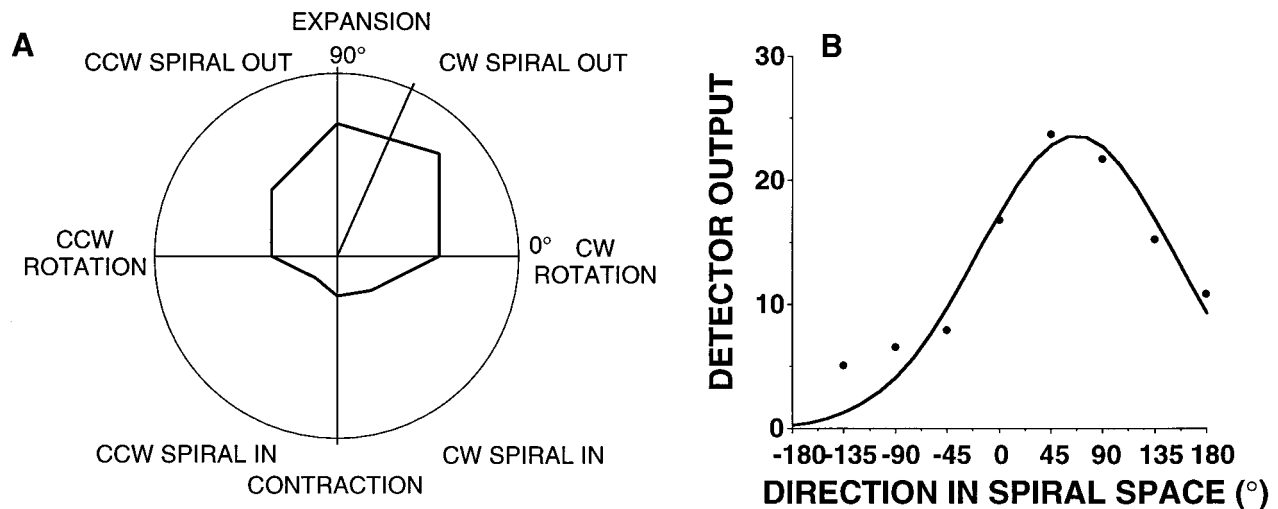


Figure 10. Spiral tuning. *A*, Polar plot of detector responses using the spiral space representation of Graziano et al. (1994, their Fig. 7). The polar angle represents the stimulus type as described in the text. This detector responds best to expanding clockwise spiral stimuli. The *solid line* (also see Figs. 11, 13) indicates the preferred spiral direction. *B*, The same tuning curve plotted in Cartesian coordinates. The *continuous curve* is the best fitting Gaussian from which the preferred spiral direction, bandwidth, and goodness-of-fit were derived. The example detector was chosen to illustrate the existence of spiral-tuned detectors in the frontoparallel configuration.

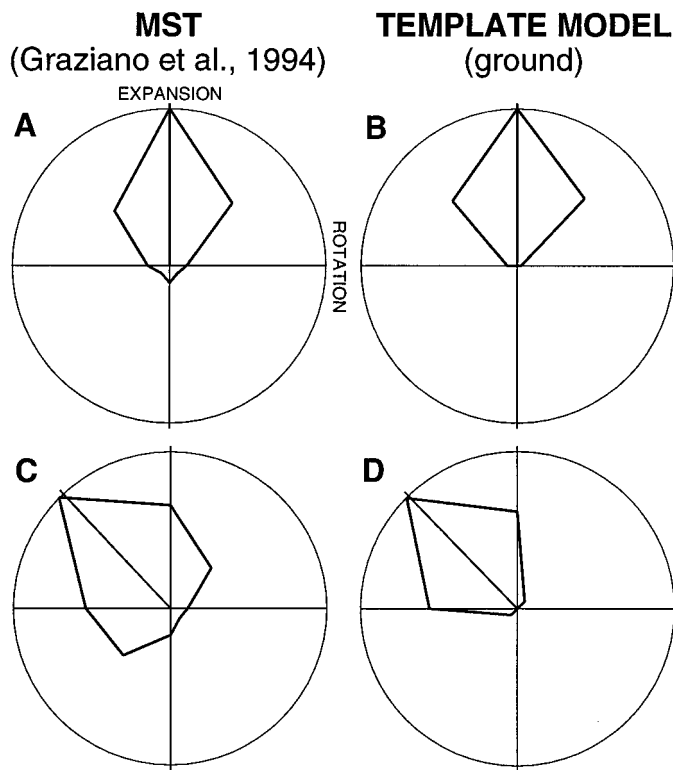


Figure 11. Comparison of the spiral tuning of MSTd neurons and ground detectors. Polar plots follow the same convention as that in Figure 10*A*. *A*, *C*, The responses of the two MSTd example neurons from Graziano et al. (1994, their Fig. 8*A*, *C*). *B*, *D*, The responses of two model detectors chosen by surveying the population for responses that matched those of the example neurons.

