

# Visual–vestibular estimation of the body's curvilinear motion through the world: A computational model

John A. Perrone

School of Psychology, University of Waikato,  
Hamilton, New Zealand



**Motion along curved paths (curvilinear self-motion) introduces a rotation component to the radial expanding patterns of visual motion generated in the eyes of moving animals with forward-facing eyes. The resultant image motion (vector flow field) is no longer purely radial, and it is difficult to infer the heading direction from such combined translation-plus-rotation flow fields. The eye need not rotate relative to the head or body during curvilinear self-motion, and so there is an absence of efference signals directing and indicating the rotation. Yet the eye's rotation relative to the world needs to be measured accurately and its effect removed from the combined translation–rotation image motion in order for successful navigation to occur. I demonstrate that to be able to account for human heading-estimation performance, the precision of the eye-in-world rotation velocity signal needs to be at least 0.2°/s. I show that an accurate estimate of the eye's curvilinear motion path through the world can be achieved by combining relatively imprecise vestibular estimates of the rotation rate and direction with visual image-motion velocities distributed across the retina. Combined visual–vestibular signals produce greater accuracy than each on its own. The model can account for a wide range of existing human heading- and curvilinear-estimation psychophysical data.**

## Introduction

Locomotion is essential for many species of animal. If the animal has some form of optics that creates an image from the light surrounding it, then self-motion will generate patterns of image motion on the photoreceptor array processing the image.

It has long been known that this two-dimensional image motion contains information about our motion through the world (self-motion), and about the relative depth of points in the world (Gibson, 1950; Ono & Wade, 2005). Humans have an amazing ability to navigate through cluttered environments using a single eye. We can extract depth in fractions of a second from just the 2-D retinal image motion, whereas the artificial

vision systems used in robots and autonomous cars currently rely on multiple active sensors (e.g., lidar) to learn about the environment in front of them. Understanding how humans achieve this skill of depth from self-motion would help in the development of sensors for computer-vision-based obstacle-avoidance systems.

For humans and many other animals with forward-facing eyes, the pattern of motion created during forward translational movement of the body looks radial in nature, with the center of the expansion pattern (*focus of expansion*, FOE) corresponding to the heading direction (Gibson, 1950). These global patterns of image motion (called the *flow field*) can be represented using velocity vectors, as depicted in Figure 1a.

This is for the case in which the eye is directed 10° to the right of the instantaneous direction of heading. If we have access to the vectors in such a flow field, then it is straightforward to extract the relative depth of the points using basic trigonometry (Koenderink & van Doorn, 1975; Longuet-Higgins & Prazdny, 1980). The two requirements for extracting this depth information are a rotation-free flow field and the instantaneous heading direction of the eye or camera. These are closely related, because if the flow field contains a rotation component it is difficult to determine the heading.

## Template models for determining heading direction

There is a long history of attempts to extract information about the body's own motion from a two-dimensional flow field (Cutting, 1986; Koenderink & van Doorn, 1975; Lee, 1974; Longuet-Higgins & Prazdny, 1980; Nakayama & Loomis, 1974; R. Warren, 1976; for reviews, see Britten, 2008; Lappe, Bremmer, & van den Berg, 1999; W. H. Warren, 2003).

The primate visual brain area MSTd (dorsal medial superior temporal) has long been associated with self-

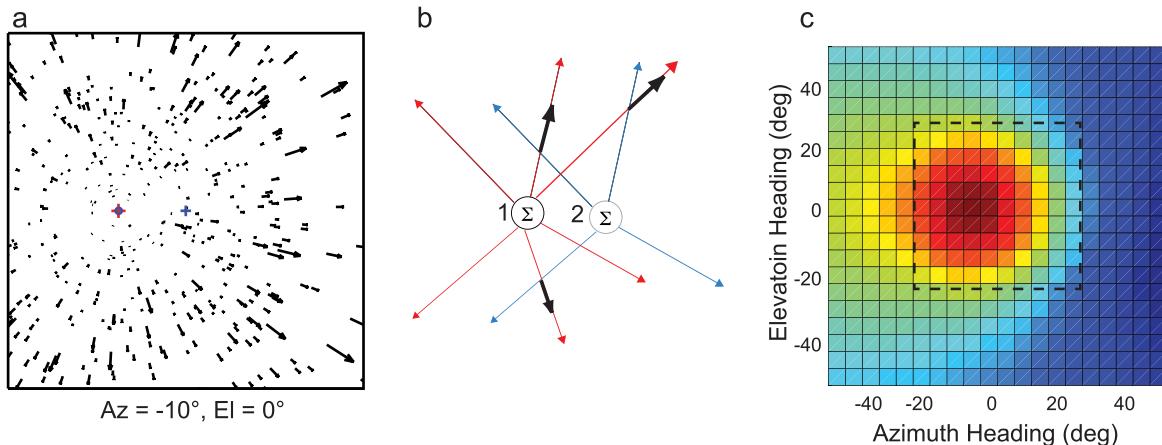
Citation: Perrone, J. A. (2018). Visual–vestibular estimation of the body's curvilinear motion through the world: A computational model. *Journal of Vision*, 18(4):1, 1–32, <https://doi.org/10.1167/18.4.1>.

<https://doi.org/10.1167/18.4.1>

Received August 10, 2017; published April 2, 2018

ISSN 1534-7362 Copyright 2018 The Authors





**Figure 1.** Image motion generated on the back of the eye during forward translation, and detectors for locating the point of expansion (heading direction). (a) Vector flow field for case in which the body is moving at an angle of  $-10^\circ$  relative to the line of sight (blue cross). The red cross signifies the focus of expansion and corresponds to the direction of heading. (b) Heading detectors based on MSTd neurons for locating the focus of expansion. (c) Heading map representing the activity of different heading detectors (small squares) in response to the flow field. Red indicates the most active units. The range of headings for the detectors can extend beyond the field of view of the input image (dashed square), but for illustration purposes only a limited range is depicted here.

motion estimation, and the early studies of the neuron properties in MSTd revealed that some MSTd neurons respond well to the expanding patterns of image motion generated during forward motion (Britten & van Wezel, 1998; Duffy & Wurtz, 1991; Saito et al., 1986; Tanaka et al., 1986). Soon after their discovery it was suggested that these neurons may be integrating the signals from directionally selective neurons in an earlier part of the primate visual-motion pathway (middle temporal, MT/V5), with their preferred directions aligned with the radial direction out from the FOE location (Tanaka, Fukada, & Saito, 1989). In parallel with these electrophysiological discoveries, a computer-based template model of heading estimation was proposed, made up of radially tuned sets of 2-D motion sensors (Perrone, 1987; Perrone, 1992). Through the use of such computer models, it was verified that the MSTd neurons in the primate visual system have the required properties for estimating heading from 2-D flow fields (Perrone & Stone, 1994; Perrone & Stone, 1998).

The principle behind the MSTd-based heading model is shown in Figure 1b. Each of the two circles (1 & 2) represents a separate model MSTd neuron tuned to a different direction of heading. The black arrows are sample image-velocity vectors ( $|V_i|, \theta_i$ ) located at positions  $(x_i, y_i)$ . If the heading tuning of MSTd unit #1 coincides with the position of the FOE  $(x_H, y_H)$ —which indicates actual heading direction ( $Az_H, El_H$ )—then the velocity vectors will be closely aligned with the radial directions  $\alpha_i$  joining  $(x_i, y_i)$  and  $(x_H, y_H)$ . For an MSTd unit tuned to another heading direction that does not coincide with the actual heading (MSTd unit #2), the radial directions do not line up as well with the

direction of the velocity vectors and there will often be a difference  $\beta_i$  between  $\theta_i$  and  $\alpha_i$ , and so the projection of the velocity vector onto the incorrect radial direction (the dot product) will be less than for the case in which  $\beta_i = 0^\circ$ . The sum of the projections ( $\Sigma V \cos \beta_i$ ) will be maximum when the actual heading lines up with the heading tuning of the detector (Perrone, 1992). This is why these models are referred to as template models: They are looking for the radial pattern in the model set of MSTd units that matches the one generated while the eye is translating in the direction ( $Az_H, El_H$ ).

Figure 1c shows the output from a set of model MSTd units tuned to a range of azimuth and elevation heading directions ( $-50^\circ$  to  $50^\circ$ , in  $5^\circ$  steps) when tested with the flow field shown in Figure 1a (heading: azimuth  $Az_H = -10^\circ$ , elevation  $El_H = 0^\circ$ ). The most active unit is the one tuned to  $(-10^\circ, 0^\circ)$ , and so the model is able to extract a heading direction from 2-D patterns of image motion generated during pure translation of the observer. These template models have been shown to be robust to the effect of input noise (perturbation of velocity-vector directions) and to account for a wide range of MSTd neuron properties (Perrone & Stone, 1994; Perrone & Stone, 1998).

Unfortunately, these basic heading-detector models fail when the eye rotates during the forward translation of the observer. The eye rotation can arise from movements of the eye in the head, movements of the head-eye system relative to the body, or movement of the body along curved paths (curvilinear self-motion).

Figure 2a shows the flow field for the same situation as depicted in Figure 1a, but at the same time that the eye is translating forward, it rotates to the right about the vertical axis at  $5^\circ/\text{s}$ . The image motion is now more

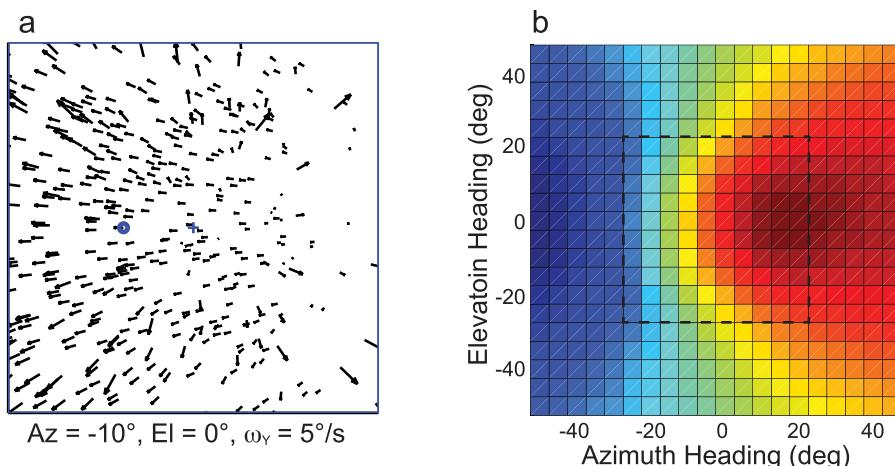


Figure 2. Vector flow field when rotation is present. (a) Heading is in direction  $(-10, 0)$  and rotation is  $5^\circ/\text{s}$  to the right. (b) Heading-detector map without rotation compensation.

complex and there is no obvious FOE in the location specified by the actual heading direction (circle to left of center). Each image-motion vector  $V_T$  in the flow field is perturbed by the image-motion vector  $V_R$  created with just the rotation, such that the resultant motion is given by the vector sum of the two ( $V_{T+R}$ ). An array of MSTd-like heading templates (Figure 2b) incorrectly signals that the heading direction is at  $(20^\circ, 0^\circ)$ . The addition of flow vectors caused by rotation is what makes the heading-estimation problem so hard (Regan & Beverley, 1982).