tor to this difference is that the simulation stimuli were perfectly centered on the receptive field of the detectors, i.e., were presented exactly at (0°, 0°). During single-unit experiments, this is not possible. When the stimuli are randomly centered in a 20° × 20° box centered in the receptive field, the percentage of spiral-

tuned detectors increases to ~11%. Nonetheless, the proportion of spiral-tuned detectors in the frontoparallel configuration appears lower than that found by Graziano et al. (1994).

In the ground configuration, spiral-tuned detectors are common. The distribution of preferred spiral directions (Fig. 12*C*) is similar to that of the sample of MSTd neurons of Graziano et al. (1994) (Fig. 12*A*). Figure 12*C* is based on a random sample of 200 ground detectors of which 148 met their selection criterion ($r > 0.9$). The inclusion of a single map of backward-tuned detectors produced a ratio of contraction to expansion tuning similar to that found in MSTd. Although the preferred spiral directions still cluster near expansion, consistent with the data from Graziano et al. (1994) as well as from many other studies (Tanaka and Saito, 1989; Duffy and Wurtz, 1991a), a reasonable proportion of the ground detectors (27%) are spiral tuned. The mean bandwidth of the sample is 54°. Furthermore, this sample is representative of the entire ground detector population (76% with $r > 0.9$; 27% spiral tuned; mean $\sigma = 52^\circ$). The properties for their sample of MSTd neurons are similar (86% with $r > 0.9$; 35% spiral tuned; mean $\sigma = 61^\circ$). We conclude that a simple parametric modification of the depth-sampling parameters makes spiral tuning nearly as common among template-model detectors as among MSTd neurons. Perhaps the spiral tuning of ground detectors should not be surprising; an examination of the optic-flow pattern in Figure 1*B* illustrates that spiral flow does indeed occur during self-motion over natural ground-plane-like layouts. The effect of depth sampling on spiral tuning suggests that the examination of more ecologically appropriate depth-sampling strategies is a worthwhile area for future exploration. Another important result of the spiral-tuning simulations is the discovery that detectors without a rotation component often display spiral tuning. Out of 288 detectors in the 0°/sec rotation map from the ground configuration, ~10% are spiral tuned with $\sigma < 50^\circ$ and $r > 0.95$. This shows that a pure expansion detector, when tested with stimuli not centered on their preferred COM, can exhibit sharp spiral tuning. In other words, spiral tuning, as defined by Graziano et al. (1994), does not require a spiral receptive-field structure.