There are two main classes of theories that have been proposed for how heading can be derived from the combined T+R flow fields: vision-based models that only make use of the flow field and nonvisual (extraretinal) models which assume that additional sources of information (e.g., oculomotor or vestibular) are available.

### Vision-based methods for recovering the translation-only flow field

If the brain has only the visual flow field to extract heading and depth, it needs to overcome the rotation problem. Vision-based theories for doing this can be divided into three main (and often overlapping) approaches: (a) Extract just the translation component T of the motion free of any rotation R (differential-motion models); (b) measure the rotation visually and remove it from the T+R field (visual-compensation models); and (c) use the motion trajectory of points in the environment to infer heading without measuring rotation separately (flow-line models).

Differential-motion models (e.g., Rieger & Lawton, 1985; Royden, 1997) extract just the translation component of the optic flow by subtracting two nearby

vectors from each other. If the vectors arise from points at different depths in the world, then the subtraction process removes the rotation component (since it is independent of the depth and common to both vectors), leaving a component that tends to lie along the radial direction out from the FOE. A template-type model can then be used to determine the heading from the R-free expansion field.

In the second category of visual models (visual-compensation models), a number of different strategies have been suggested for measuring the rotation that is occurring in a combined T+R flow field. I have suggested a scheme for detecting rotation using networks of MT-like planar-motion detectors (Perrone, 1992). Beintema and van den Berg (1998) have proposed more elaborate rotation-tuned templates for detecting the rotation, and more recently it has been suggested that rotation information can be extracted from dynamic perspective cues (H. R. Kim, Angelaki, & DeAngelis, 2015; Sunkara, DeAngelis, & Angelaki, 2016).

The third category of vision-based models does not measure the rotation directly, and even argues against the idea that a separate estimation of heading occurs at all. Rather, these theories have suggested that path perception can be established independently of heading using the locomotor flow lines (Lee & Lishman, 1977) or the motion trajectories of points in the environment (N. G. Kim & Turvey, 1999; Wann & Swapp, 2000). It is not obvious in these schemes how the pure translation flow field would then be derived from the T+R flow in order to establish the depth. In addition, the experimental data mainly support the idea that the detection of the visual rotation is a key step in the determination of one's curved path (J. C. Cheng & Li, 2012).

Each of these vision-based approaches to the rotation-removal problem has different weaknesses,

but they all suffer from one main problem: They use only a limited part of the visual information available. The differential-motion models specifically include a vector-subtraction stage. For large regions of the visual field where nearby points in the world have small depth variation, this vector-subtraction step leaves minimal flow information. In addition, the subtraction leaves only a part of the  $V_{T+R}$  vector at each image location depending on the depth separation of the points. The remaining flow information used to determine the heading (and depth) represents only a fraction of the original flow field.

The current visual-rotation-detection models also use only a part of each  $V_{T+R}$  vector to determine the rotation. The Perrone (1992) scheme used a projection of the vectors onto candidate template rotation directions, and except for very distant points, this was only a small part of the full vector. The Beintema and van den Berg (1998) templates only measure the rotation for directions orthogonal to the radial-translation vectors. The dynamic-perspective-cue models (H. R. Kim et al., 2015; Sunkara et al., 2016) rely on small deviations from planar motion that become significant only at the edges of the flow field and when the field of view is relatively large. The visual motion in the central region of the flow field lacks the parallax information used for determining the rotation.

Given that the measurement of image velocity is difficult and prone to noise (Perrone, 2012), it makes sense to use as many motion estimates as possible over the largest area of the visual field that is practical when attempting to extract and measure the rotation. When couched in terms of a signal-to-noise problem, the measurement of rotation using all of the visual motion present in the flow field has an advantage over schemes that use only a portion of that information. I have overcome this major limitation of earlier approaches and found a previously undiscovered means of measuring the rotation visually using the full flow vector at each location across the whole image. This forms the basis of the model of curvilinear path detection I present in this article.

## Extraretinal methods for recovering the $V_T$ flow field and heading

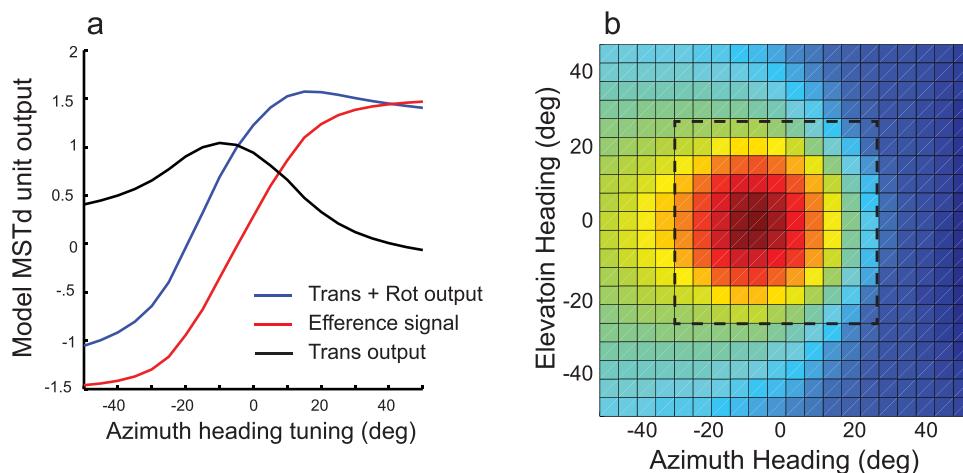
In parallel with the vision-based theories that were proposed for solving the rotation problem, a number of theories have been suggested that do not use visual flow for estimating the rotation but rather assume that the brain has access to signals indicating how much rotation of the eye has occurred (e.g., Banks, Ehrlich, Backus, & Crowell, 1996; Crowell, Banks, Shenoy, & Andersen, 1998; Freeman, Champion, & Warren, 2010; Gu, Watkins, Angelaki, & DeAngelis, 2006; Perrone &

Krauzlis, 2008; Telford, Howard, & Ohmi, 1995). For the Figure 2 case, if the 5°/s yaw rotation has occurred as a consequence of an eye rotation in the head or a head rotation relative to the body, then an efference (corollary discharge) signal is available to the visual system to remove the effect of the rotation (Sommer & Wurtz, 2008; Sperry, 1950; von Holst & Mittelstaedt, 1950). We have previously shown that with a known rotation rate and direction signal (e.g., 5°/s about the vertical axis), it is possible to carry out the equivalent of local vector subtraction ( $V_{T+R} - V_R$ ) to recover the pure translation vector  $V_T$  by adding a known efference signal to each of the MSTd heading templates (Perrone & Krauzlis, 2008).

A particular rotation of the eyes or head causes a known amount of activity in each of the heading templates. Because the rotation is independent of the depth of the point in the world, the image motion created by just the rotation ( $V_R$ ) depends only on the rotation rate and direction. Therefore, for a particular location in the image, the projection of the  $V_R$  vector onto the radial direction specified by a particular heading template can be calculated. The rotation vector could also be subtracted directly from the  $V_{T+R}$  vector at that location, but psychophysical evidence suggests that such local vector subtraction does not occur in the human visual system (Beintema & van den Berg, 2001). In order to allow for this constraint, we showed that local vector subtraction is equivalent to a particular amount of activity being subtracted from the individual heading templates (Perrone & Krauzlis, 2008). Figure 3 shows this process being applied to the template activity generated by the flow field shown in Figure 2a.

The activity in a set of heading templates tuned to a range of heading directions in response to the combined  $T+R$  flow (Figure 2a) is shown as the blue curve in Figure 3a. As indicated in Figure 2b, this produces an incorrect heading estimate (20° to the right). The amount of activity generated in each template in response to the 5°/s yaw rotation is known from the efference signal that generated the rotation, and it is represented by the red curve in the figure. This activity distribution is subtracted from the  $T+R$  distribution to produce the black curve and the correct heading (−10°, 0°). The full 2-D heading-template activity map is shown in Figure 3b. This form of efference-compensation model can successfully remove the effect of the rotation and recover the correct heading direction and  $V_T$  flow field for depth estimation (Perrone & Krauzlis, 2008).

This mechanism was assumed to work with either eye movements relative to the head or head movements relative to the body when there is a vestibular efference signal available (Angelaki & Cullen, 2008). There is some limited evidence showing a modification of the responses of MSTd neurons when eye or body-relative



**Figure 3.** Compensation for effect of rotation using an efference signal. (a) A particular rotation rate produces a known amount of activity in the heading detectors (red curve). This can be subtracted from the combined translation and rotation signals in the templates (blue curve) to leave the correct pure translation distribution (black curve). (b) Two-dimensional map of heading-detector activity distribution after the compensation has occurred. The correct heading direction is now indicated by the peak (dark-red region).

head movements occur (Bradley, Maxwell, Andersen, Banks, & Shenoy, 1996; Gu et al., 2006), and with human heading estimation when actual eye, head, and body rotations occur (Bertin & Berthoz, 2004; Crowell et al., 1998). Figure 3a indicates why it could be difficult to obtain electrophysiological evidence for this compensation mechanism by looking at only the activity from single neurons. If a single heading detector that was tuned to around  $-5^\circ$  azimuth was selected for recording, very little change in its activity would be predicted; the precompensation activity (blue curve) is very similar to the postcompensation activity (red curve) at  $-5^\circ$ , and it would be difficult to discern any difference in the activity of the cell. Simultaneous recording across a population of MSTd neurons would be required to find evidence for the rotation-compensation mechanism.

Similarly, it is difficult to assess heading performance ability using a passive observer sitting in front of a screen. This was the standard practice for the majority of early human heading experiments, yet it represents a cue-conflict situation from the point of view of efference-compensation models; the vestibular system is indicating that no rotation is occurring, yet the visual flow is indicating rotation. This problem was identified and rectified by some early heading researchers (Telford et al., 1995), but the bulk of the research involves visual–vestibular cue conflict. This is an important issue that needs to be considered when assessing the different models, and I will be raising it again on a number of occasions throughout this article.

In the efference-compensation scheme already described, the rotations are assumed to occur in a coordinate system that is fixed to the head or body. The brain sends commands to move the eye or head and so

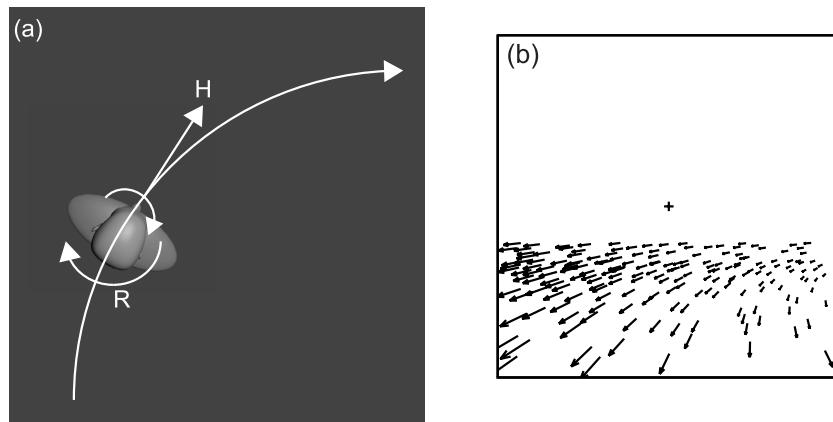
it has access to signals indicating that the eye or head is rotating at a particular rate and direction (Sommer & Wurtz, 2008). The effect of the rotation can be removed from the flow field as per Figure 3. Therefore, for rotations of the eyes relative to the head or the eye–head system relative to the body, the establishment of heading direction relative to the eye-gaze direction is possible in a straightforward manner.

### Rotation introduced by motion along curvilinear paths

For forward motion along curved paths, however, the problem is not as simple (Figure 4). The eye–head–body system undergoes a rotation relative to the world ( $R_w$ ) and so the visual-image motion on the retina is perturbed in the same way as when just the eye or head rotates (Figure 2a), but the brain has not sent any signals indicating that the eye should move relative to the head or the head relative to the body. Curvilinear self-motion generates image rotation, but there is no efference signal available to cancel it. The fact that we can solve the rotation problem while driving indicates that proprioceptive signals from our lower limbs are not essential for this process.

The actual instantaneous heading direction ( $H$  in Figure 4a) generates an FOE in the direction of heading (cross in Figure 4b), but this is obliterated by the image motion created by  $R_w$ .

A human psychophysical experiment examining the compensation of heading estimation during head turns indicated that for rotation cancellation to occur, the head had to be turned actively relative to the body



**Figure 4.** Motion along curved paths and curvilinear rotation. (a) As an observer moves along a curvilinear path with eyes aligned with the heading vector (tangent to curve,  $H$ ), the eye–head–body undergoes rotation  $R$ . (b) Vector flow field representing image motion that occurs while heading toward the cross and moving along a curved path that generates image rotation to the left. The heading is now difficult to detect.

(Crowell et al., 1998). For motion along curved paths, the eyes and head can remain fixed relative to the body, and so these head- or body-referenced sources of compensation are absent; yet the rotation is somehow detected and canceled successfully by the brain (Bertin, Israel, & Lappe, 2000; Li & Cheng, 2011; Rieger, 1983; Stone & Perrone, 1997; W. H. Warren, Mestre, Blackwell, & Morris, 1991). How this occurs remains a mystery. To determine rotation, does the visual system use the visual motion present in the image, some extraretinal source, or both?

As we move along curvilinear paths, distant points have image velocities that tend to be uniform and that can indicate the amount of rotation that is occurring (see points on horizon in Figure 4b). There have been a number of schemes suggested for how to register this rotation visually (see earlier). Despite the availability of visual information, the process of detecting the rotation need not be limited to just vision.

Many animals, including humans and nonhuman primates, have a vestibular system made up of the otoliths and semicircular canals to convey information about rotation and acceleration of the head relative to gravity (Angelaki & Cullen, 2008; Goldberg et al., 2012; Howard, 1982; Imai, Moore, Raphan, & Cohen, 2001). While there are countless studies on the vestibular system in general, unfortunately there is a relative paucity of research on human or nonhuman primate vestibular responses during curvilinear self-motion, because the experiments require mobile robots (Bertin & Berthoz, 2004; Bertin & Israel, 2005; Ivanenko, Grasso, Israel, & Berthoz, 1997) or large (and expensive) moving-base simulators to measure sensitivity correctly (Chen, DeAngelis, & Angelaki, 2011; Z. Cheng & Gu, 2016; Crane, 2014; Gu et al., 2006; MacNeilage, Turner, & Angelaki, 2010; Nooij, Nesti, Bulthoff, & Pretto, 2016; Takahashi et al., 2007).

MacNeilage, Turner, and Angelaki (2010) used such a moving-base simulator and established through the use of a rigorous signal-detection method that the mean angular rotation-rate threshold for humans was  $1.45^\circ/\text{s}$  (see also Grabherr, Nicoucar, Mast, & Merfeld, 2008). Is this accurate enough for the compensation and removal of the rotation from the combined rotation-and-translation flow field ( $V_{T+R} - V_R$ )? This is an empirical question that can be answered using models of heading estimation such as already described (Perrone & Krauzlis, 2008), and I address it later.

It is doubtful that the brain has limited itself to just a visual or just a vestibular solution to the curvilinear rotation-estimation problem, and many suggestions have been proposed for schemes that synthesize or integrate visual and nonvisual sources of information to solve the heading-rotation problem (e.g., Beintema & van den Berg, 1998; Freeman et al., 2010; Gu et al., 2006; Lappe, 1998; Pack, Grossberg, & Mingolla, 2001). We have previously suggested a scheme in which parts of the heading template maps (see Figure 2b) are inhibited depending on extraretinal signals (Perrone & Stone, 1994), but the details of how this might occur were never specified. Since that time, many more experiments have been carried out examining the neural signals in the primate visual system during combined visual–vestibular movement throughout a number of visual areas (Chen et al., 2011; Z. Cheng & Gu, 2016; Gu et al., 2006; Takahashi et al., 2007). This has led to a variety of models for how integration of the visual and vestibular signals could occur to aid heading perception (e.g., Butler, Smith, Campos, & Bulthoff, 2010; Fetsch, Turner, DeAngelis, & Angelaki, 2009; Gu et al., 2006), with many adopting a Bayesian approach (Niehorster, Cheng, & Li, 2010; Pouget, Deneve, & Duhamel, 2002).

The explanatory power of these multisensory models is limited, however, because they do not include a detailed model of the visual-flow-field processing stage. This means that they cannot be tested with the same stimuli used for human and animal psychophysical studies of heading estimation. In this article I will outline a model for how heading can be determined in the presence of curvilinear self-motion that includes a visual-flow-field front-end. I will show how the model can be used to explain a range of human psychophysical data on heading estimation.

I begin with the assumption that the visual system needs to compensate for the rotation that occurs during combined T+R self-motion in order to derive a pure translation field from which to determine heading direction and depth information. I will then determine how precise the rotation estimates need to be in order to account for human heading-estimation performance (Experiment 1). Based on the specification requirements of the rotation detectors, I will develop and outline a scheme for estimating rotation visually from the combined T+R flow field that occurs as we move along curvilinear paths (Model design). I then show that visual estimation of rotation on its own is not precise enough, and that vestibular signals need to be included to reliably estimate heading and rotation during curvilinear self-motion (Experiment 2). I then show how this new model can account for a range of existing human psychophysical data from self-motion perception experiments (Experiments 3–6).

## Experiment 1

### Precision requirements for rotation estimation during curvilinear self-motion

There exist many sets of human data indicating what constitutes accurate heading performance (W. H. Warren, 2003). Both the bias in heading estimates and their precision (as a threshold measure) have been studied. I will use the latter measure, because thresholds are the most commonly reported psychophysical data in heading experiments and they are amenable to a noise analysis. Given a collection of vectors across the image ( $V_{T+R}$ )<sub>i</sub>, how precise must the estimate of  $V_R$  be in order that  $(V_{T+R} - V_R)$ <sub>i</sub> results in sufficiently precise  $V_T$ <sub>i</sub> values across the image to provide heading estimates ( $Az_H$ ,  $El_H$ ) with a precision that matches human psychophysical data? The threshold  $\sigma_H$  for heading estimation with and without added rotation has been established as being in the region of 1.0°–1.5°, and this band is also often cited as the precision required for safe navigation (Cutting, 1986). Studies that have examined the threshold for heading during

curvilinear self-motion are less common, but the indication is that it is also in the 1.0°–1.5° range for moderate rotation rates (Banks et al., 1996; Stone & Perrone, 1997; W. H. Warren, 2003; W. H. Warren et al., 1991). I will use the study by W. H. Warren et al. (1991) as a representative example of these studies, because it used a range of stimulus conditions and explicitly reported on the heading thresholds. That study found heading thresholds in the region of 1.0°, but that depended on the type of simulated 3-D environment and the rate of rotation. Higher rates of rotation generated higher thresholds (~1.5°), as did ground planes.

Noise will be added to the rotation-compensation stage of the template-based model (Figure 2) to determine how precise the visual and/or vestibular signals need to be in order to produce heading thresholds less than or equal to the 1.0° limit determined by W. H. Warren et al. This analysis is agnostic as to the source of the compensation signal; it could be visual or vestibular or both. As mentioned, the 1.0° heading threshold pertains to what humans can achieve under cue-conflict conditions; we would expect the performance to be even better when the correct vestibular signal is available to the observer. I will show that, for the low rates of rotation used by W. H. Warren et al. the discrepancy between the visual and visual–vestibular estimates of heading is not large, and we can use the 1.0° threshold value as a reasonable benchmark against which the model performance can be compared. The model heading estimates need to achieve at least this level of precision if the model is to be considered an analogue of human curvilinear self-motion perception.