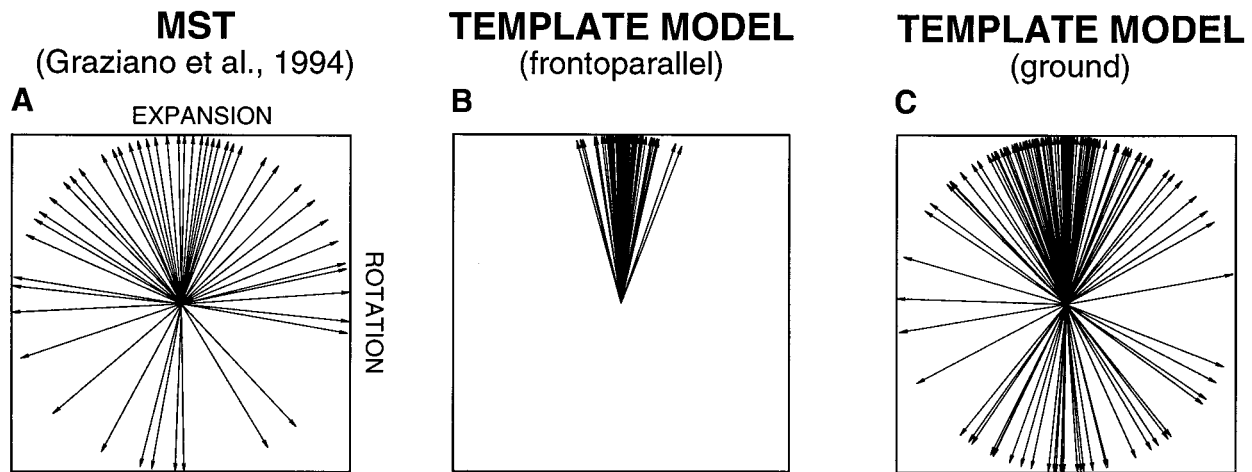


Figure 12. Spiral-tuning test. *A*, Replotted data from Graziano et al. (1994, their Fig. 9) showing the preferred spiral direction for their sample of 57 MSTd neurons. *B*, The distribution of preferred spiral directions of a random sample of 79 frontoparallel detectors. The lines in the model plots have been jittered by a small amount in the range $\pm 2^\circ$ to reduce the amount of overlap. *C*, The distribution of preferred spiral directions of a random sample of 148 ground detectors.

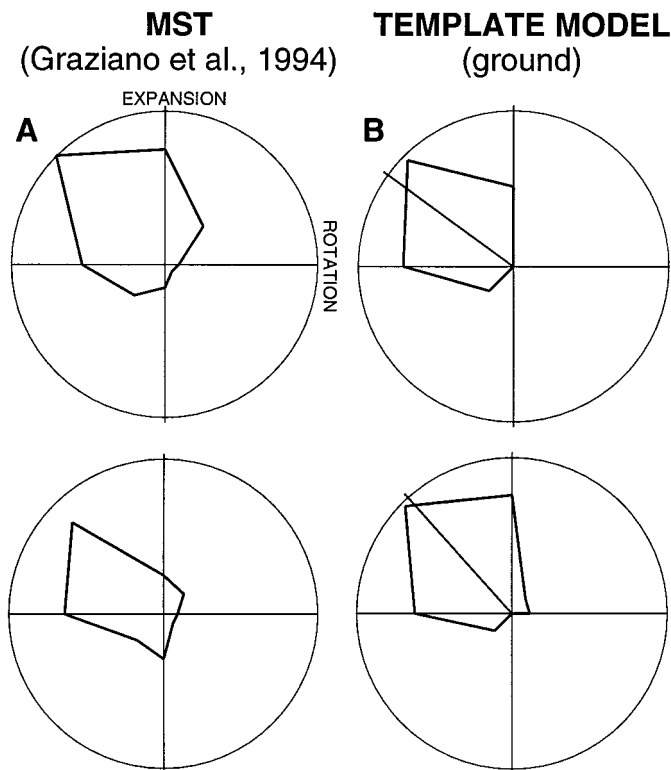


Figure 13. Spiral-invariance test. *A*, Replotted data from Graziano et al. (1994, their Fig. 14) showing the spiral-tuning curves of a single MSTd neuron obtained for stimulus presentation at two vertical displaced positions shifted by 8.25° . *B*, Results for the same test performed on an example ground detector chosen by surveying the population for responses that matched that of the example neuron.

Spiral invariance

Graziano et al. (1994) also tested position invariance using a spiral-tuning criterion, or “spiral invariance.” Using the ground configuration, we repeated their spiral-invariance test by measuring spiral tuning with stimulus sets presented at two locations in

the receptive field and generating a different tuning curve for each location. The first location was in the center of the field, and the second was 8.25° below the first. The stimulus size was reduced to 16.5° diameter to match the methods of Graziano et al. (1994).

Figure 13*A* (*top, bottom*) replots the responses of one of the spiral-tuned neurons of Graziano et al. (1994). Even though some change in bandwidth and shape is apparent, they found that for the 22 MSTd neurons tested, the preferred spiral direction shifted on average by only 10.7° . Figure 13*B* (*top, bottom*) shows the tuning curves for the spiral-tuned model detector (-25° , 6.5° , $4^\circ/\text{sec}$). The preferred tuning directions for the two vertically displaced positions were 144° ($\sigma = 46^\circ$; $r = 0.99$) and 132° ($\sigma = 50^\circ$; $r = 0.99$), indicating a shift of 12° . The small change in shape and preferred direction is comparable with that found for MSTd neurons. For the entire population of detectors (for pairs of curves with $r > 0.9$), the median shift in preferred spiral tuning across the two vertically displaced stimulus positions was 14.0° (although the mean was 26.3° because the distribution is skewed).