When discussing compensation signals from vestibular sources, it is important to note that the vestibular rotation threshold (1.45°/s) described previously was derived using a one interval presentation in which the observer was asked to make a two-alternative forced-choice decision as to whether it was a translation or translation-plus-rotation condition (MacNeilage, Turner, & Angelaki, 2010). The W. H. Warren et al. (1991) study used a yes/no procedure to determine the heading thresholds. In order to bring these into line, the two-alternative forced-choice threshold needs to be divided by  $\sqrt{2}$  (1.41) to make it compatible with yes/no-derived thresholds and to reflect the standard deviation of the underlying signal distribution (Green & Swets, 1974). Therefore, from now on I will use 1°/s as the vestibular rotation-rate signal's standard deviation ( $\sigma_r$ ). It has been reported that the rotation threshold is a function of the frequency of the test stimulus, but the 1°/s value is close to the estimate for an intermediate test frequency (0.2 Hz) that was obtained from a model fit to a range of test frequency data (Grabherr et al., 2008).

By adding noise (normally distributed with standard deviation  $\sigma_r$ ) to the rotation-compensation stage ( $V_{T+R} - V_R$ ), we can estimate the variability (standard deviation) in the heading estimates and compare this to human performance. This can be done separately for the  $V_R$  vector magnitude (rate) and for the direction. This estimation assumes that a mechanism similar to that shown in Figures 1 and 2 underlies human heading perception, with some type of vector-subtraction process involved (Perrone & Krauzlis, 2008). Given that the technique outlined in Figure 2 uses all the available motion information and noise-free vectors, it could be considered an ideal-observer model with derived  $\sigma_H$  values representing optimum heading performance. Note that estimates of  $V_R$  can also lack accuracy and can be biased to higher or lower rates. This would predict biases in heading estimates; these will be discussed separately later when the full model is tested.

## Method

Since random-dot clouds are the most commonly tested 3-D environments in heading experiments, the conditions used in the W. H. Warren et al. (1991) "Cloud" experiment 6 (see their figure 9a) were simulated. The observer speed was 1.9 m/s and the dots ranged from 10 to 40 m from the eye. The flow field that existed at the end of 3.7-s trial was used as input to the model and was generated using standard flow-field equations (see Perrone & Stone, 1994). The motion occurring over a window of 1/15 s (15-Hz sample rate) was used to represent the image flow. The field of view matched that used in the W. H. Warren et al. study and was set at 40° horizontal × 30° vertical. Heading direction was always (0°, 0°), and simulated rotation about the vertical axis was added (1.36°/s), replicating the W. H. Warren et al. radius condition of 50 eye heights.

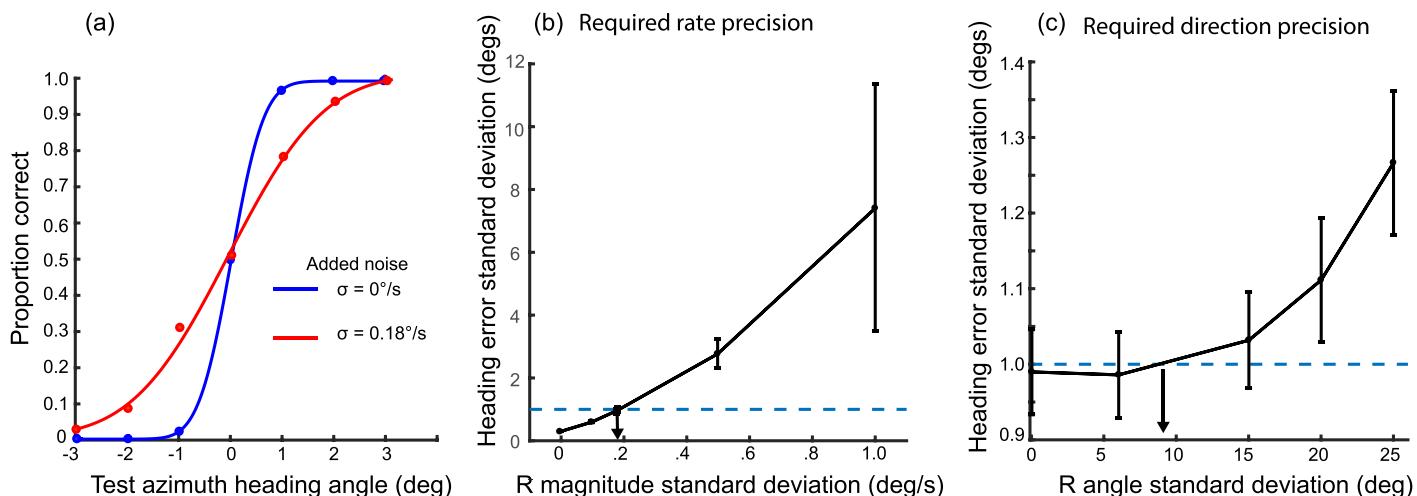
With just rotation, flow vectors are generated on the image plane with magnitude and direction ( $|V_R|, \phi_R$ ). From each vector ( $|V_{T+R}|, \theta_{T+R}$ ) in the test flow field is subtracted the equivalent of a compensation vector ( $|V_R| + N_r, \phi_R + N_d$ ), using the rules of vector subtraction.  $N_r$  and  $N_d$  are noise components for rate and direction, respectively, drawn from a normal distribution with means ( $\mu_r, \mu_d$ ) = (0, 0)°/s and standard deviations ( $\sigma_r, \sigma_d$ ). The vector subtraction occurs at the level of the heading detectors (see Perrone & Krauzlis, 2008), and the noise is assumed to arise from a single rotation-detecting unit. The same level of noise is therefore assigned to each heading detector. It is assumed that this is the noise that is applied during the epoch of the flow field (1/15 s in the model simulations). The noise occurring at the local vector level and over

many frames was simulated by running multiple trials (30) and generating a different flow field for each heading tested. Values of (0, 0.1, 0.18, 0.5, 1.0)°/s were used for  $\sigma_r$ . The  $\sigma_d$  values tested were (0°, 6°, 15°, 20°, 25°). When the  $\sigma_d$  values were tested,  $\sigma_r$  was set to 0.18°/s.

In order to derive the threshold for the heading template model, a receiver-operating-characteristic analysis was used to generate "modelometric" functions (Figure 5a) in the same way that neurometric (Britten, Shadlen, Newsome, & Movshon, 1992) and oculometric (Gegenfurtner, Xing, Scott, & Hawken, 2003; Kowler & McKee, 1987) functions have been used to find thresholds. Following the technique used by Gegenfurtner et al. (2003) to derive thresholds from eye-tracking data, 30 different flow fields were created for each of a range of azimuth heading directions (-3°, -2°, -1°, 0°, 1°, 2°, 3°) for a total of 210 simulation trials. Then for a range of criterion levels ( $c_i = -3^\circ$  to  $3^\circ$  in 0.1° steps), the proportion of times that the model's heading estimate was greater than  $c_i$  (hits) was plotted against the number of times the model's estimates for the 0° test input was greater than  $c_i$  (false alarms). This procedure produces a receiver-operating-characteristic curve (Gegenfurtner et al., 2003; Green & Swets, 1974) for each of the test heading directions. The area under each curve provides a proportion-correct value equivalent to that found from a two-alternative forced-choice psychophysical procedure (Green & Swets, 1974). An example of the derived proportion-correct values from the simulation is shown in Figure 5a, fitted using a cumulative Gaussian function (solid curves). The standard deviation of this function was divided by  $\sqrt{2}$  to make it compatible with a yes/no psychophysical procedure (Green & Swets, 1974) and with the data from W. H. Warren et al. (1991). This scaled value was used as the heading-threshold estimate for the model.

## Results

The threshold values for each of the  $\sigma_r$  and  $\sigma_d$  values listed previously are plotted in Figure 5b and 5c, with the error bars representing the standard deviation derived from 12 runs of the threshold-estimation procedure. For rate sensitivity (Figure 5b), it is apparent that the rotation precision (1.0°/s) that can be derived from a purely vestibular source (MacNeilage, Turner, & Angelaki, 2010) is insufficient to account for the human heading-threshold performance found by W. H. Warren et al. (1991; 1.0°, dashed line). It would result in heading discrimination thresholds of around 7° at the test rotation rate used in their experiment (1.36°/s). The rotation-rate discrimination performance needs to be around 0.2°/s to explain the human data (see arrow in Figure 5b).



**Figure 5.** Estimating the precision of  $\sigma$  required to account for human heading-discrimination thresholds from W. H. Warren et al. (1991, experiment 6; dashed line). (a) “Modelometric” curves from heading model tests found using a receiver-operating-characteristic signal-detection technique. Actual curvilinear rotation rate was  $1.36^\circ/\text{s}$  in the  $180^\circ$  direction. Blue solid points are for the case where no noise is assumed in the extraretinal vestibular signal that removes the effect of the rotation. The threshold (found from the standard deviation of a fitted cumulative Gaussian function; see curve) was  $0.4^\circ$ . The red points are for the case where the vestibular-signal rate threshold  $\sigma_r$  was set at  $0.18^\circ/\text{s}$ . The model’s heading threshold in this case was  $1.04^\circ$ . (b) Heading precision estimates for a range of noise levels (standard deviation of the noise added to the magnitude of the extraretinal rotation-cancellation signal). The rotation-compensation rate signal is perturbed with noise drawn from a normal distribution with different  $\sigma$  values. The required precision to account for human heading data is shown by the arrow. Error bars are the standard deviations from 12 simulation sessions (30 trials each) for each condition. (c) Precision estimates for rotation direction.

For the direction component  $\phi$  of  $V_R$  (Figure 5c), the precision of the estimates is a little less demanding; for the low rotation rate used in the test, at least, a modest  $\sigma_r$  value of  $8^\circ$  to  $10^\circ$  (arrow in Figure 5c) still produces heading-discrimination values less than or equal to the  $1.0^\circ/\text{s}$  human-data line. Currently there are no human data indicating what the vestibular direction-sensitivity  $\sigma_d$  equivalent value is for the  $\sigma_r = 1.0^\circ/\text{s}$  found by MacNeilage, Turner, & Angelaki (2010).

This analysis demonstrates that a purely vestibular solution to the problem of detecting and compensating for the rotation that occurs during motion along curvilinear paths is not possible given the reported precision of the vestibular rate signals (MacNeilage, Turner, & Angelaki, 2010). If we are to account for human heading-discrimination thresholds in the region of  $1.0^\circ/\text{s}$ , the rate estimates need to be fairly precise ( $\sigma_r \leq 0.2^\circ/\text{s}$ ). The direction component of the rotation is less demanding but still requires a precision of around  $10^\circ$  if heading performance is not to be overly compromised (Figure 5c).

I have now established the performance specifications for a system that determines heading in the presence of rotation created by motion along curvilinear paths via a compensation mechanism similar to that shown in Figure 2. Irrespective of how the rotation is measured, it needs to fit these specifications if a form of vector subtraction is the method used to recover the correct heading direction. I have demonstrated that

vestibular signals alone cannot provide the required level of precision. I will now develop a solution for estimating the rotation using both visual and vestibular signals and show that it is able to meet the specifications.

## Model design

### Visual estimation of rotation

I will begin with an outline that highlights the principle behind the model, and then I will provide a detailed description of the specific mechanisms. Consider a single vector in a flow field such as that depicted in Figure 2a with the rotation arising from yaw and pitch of the eye-head-body system relative to the world as the body moves along a curvilinear path. Figure 6a represents in vector form the image motion occurring at a retinal location  $(x, y)$  during a combined forward translation of the eye (creating image motion  $V_T$ ) and a body rotation about the vertical and horizontal world axes. Let the yaw and pitch rates equal  $(\omega_Y, \omega_P)$ , such that the body rotation generates image motion  $V_R$ . Roll rotation around the axis coinciding with the heading direction ( $\omega_{\text{Roll}}$ ) is not considered in the model because it creates a rotation vector orthogonal to the pure translation vector  $V_T$ , and hence does not change the length of this vector. It can therefore be ignored in the equations to follow.

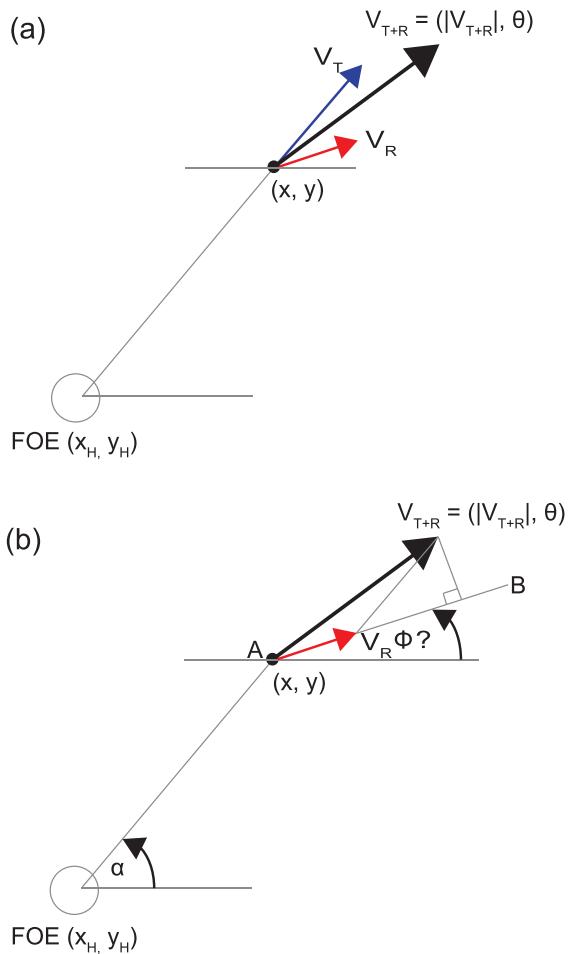


Figure 6. Image motion occurring on the retina at location  $(x, y)$  during motion along a curvilinear path. (a) For forward motion on its own (pure translation), the image motion is given by the blue vector. The length is a function of the distance of the point, and the direction is radially oriented out from the focus of expansion, which corresponds to the heading direction. For a particular unknown body rotation caused by motion along a curvilinear path, the image motion is given by the red vector. The resulting combined image motion (translation + rotation) is the vector sum of the blue and red vectors and is shown as the black vector in the figure. (b) The curvilinear rotation-detection problem. If only the black vector is available to the visual system, how can the red vector be estimated?

For a curved imaging surface such as the eye, the  $V_R$  vectors all have the same magnitude and direction. My technique for deriving the rotation is designed to be tested with video sequences derived from standard video cameras and to be used for computer-vision systems with regular cameras. I therefore assume a planar projection from the world onto the image plane. For wide-angle lenses this creates a pincushion distortion when a rotation vector field is projected onto it and the vectors are not all identical. The value of  $V_R$  varies slightly depending on which part of the image the motion is occurring in. However, the

amount of distortion is known, given knowledge of the optics (e.g., the field of view of the camera) as well as the vector location, and can be compensated for prior to the extraction of  $V_R$ . For the field-of-view values tested in this article the distortion is very small and has minimal impact on the estimates of rotation. The vectors in the flow field are assumed to be derived from video image sequences using some sort of flow-extraction algorithm (e.g., Perrone, 2012). If the model is to be used for wide field-of-view tests ( $>90^\circ$ ), it is recommended that a pincushion compensation algorithm be applied to the input image sequence (e.g., Bouguet, 2015). For extremely wide-angle inputs ( $>120^\circ$ ), some sort of hemispheric sensor system could be considered. In order to demonstrate the optimum performance of the model in the experiments reported later, I have applied a distortion-compensation algorithm to the flow field, and all of the  $V_R$  vectors are close to being identical.

The body rotation ( $\omega_Y, \omega_P$ ) creates image motion at  $(x, y)$  given by the vector  $V_{T+R}$ . The magnitude of the vector  $V_T$  is unknown because it depends on the (unknown) distance of the point in the world generating the image motion at  $(x, y)$ . We do know that the direction of  $V_T$  is pointed outward from the FOE location (Gibson, 1950) and is aligned with the radial direction out from a retinal position  $(x_H, y_H)$  corresponding to the unknown heading direction ( $Az_H, El_H$ ). Figure 6b depicts the problem that must be solved: Given  $V_{T+R}$ , is it possible to find  $V_R$  (and hence  $V_T$ )? For the compensation mechanism, we require just  $V_R$ , and not the rotation relative to the world; the removal of the rotation is all done in a retinal coordinate system (Perrone & Krauzlis, 2008). If we can find  $V_R$ , it is possible to recover  $(\omega_Y, \omega_P)$ , but it requires a transformation from retinal coordinates to world coordinates (Koenderink & van Doorn, 1975; Longuet-Higgins & Prazdny, 1980).

As previously mentioned, the magnitude of  $V_T$  is dependent on unknown variables such as the depth of the points in the world, but we can derive it from  $V_{T+R}$  once we have estimated  $V_R$ , because they are linked by the rules of vector addition. In the treatment to follow, all angles are measured relative to arbitrary horizontal and vertical axes on the image plane, with  $0^\circ$  corresponding to the rightward horizontal direction.

The angle of the vector  $V_{T+R}$  is equal to  $\theta$  and its magnitude is  $|V_{T+R}|$ . Let  $\alpha$  be the angle between the unknown FOE position  $(x_H, y_H)$  and the location of the  $V_{T+R}$  vector  $(x_i, y_i)$ . Let the angle of the (unknown)  $V_R$  vector be  $\phi$ . Note that this is defined in retinal coordinates. Using trigonometry, it can be shown that the magnitude of  $V_R$  is related to  $V_{T+R}$ ,  $\alpha$ , and  $\phi$  via the following equation:

$$|V_R| = |V_{T+R}| \cos(\theta - \phi) - \frac{|V_{T+R}| \sin(\theta - \phi)}{\tan(\alpha - \phi)}. \quad (1)$$

This can be simplified to

$$|V_R| = \frac{|V_{T+R}| \sin(\alpha - \theta)}{\sin(\alpha - \phi)}. \quad (2)$$

Figure 6b and Equation 1 highlight the fact that the magnitude of the  $V_R$  vector is primarily determined by the cosine component of the  $V_{T+R}$  vector projected onto the axis defined by the direction of the  $V_R$  vector ( $AB$  in Figure 6b). The projected vector length is given by  $|V_{T+R}| \cos(\theta - \phi)$ , and  $V_R$  can be found from this by subtracting a correction factor that is a function of  $|V_{T+R}| \sin(\theta - \phi)$  and a tangent term dependent on the heading and rotation directions (Equation 1). For very distant points, the magnitude of the  $V_T$  vector is very small and  $V_{T+R} \approx V_R$ . Therefore, the rotation could be determined approximately by summing just the  $|V_{T+R}| \cos(\theta - \phi)$  values across many vectors and a (constrained) set of potential  $\phi_i$  values (e.g.,  $0^\circ$ – $330^\circ$  in  $30^\circ$  steps).  $V_R$  could then be determined by finding the  $\phi_i$  value with the maximum activity. This is the scheme I originally suggested for purely visual estimation of rotation (Perrone, 1992). For some scenes with a dominance of distant points, it provides reasonably accurate results (see figure 9 in Perrone, 1992). However, for many situations the rotation rate is overestimated, and large errors for  $V_R$  and  $\phi$  can occur. For this reason, the full and exact version of the rotation equation with the  $\alpha$  and  $\phi$  values has been adopted.

There are two unknown values on the right side of Equation 2:  $\phi$  and  $\alpha$ . The first,  $\phi$ , is constrained to lie between  $0^\circ$  and  $360^\circ$  and, as will be shown later, can be subsampled relatively sparsely ( $30^\circ$  steps) without a great loss of precision in the rotation-direction estimates. The second unknown,  $\alpha$ , is directly related to the vector location ( $x_i, y_i$ ) and the (unknown) heading direction ( $Az_H, El_H$ ):

$$\alpha_i = \arctan(y_i - FOE_y / x_i - FOE_x), \quad (3)$$

where  $(x_i, y_i)$  is the image location of the vector and  $FOE_x$  and  $FOE_y$  are the  $x$ - and  $y$ -coordinates of the unknown FOE that results from movement along the unknown heading direction ( $Az_H, El_H$ ).

Equation 2 also signals that  $V_R$  is undefined for particular values of  $\alpha$  and  $\phi$  such that  $\alpha - \phi = 0$ . We can examine the way  $V_R$  changes across different values of  $\phi$  in order to gain an intuition as to how the new curvilinear detection algorithm works by considering the simple situation depicted in Figure 7a. I will begin with some very strong simplifying assumptions. Only four points are assumed to be visible in the visual field

(origins of vectors in Figure 7a), and the forward motion is set so that it is directed toward the center of the field ( $Az_H, El_H = (0^\circ, 0^\circ)$ ). This latter assumption is in place solely to illustrate one particular property of the new mechanism, and it will be removed later. At the same time that it is moving forward, the body is assumed to be rotating to the right at  $4^\circ/\text{s}$  about a vertical axis through the center of rotation of the head/body ( $\omega_Y = 4^\circ/\text{s}$ ,  $\omega_P = 0^\circ/\text{s}$ ); this generates the image motion indicated by the red vectors.