Hence, over a relatively short distance, model detectors exhibit spiral invariance as defined by Graziano et al. (1994). The reason is that the change in position used in their study ($\sim 8^\circ$) is small compared with the bandwidth of heading tuning of the detectors and of MSTd neurons ($\sim 30\text{--}40^\circ$, see Fig. 7). Indeed, if the detector population is tested using 50° position shifts, the median change in the preferred spiral direction increases to $\sim 65^\circ$. Recently, however, Geesaman and Andersen (1996) have published the results of testing a single MSTd neuron for spiral invariance using 50° shifts and found preferred spiral direction changes up to $\sim 52^\circ$ (found by fitting Gaussians to the *solid-square* curves in the *bottom two panels* of their Fig. 16), which is not inconsistent with our simulations. Strict spiral invariance is therefore not a property of either template-model detectors or MSTd neurons, but a limited spiral invariance can manifest itself when small test patches are moved over small distances.

DISCUSSION

Our results demonstrate that the characteristic visual receptive-field properties of MST neurons (multicomponency, wide-field spatial integration, sensitivity to nonpreferred flow, COM tuning,

limited position invariance, spiral tuning, and limited spiral invariance) can be explained by a template model of heading estimation. Because the model detectors were designed to estimate heading and not crafted to mimic MST responses, their physiological properties are emergent. Furthermore, both depth-sampling configurations produced nearly identical results, except for their spiral tuning. More pointedly, the sensitivity to nonpreferred flow, COM tuning, and limited position invariance of MST neurons can be quantitatively explained without changing any of the parameters set by Perrone and Stone (1994). Changes in depth sampling and the inclusion of backward-tuned detectors were however used to make spiral tuning as common as in MSTd.

We have also clarified the issue of position invariance. Graziano et al. (1994) called a response “position invariant” if small shifts in the stimulus location did not much change their directional selectivity along the preferred cardinal axis of motion and “spiral invariant” if shifts did not much change the preferred spiral direction. Duffy and Wurtz (1991b) examined another form of limited position invariance defined using a directional-preference criterion. Our simulations demonstrate that both limited forms of invariance are fully consistent with individual MSTd neurons directly encoding heading (see also Zhang et al., 1993). Indeed, the model predicts that responses will not be strictly invariant and that large shifts in stimulus location will often produce large changes in response amplitude, consistent with the MSTd data.

Refinements to the template model

Our model represents a “proof of principle” showing that a template-like computational strategy (surely more complex than ours) could underlie heading estimation from optic flow within MSTd. Nonetheless, the model will need refinement if it is to be used as a more complete descriptor of primate heading estimation or MSTd receptive fields. A number of refinements are motivated by the fact that cues, other than optic flow, could be helpful in self-motion estimation. The disparity signals both within MT (Maunsell and Van Essen, 1983b; Bradley et al., 1995) and MST (Roy and Wurtz, 1990; Roy et al., 1992) could be used to enhance model performance by providing depth cues independent of flow. Oculomotor or vestibular signals could be used to weight detector responses within heading maps (Perrone and Stone, 1994; Bradley et al., 1996), or eye movement signals could dynamically alter the MT inputs to MST (Perrone, 1992). There are signals related to eye movements within MST (Newsome et al., 1988; Thier and Erickson, 1992a; Siegel and Read, 1994; Bremmer et al., 1997), and eye movements can alter MST responses to optic flow (Duffy and Wurtz, 1994; Bradley et al., 1996). Such oculomotor signals could compensate for rotation in the optic flow as they have been shown to assist in path estimation (Royden et al., 1994). However, oculomotor compensation for rotation appears at best only partial within MST (Bradley et al., 1996) and is not necessary for accurate heading estimation (Rieger and Toet, 1985; Cutting, 1986; Stone and Perrone, 1993, 1997a). Vestibular responses in MST neurons are also beginning to be explored within the context of heading perception (Thier and Erickson, 1992a,b; Duffy, 1996; Pekel et al., 1996; Shenoy et al., 1996). Finally, higher order optic-flow properties (e.g., acceleration) could provide important self-motion information (Rieger, 1983; Perrone, 1996).