Figure 7b is a plot showing the values of  $|V_R|$  derived from Equation 2 for each of the four different vectors in Figure 7a and for a range of candidate  $\phi$  values ( $x$ -axis). The value of  $\alpha$  for each vector is assumed to be known exactly in this plot, and  $\phi$  has been sampled finely in  $1^\circ$  steps so that the curves are smooth. Also, combinations of  $\alpha$  and  $\phi$  that result in undefined values of  $|V_R|$  (see earlier) have been avoided by not considering  $|V_R|$  values greater than  $12^\circ/\text{s}$ . All of the curves cross (intersect) at the actual curvilinear rotation rate ( $4^\circ/\text{s}$ ) and the angle  $180^\circ$  from the actual plane of rotation ( $0^\circ$ ). This  $180^\circ$  flip is because the algorithm is detecting the image-motion direction, which is opposite to the actual direction of rotation.

By binning the values of  $|V_R|$  and  $\phi$  into a set of candidate rates (e.g.,  $0^\circ/\text{s}$  to  $16^\circ/\text{s}$  in  $1^\circ/\text{s}$  steps) and directions (e.g.,  $30^\circ$  steps), it is possible to see that the most commonly occurring value of  $|V_R|$  from Equation 2 occurs at the correct rate and direction (Figure 7c). The technique relies on the fact that the correct  $|V_R|$  value and direction occur the most often. Because image motion that results from rotation is independent of the depth of the points in the world (Koenderink & van Doorn, 1975; Longuet-Higgins & Prazdny, 1980), it is constant across the image (assuming the wide-angle edge-distortion effects have been corrected), as can be seen by the red vectors in Figure 7a. Each image location has the same 2-D-motion rotation component.

In terms of a neural implementation of the algorithm, each square in the 3-D plot of Figure 7c could be considered a neural element (curvilinear neuron) that is tuned to a particular  $|V_R|$  and  $\phi$  value. The tuning would come about via a particular (sine) weighting (based on Equation 2) being applied to the connection from the motion sensors at particular  $(x, y)$  image locations and the curvilinear detector unit. Neurons tuned to  $|V_R|$  and  $\phi$  would generate the most activity when the actual curvilinear rotation matched  $|V_R|$  and  $\phi$ .

This simple implementation of the curvilinear algorithm assumed that the heading direction was known (i.e., the values of  $\alpha$  in Equation 2 were known). This is obviously a big assumption and an unrealistic one, because we are attempting to find the rotation so that we can derive the heading (see Figure 3) and so that the algorithm becomes circular. This assumption can be removed by testing a number of candidate

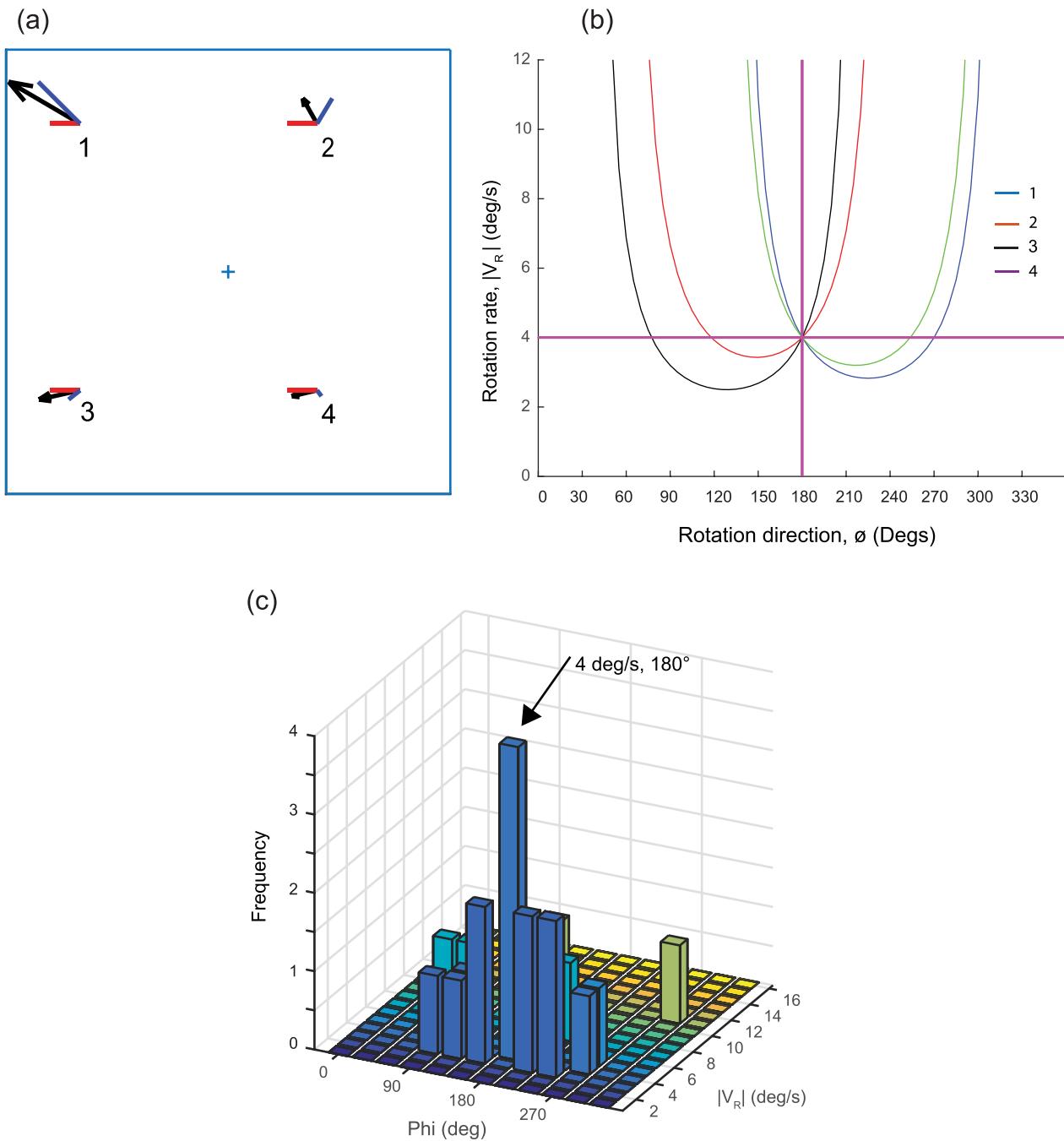


Figure 7. Illustration of the mechanism underlying the curvilinear rotation-estimation model. (a) Vector input flow field with just four vectors (black). Each vector is created from the vector sum of a translation vector (blue) expanding out from the heading direction (cross) and a rotation vector (red) created by rotation of the observer about a vertical axis to the right. (b) Graph showing the output of Equation 2 for different values of  $\phi$  while  $\alpha$  is set to a single value based on heading ( $Az_H, El_H = (0^\circ, 0^\circ)$ ). Each curve is for a different vector in the flow field. The curves cross at the correct rotation rate and direction because the rotation vector is common to all locations. (c) Binned  $R$  values in 3-D histogram format. The most frequently occurring value (peak) occurs where the curves cross in (b), and signifies the correct rotation.

heading directions in the same way that heading estimation has been carried out (see Figures 1 and 2). We have previously shown (Perrone, 1992; Perrone & Stone, 1994; Perrone & Stone, 1998) that heading can be sampled with a sparse array of candidate directions.

For the simulations presented in the remainder of the article I will use a set of azimuth and elevation values that range from  $-80^\circ$  to  $80^\circ$  in  $10^\circ$  steps (a  $17 \times 17$  regular array), but other configurations are possible (see Perrone & Stone, 1994).

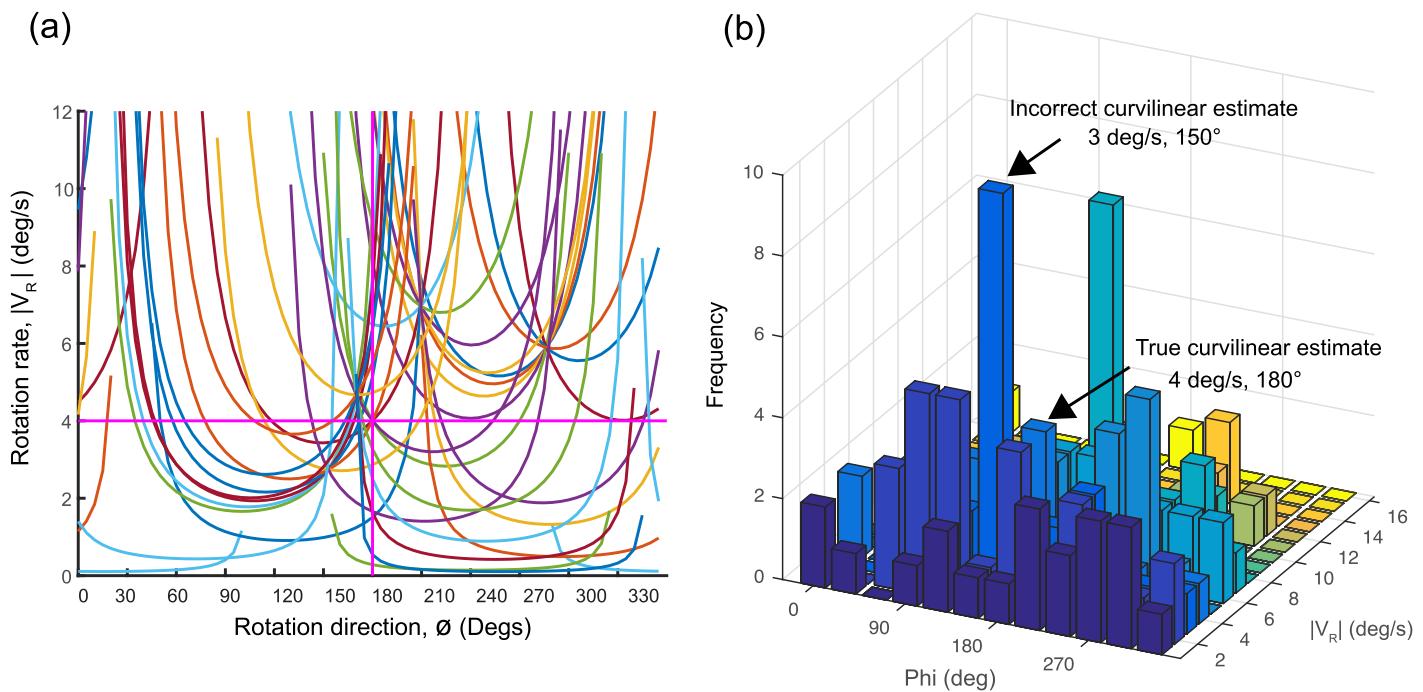


Figure 8. Multiple values of  $\alpha$  in Equation 2. (a) Plot of  $\phi$  versus  $|V_R|$  as in Figure 7, but with heading not constrained to a single value. (b) Histogram of binned  $R$  values. Peak occurs at incorrect rotation value.

For the following test, a subset of the full  $17 \times 17$  array of azimuth and elevation heading values was used ( $Az_H = -60^\circ$  to  $60^\circ$  in  $30^\circ$  steps,  $El_H = 0^\circ$ ) to prevent the plot becoming too dense with lines. If we base the values of  $\alpha$  in Equation 2 on this smaller set of ( $Az_H$ ,  $El_H$ ) values (using Equation 3), we obtain the  $\phi$ -versus- $|V_R|$  plot shown in Figure 8.

Again, this is only for the same four vectors shown in Figure 7a, but it is obvious that there is no single point of intersection and the distribution of  $|V_R|$  values is much more complex. This is for the case of a subset of the full  $17 \times 17$  array of possible ( $Az_H$ ,  $El_H$ ) values, but even with this small number, the  $\phi$ -versus- $|V_R|$  plot has many more overlapping curves. Applying the same binning technique that was used in Figure 7 reveals that the peak activity occurs at the incorrect ( $|V_R|$ ,  $\phi$ ) location of  $(3^\circ/\text{s}, 150^\circ)$ . This is an extreme case, and it will be shown later that a larger number of vectors will usually result in the correct peak being identified, but I am using it to illustrate how vision alone (at least for a small number of vectors) is not adequate for curvilinear detection based on the algorithm specified in Equation 2.

As pointed out previously, we have vestibular information from the otoliths and semicircular canals that could be used to aid our estimation of the curvilinear rotation that is occurring as the body moves along a circular path. I now demonstrate how relatively imprecise values for the angle of rotation and the rate of rotation can help improve the estimates for ( $|V_R|$ ,  $\phi$ )

derived from the curvilinear detection mechanism just outlined (Equation 2).

If we have some indication as to the angle of the rotation direction from a vestibular signal, it can be used to constrain the solution space regarding the direction of the rotation  $\phi$ . The semicircular canals, possibly in conjunction with the otoliths (Angelaki & Cullen, 2008), could signal that rotation is occurring in a plane somewhere in the region  $\phi \pm \gamma$  (Figure 9a). If, for example, the actual plane of rotation =  $0^\circ$  (i.e.,  $\phi = 180^\circ$ ) and  $\gamma = 30^\circ$ , then the solution space could be constrained to the region  $180^\circ \pm 30^\circ$  (see vertical dashed lines in Figure 9a). This constraint greatly reduces the possible curvilinear solutions and eliminates many of the incorrect regions of overlap in the ( $|V_R|$ ,  $\phi$ ) space (compare Figures 8a and 9a).

Another possible vestibular constraint comes from the semicircular-canal signals indicating the rate of rotation (see horizontal dashed lines in Figure 9b). If the vestibular system were able to indicate the rotation rate within some band of the true value (Figure 9b), then this could constrain the range of possible solutions being generated visually (top and bottom horizontal lines). This constraint in the rate estimates helps to eliminate the erroneous junctions in the ( $|V_R|$ ,  $\phi$ ) solution space.

It is also possible that signals from the otoliths could be used to constrain the heading direction and hence the range of  $\alpha$  values used. There is psychophysical evidence that a vestibular signal indicating forward translation can influence the perception of curvilinear

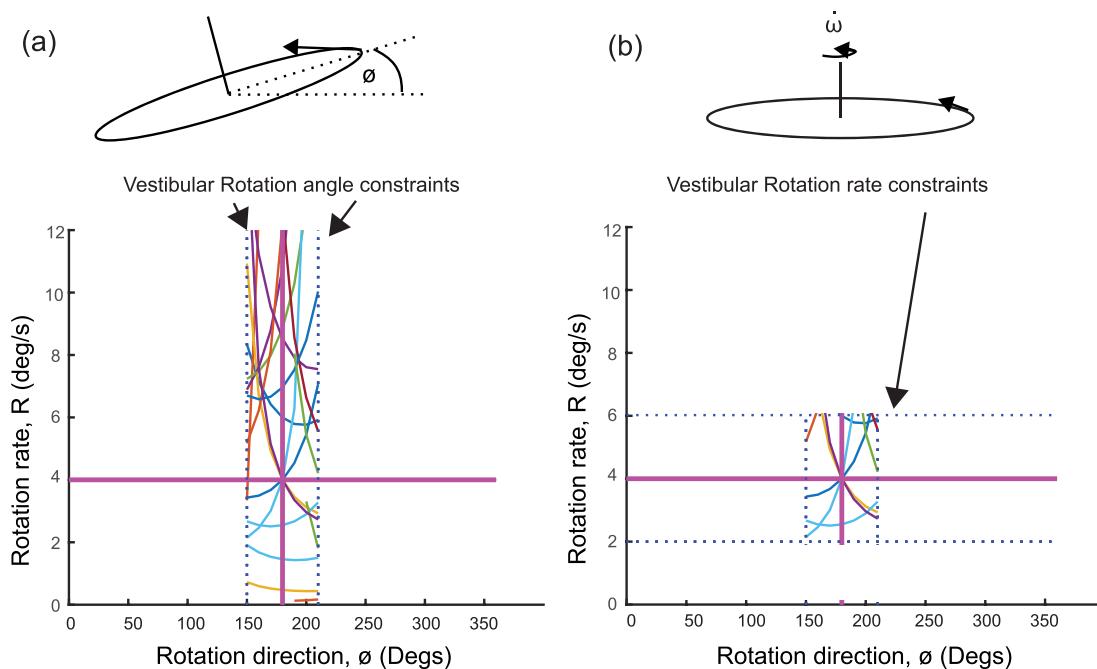


Figure 9. Using vestibular signals to constrain the vision-derived curvilinear rotation outputs. (a) If some information is available via the vestibular system as to the direction of the curvilinear rotation  $\phi$ , then parts of the solution space can be eliminated. The width of the constraint boundaries depends on the precision of the vestibular signals. (b) Vestibular information concerning the rotation rate can similarly constrain the visual solution.

paths (Bertin & Berthoz, 2004). However psychophysical measurements of heading-discrimination ability from purely vestibular sources indicate a range of estimation thresholds from medium values in the region of  $6^\circ$ – $9^\circ$  (MacNeilage, Banks, DeAngelis, & Angelaki, 2010; Telford et al., 1995) through to very high thresholds of  $30^\circ$  (Nooij et al., 2016). If vestibular heading were to be included as a constraint, it could not be a very narrow one based on human data. Conflicting evidence from monkey studies indicates quite precise heading-discrimination thresholds (vision:  $1.2^\circ$ ; vestibular:  $1.1^\circ$ ; both:  $0.4^\circ$ ), with a strong influence of a vestibular heading signal (Fetsch et al., 2009; Fetsch, DeAngelis, & Angelaki, 2010). I have found the addition of an  $\alpha$  constraint to have minimal impact on the performance of the curvilinear model over the relatively narrow range of heading directions currently tested. It may become more useful when heading is allowed to span a wider range. For now, I have omitted this possible source of vestibular information, but I will revisit the option once more human psychophysical data are available.

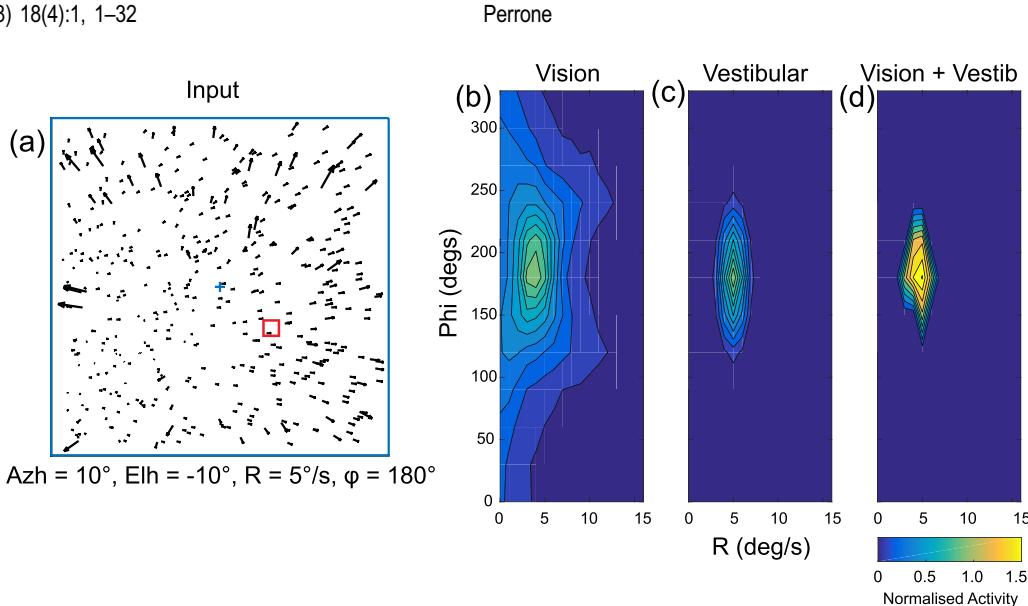
All in all, it can be seen that relatively imprecise vestibular signals could help refine the visual estimate of the rotation. It should be noted that these signals would be available only during transient changes in rotation. They would weaken and fall off during constant motion along a curved path (Guedry & Lauver, 1961). Just how accurate the final visual-

vestibular rotation estimates need to be for accurate heading estimation will be addressed later. Before this can be done, the parameters of the model need to be specified, as well as the nature of the readout that will be used to obtain the final curvilinear rotation estimate from the population activity across the ( $|V_R|$ ,  $\phi$ ) units.