Alternate models

Orban et al. (1992) provided physiological evidence against full decomposition models of heading estimation. Moreover, models

that use differential motion for decomposition (e.g., Rieger and Lawton, 1985; Hildreth, 1992; Royden, 1997) predict systematic errors across depth discontinuities in the layout that are qualitatively different from the observed small psychophysical errors related to the trajectory and unrelated to the layout (Stone and Perrone, 1993). In addition, expansion stimuli devoid of depth variation produce little response in differential-motion units yet can generate vigorous MST responses (e.g., Duffy and Wurtz, 1991a,b).

Lappe and Rauschecker (1993) proposed a two-layered partial decomposition model that predicts that expansion-tuned MST units will show sigmoidal tuning along a specific 1D axis and no variation along the orthogonal axis. After 2D bell-shaped heading tuning was found in many MST neurons (Duffy and Wurtz, 1995), Lappe et al. (1996) added a third layer to explain this finding. Nevertheless, the majority of their model units are sigmoidally tuned, and the apparent sigmoidal tuning of some MST neurons when tested only over a limited range does not provide strong support for their model. Their sigmoidal MST neurons may simply be tuned to planar motion or may not have been tested at high enough eccentricity to reveal a peripheral peak in the response curve. Lastly, the fact that many MST neurons maintain their direction preference over their entire receptive field (Duffy and Wurtz, 1991b; Lagae et al., 1994) is hard to reconcile with the fact that all Lappe–Rauschecker (1993) sigmoidal units systematically reverse their direction preference across their receptive fields.

The Lappe–Rauschecker model also requires image speed and direction (V_x , V_y) from its first layer to perform the vector computations essential to their approach. Originally (Lappe and Rauschecker, 1993, 1995), the output of their first layer units was explicitly proportional to speed [Lappe and Rauschecker (1993), their Eq. 2.3, p. 379], which is incompatible with the properties of MT neurons. Recently, they have proposed a more realistic distributed population code that uses a small basis set of MT-like units to encode velocity (Lappe et al., 1996). Yet, it remains unclear how this approach resolves which MT neurons to use and which to ignore when recovering velocity from the many active neurons with a near continuum of direction and speed preferences at each location. The template model however does not require velocity as input, makes explicit decisions as to which MT units at each location provide input to each detector (Eq. 3), and uses the full range of preferred directions and speeds represented within MT. Lastly, both the Lappe–Rauschecker and the Perrone and Stone (1994) template models assume gaze stabilization. Should this assumption prove too restrictive (Crowell, 1997), the template model can revert to its unrestricted version (Perrone, 1992) while remaining consistent with MST data. However, it may be difficult for the Lappe–Rauschecker model to revert to its unrestricted version (Heeger and Jepson, 1992), given the findings of Orban et al. (1992).

More recently, Zemel and Sejnowski (1998) proposed a “multiple-cause” model of MST. Its hidden units are proposed to encode optic flow within MST using a sparse, distributed representation that could be used to facilitate image segmentation and object- or self-motion estimation by read-out units in an area beyond MST. Many of their hidden units showed spiral tuning and spiral invariance, although the stimulus conditions were different from those of Graziano et al. (1994). Unfortunately, they did not quantitatively assess the sensitivity to nonpreferred flow, the COM tuning, and the position invariance of their hidden units, and no evaluation was performed on the physiological

plausibility of the read-out units that actually perform the segmentation and heading estimation. Lastly, their model predicts that the response of an MST neuron to its preferred stimulus or a piece of it (optic flow caused by the relative motion between an object and the observer) will be largely immune to other flow present in its receptive field (caused by motion relative to a different moving object). The template model predicts otherwise. Future experiments will be needed to resolve this issue.

The concept of optic-flow templates has been around for some time in the fields of insect vision (e.g., Horridge, 1991; Krapp and Hengstenberg, 1996) and primate self-motion estimation (Saito et al., 1986; Tanaka et al., 1986, 1989; Perrone, 1987, 1990; Glünder, 1990; Hatsopoulos and Warren, 1991). However, for primates, it has been less well accepted because of weaknesses in the early designs. For example, template models could not accurately process rotation, whereas decomposition models could. Furthermore, the specifics of how the templates would be constructed was not formalized, whereas decomposition models were often presented with formal mathematical proofs. In addition, the number of templates required to solve the general self-motion problem was assumed to be almost infinite (this problem is worse for multiple-cause models). Our template model overcomes many of these shortcomings and demonstrates that robust heading estimation is possible with a restricted number (~1000) of templates. Perhaps a less restricted model could perform even better in heading estimation, as well as exhibit a wider range of MSTd response properties (e.g., roll tuning). We conclude that the template model remains a viable descriptor of MSTd visual response properties and defines a simple and specific set of MT to MST connections sufficient to achieve these properties.

REFERENCES

- Albright TD (1984) Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol* 52:1106–1130.
- Bradley DC, Qian N, Andersen RA (1995) Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature* 373:609–611.
- Bradley DC, Maxwell M, Andersen RA, Banks MS, Shenoy KV (1996) Mechanisms of heading perception in primate visual cortex. *Science* 273:1544–1547.
- Bremmer F, Ilg UJ, Thiele A, Distler C, Hoffmann KP (1997) Eye position effects in monkey cortex. I. Visual and pursuit related activity in extrastriate areas MT and MST. *J Neurophysiol* 77:944–961.
- Crowell JA (1997) Testing the Perrone and Stone model of heading estimation. *Vision Res* 37:1653–1671.
- Crowell JA, Banks MS (1993) Perceiving heading with different retinal regions and types of optic flow. *Percept Psychophys* 53:325–337.
- Crowell JA, Banks MS (1996) Ideal observer for heading judgments. *Vision Res* 36:471–490.
- Cutting JE (1986) Perception with an eye for motion. Cambridge, MA: Bradford.
- Duffy CJ (1996) Real movement responses of optic flow neurons in MST. *Soc Neurosci Abstr* 22:1692.
- 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.
- 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.
- Duffy CJ, Wurtz RH (1994) Optic flow responses of MST neurons during pursuit eye movements. *Soc Neurosci Abstr* 20:1279.
- Duffy CJ, Wurtz RH (1995) Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *J Neurosci* 15:5192–5208.
- Geesaman BJ, Andersen RA (1996) The analysis of complex motion patterns by form/cue invariant MSTd neurons. *J Neurosci* 16:4716–4732.
- Gibson JJ (1950) The perception of the visual world. Boston: Houghton Mifflin.
- Glünder H (1990) Correlative velocity estimation: visual motion analysis, independent of object form, in arrays of velocity-tuned bilocal detectors. *J Opt Soc Am A* 7:255–263.
- Graziano MSA, Andersen RA, Snowden RJ (1994) Tuning of MST neurons to spiral motions. *J Neurosci* 14:54–67.
- Hatsopoulos N, Warren WH (1991) Visual navigation with a neural network. *Neural Networks* 4:303–317.
- Heeger DJ, Jepson AD (1992) Subspace methods for recovering rigid motion I: algorithm and implementation. *Int J Comp Vision* 7:95–177.
- Hildreth EC (1992) Recovering heading for visually-guided navigation. *Vision Res* 32:1177–1192.
- Horridge GA (1991) Ratios of template responses as the basis of semi-vision. *Philos Trans R Soc Lond [Biol]* 331:189–197.
- 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.
- Koenderink JJ, van Doorn AJ (1987) Facts on optic flow. *Biol Cybern* 56:247–254.
- Komatsu H, Wurtz RH (1988) Relation of cortical areas MT and MST to pursuit eye movements. I. Location and visual properties of neurons. *J Neurophysiol* 60:580–603.
- Krapp HG, Hengstenberg R (1996) Estimation of self-motion by optic flow processing in single visual interneurons. *Nature* 384:463–466.
- Lagae L, Maes H, Raiguel S, Xiao D-K, Orban GA (1994) Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. *J Neurophysiol* 71:1597–1626.
- Lappe M, Rauschecker JP (1993) A neural network for the processing of optic flow from ego-motion in man and higher mammals. *Neural Comput* 5:374–391.
- Lappe M, Rauschecker JP (1995) An illusory transformation in a model of optic flow processing. *Vision Res* 35:1619–1631.
- Lappe M, Bremmer F, Pekel M, Thiele A, Hoffmann K-P (1996) Optic flow processing in monkey STS: a theoretical and experimental approach. *J Neurosci* 16:6265–6285.
- Longuet-Higgins HC, Prazdny K (1980) The interpretation of moving retinal images. *Proc R Soc Lond [Biol]* 208:385–387.
- Maunsell JHR, Newsome WT (1987) Visual processing in monkey extrastriate cortex. *Annu Rev Neurosci* 10:363–402.
- Maunsell JHR, Van Essen DC (1983a) Functional properties of neurons in the middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, orientation. *J Neurophysiol* 49:1127–1147.
- Maunsell JHR, Van Essen DC (1983b) Functional properties of neurons in the middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol* 49:1148–1167.
- 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.
- 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 USA* 89:2595–2599.
- Pekel M, Lappe M, Schmidt M, Hoffman K-P (1996) Vestibular responses during linear motion in area MST of the macaque monkey. *Soc Neurosci Abstr* 22:1618.
- Perrone JA (1986) Anisotropic responses to motion toward and away from the eye. *Percept Psychophys* 39:1–8.
- Perrone JA (1987) Extracting 3-D egomotion information from a 2-D flow field: a biological solution? *Opt Soc Am Tech Digest Series* 22:47.
- Perrone JA (1990) Simple technique for optical flow estimation. *J Opt Soc Am A* 7:264–278.
- Perrone JA (1992) Model for the computation of self-motion in biological systems. *J Opt Soc Am A* 9:177–194.
- Perrone JA (1994) Simulating the speed and direction tuning of MT neurons using spatiotemporal tuned V1-neuron inputs. *Invest Ophthalmol Vis Sci [Suppl]* 35:2158.
- Perrone JA (1996) Generating acceleration sensitive motion sensors from sets of spatio-temporal filters. *Invest Ophthalmol Vis Sci [Suppl]* 37:S750.
- Perrone JA, Stone LS (1994) A model of self-motion estimation within primate extrastriate visual cortex. *Vision Res* 34:2917–2938.
- Rieger JH (1983) Information in optical flows induced by curved paths of observation. *J Opt Soc Am* 73:339–344.

- Rieger JH, Lawton DT (1985) Processing differential image motion. *J Opt Soc Am A* 2:354–360.
- Rieger JH, Toet L (1985) Human visual navigation in the presence of 3D rotations. *Biol Cybern* 52:377–381.
- Roy JP, Wurtz RH (1990) The role of disparity-sensitive cortical neurons in signalling the direction of self-motion. *Nature* 348:160–162.
- Roy JP, Komatsu H, Wurtz RH (1992) Disparity sensitivity of neurons in monkey extrastriate area MST. *J Neurosci* 12:2478–2492.
- Royden CS (1997) Mathematical analysis of motion-opponent mechanisms used in the determination of heading and depth. *J Opt Soc Am A* 14:2128–2143.
- Royden CS, Crowell JA, Banks MS (1994) Estimating heading during eye movements. *Vision Res* 34:3197–3214.
- 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.
- Shenoy KV, Bradley DC, Andersen RA (1996) Heading computation during head movements in macaque cortical area MSTd. *Soc Neurosci Abstr* 22:1692.
- Siegel RM, Read HL (1994) Egocentric motion from optic flow and eye position in area 7a of the behaving macaque. *Soc Neurosci Abstr* 20:1278.
- Stone LS, Perrone JA (1993) Human heading perception cannot be explained using a local differential motion algorithm. *Invest Ophthalmol Vis Sci [Suppl]* 34:1229.
- Stone LS, Perrone JA (1994) A role for MST neurons in heading estimation. *Soc Neurosci Abstr* 20:772.
- Stone LS, Perrone JA (1997a) Human heading estimation during visually simulated curvilinear motion. *Vision Res* 37:573–590.
- Stone LS, Perrone JA (1997b) Quantitative simulations of MST visual receptive field properties using a template model of heading estimation. *Soc Neurosci Abstr* 23:1126.
- Tanaka K, Saito H (1989) Analysis of the 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.
- Tanaka K, Hikosaka K, Saito H, Yukie M, Fukada Y, Iwai E (1986) Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *J Neurosci* 6:134–144.
- Tanaka K, Fukada Y, Saito H (1989) 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.
- Thier P, Erickson RG (1992a) Responses of visual-tracking neurons from cortical area MSTl to visual, eye, and head motion. *Eur J Neurosci* 4:539–553.
- Thier P, Erickson RG (1992b) Vestibular input to visual-tracking neurons in area MST of awake rhesus monkeys. *Ann NY Acad Sci* 656:960–963.
- Ungerleider LG, Desimone R (1986) Cortical connections of area MT in the macaque. *J Comp Neurol* 248:190–222.
- Warren WH, Kurtz KJ (1992) The role of central and peripheral vision in perceiving the direction of self-motion. *Percept Psychophys* 51:443–454.
- Zemel RS, Sejnowski TJ (1998) A model of encoding multiple object motions and self-motion in area MST of primate visual cortex. *J Neurosci* 18:531–547.
- Zhang K, 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.