#### Visual–vestibular curvilinear path-estimation model

The mechanisms and different output stages of the model are shown in Figure 10. An example of a test input flow field is shown in Figure 10a. The assumed ego-speed was 1.5 m/s toward a random cloud of approximately 450 points that ranged from 2 to 30 m, and the field of view was  $60^\circ$  horizontal  $\times$   $60^\circ$  vertical. This is for the case in which the observer is looking in a direction  $10^\circ$  to the left and  $10^\circ$  above the heading direction while moving on a curved path such that the body rotation rate  $|V_R|$  is  $5^\circ/\text{s}$  to the left around a vertical axis. The image-motion direction  $\phi$  we are trying to determine is therefore  $180^\circ$ . Given the observer's line-of-sight direction, the (unknown) heading ( $Az_H$ ,  $El_H$ ) =  $(10^\circ, -10^\circ)$ , shown as the red square in Figure 10a.

Equation 2 is applied to each of the vectors in the flow field. For the previous demonstrations of the curvilinear model, the sampling of the  $\phi$  values used  $1^\circ$  steps to provide continuity to the curves in the plots, but values between  $0^\circ$  and  $330^\circ$  in  $30^\circ$  steps will now be



**Figure 10.** Different stages of the new visual–vestibular curvilinear rotation-estimation model. (a) Input vector flow field. (b) Contour plot showing normalized curvilinear rotation activity across different  $R$  and  $\phi$  values. (c) Vestibular signal in the form of a 2-D Gaussian located at the correct rotation value but with the spread based on the sensitivity of the vestibular signal. (d) Sum of visual and vestibular distributions.

used. The plane of rotation is actually constrained to the range  $0^\circ$  to  $150^\circ$  if both positive and negative values of  $V_R$  are considered, but for ease of plotting,  $V_R$  is restricted to positive values and a wider  $\phi$  range has been adopted. Much of this range for the plane of rotation is probably out of the bounds encountered during normal human locomotion (except perhaps for the directions close to the cardinals). The rate  $R$  is sampled in  $1^\circ/\text{s}$  steps from  $0^\circ/\text{s}$  to  $16^\circ/\text{s}$ , but this is arbitrary and could be sampled more or less finely and over a greater range if future data become available as to the limits of human curvilinear path detection.

As mentioned before, the value for  $\alpha$  in Equation 2 is sampled using an array of candidate heading directions spanning  $-80^\circ$  to  $80^\circ$  in  $10^\circ$  steps for both azimuth and elevation. There is no specific justification for these parameter selections other than the fact that the model performs well with these settings and they simplify the coding of the model simulations. Heading space could be sampled using nonlinear schemes that extend over a greater range (see Perrone & Stone, 1994). The model performance turns out to be robust to the choice of  $\alpha$  sampling values.

Instead of the 3-D histogram plots used earlier (e.g., Figure 8b), the activity is now depicted in each of the different curvilinear detectors as a contour plot (Figure 10b). The activity distribution is normalized by dividing the output at each ( $|V_R|$ ,  $\phi$ ) location by the peak of the distribution, so that the values range from 0 to 1.0. The input vector field produces a distribution with a peak close to the true value, but it is quite broad (Figure 10b). In order to refine the  $V_R$  estimate, a vestibular signal is added to the distribution in the form of a 2-D

Gaussian with amplitude = 1.0, mean ( $\mu_R$ ,  $\mu_\phi$ ) = ( $|V_R|$ ,  $\phi$ ), and standard deviations  $\sigma_r$  and  $\sigma_d$  for the  $R$  and  $\phi$  directions, respectively. For the Figure 10 simulation,  $\sigma_r$  was set to  $1.0^\circ/\text{s}$  based on the data from MacNeilage, Turner, & Angelaki (2010), and  $\sigma_d$  was set to  $30^\circ$ . Note that the  $\phi$  direction for the visual signals is in a retinal coordinate system. This is what is required for the vector operation required for the rotation-compensation mechanism to work. However, the signal from the vestibular system is assumed to be relative to the world (defined by gravity). If the head is tipped sideways, for example, while moving around a curved path, then the visual and vestibular  $\phi$  values need not line up. I am assuming that the appropriate transformations have occurred (Angelaki & Cullen, 2008) and that the added vestibular rotation-direction signal is aligned with the correct  $\phi$  direction.

The vestibular distribution is shown in Figure 10c and represents an imprecise signal from the vestibular system indicating a particular rotation rate and direction but with noise associated with it, so that the distribution is fairly broad. For cases in which the rate estimate is  $0^\circ/\text{s}$ , the distribution along the  $\phi$  axis is assumed to be very broad ( $\sigma_d = 180^\circ$ ), because with zero rotation, direction is not defined. The values of  $\sigma_d$  for  $R = 1^\circ/\text{s}$ – $3^\circ/\text{s}$  were similarly set to  $100^\circ$ ,  $90^\circ$ , and  $45^\circ$ , respectively, in the model to reflect a lack of direction precision at low rotation rates. These values were not found to be critical in the simulations reported in this article.

There are many ways in which the information from the two domains (vision and vestibular) could be combined, and there is a large literature on combina-

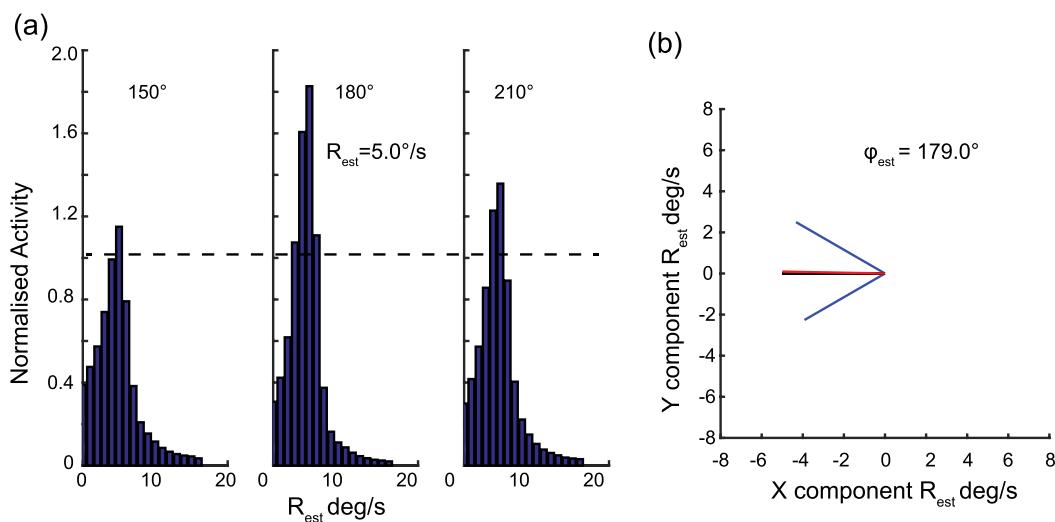


Figure 11. (a) Activity at different  $R$  values for three different  $\phi$  angles. Dashed line is threshold used to zero values for centroid-estimation stage. It is at 55% of the peak value across the whole activity distribution. (b)  $R$  centroid estimates as a function of angle represented as vectors. The model output direction estimate (red vector) is based on the angle of the resultant vector.

tion rules (Ernst & Banks, 2002; Fetsch et al., 2010; Landy, Maloney, Johnston, & Young, 1995; Rohde, van Dam, & Ernst, 2016). However, I have found that a simple addition of the two distributions is adequate to explain a lot of existing heading data. Again, future psychophysical experiments may provide additional evidence to suggest refinements to this combination stage, but for now I will adopt the simplest strategy.

Once the two activity maps have been added (Figure 10d), a clear peak appears around the correct curvilinear rotation-rate and direction values. A number of options present themselves as to how the output can be read out from the combined vision–vestibular activity (VVA) distribution. I have opted for a weighted vector-average scheme based on the commonly used population vector decoder (Georgopoulos, Schwartz, & Kettner, 1986). This choice was made because it makes greater use of the information in the distribution and accounts for some psychophysical data when visual–vestibular conflict is operating (see later).

A slice through the summed distribution in Figure 10d is shown in histogram form in Figure 11a for  $\phi$  angles of 150°, 180°, and 210°. The distributions of  $R_{\text{est}}$  values are not symmetrical, and so it is possible that an estimate for  $R$  based on the centroid could be biased upward if too much of the long tail is included. A more accurate estimate would result if some of the activity were excluded from the centroid estimate via some sort of thresholding that prevents activity below a certain level being passed to the centroid-estimation stage. On the other hand, the elimination of too much activity (by using a very high threshold value) removes information that could help determine the correct  $R$  or  $\phi$  value. I have opted for a threshold value of 0.55 of the peak

across the whole distribution (see dashed line in Figure 11a) because it produces the most accurate rotation estimates for the stimulus conditions tested. This value has been fixed and used in all of the simulations described in this article—except for one where human psychophysical data was matched (see Figure 16 later). Therefore, for each candidate rotation direction  $\phi_j$ , a weighted vector average is calculated to provide an estimate of the rotation rate  $R_{\text{est}}$ :

$$R_{\text{est}}(\phi_j) = \frac{\sum_i^{nr} R_i \text{VVA}_i}{\sum_i^{nr} \text{VVA}_i}, \quad (4)$$

where  $nr$  is the number of  $R$  values with nonzero VVA values after the threshold has been applied and  $R = 0^\circ/\text{s}$  to  $16^\circ/\text{s}$  in  $1^\circ/\text{s}$  steps. For the test shown in Figure 10, the  $R_{\text{est}}$  value comes out as  $5.0^\circ/\text{s}$  for the  $180^\circ$  direction, which is a perfect match to the true  $5^\circ/\text{s}$  value. A small number of other  $\phi$  directions also generate nonzero  $R_{\text{est}}$  outputs, and the final output is assumed to be the maximum across the 12 different  $\phi_j$  angles.

For the estimate of the rotation direction  $\phi_{\text{est}}$ , the nonzero  $R_{\text{est}}$  values for all directions  $\phi_j$  are represented as vectors (blue vectors in Figure 11b). For each angle  $\phi_j$  ( $j = 0^\circ$  to  $330^\circ$  in  $30^\circ$  steps), the vector sum of the individual vectors represented by magnitude  $R_{\text{est}}(\phi_j)$  and direction  $\phi_j$  are summed using vector summation, and the angle of the resultant vector is used as an estimate of the rotation direction  $\phi_{\text{est}}$ . This scheme makes use of interpolation between the  $30^\circ$  sampling used for  $\phi_j$  and tends to be more accurate than a mechanism that uses just the maximum  $R_{\text{est}}$  direction. For the test shown in Figure 10a, the rotation direction was found to be  $179.0^\circ$  (red vector in Figure 11b), which is very close to the true input direction ( $180^\circ$ ).

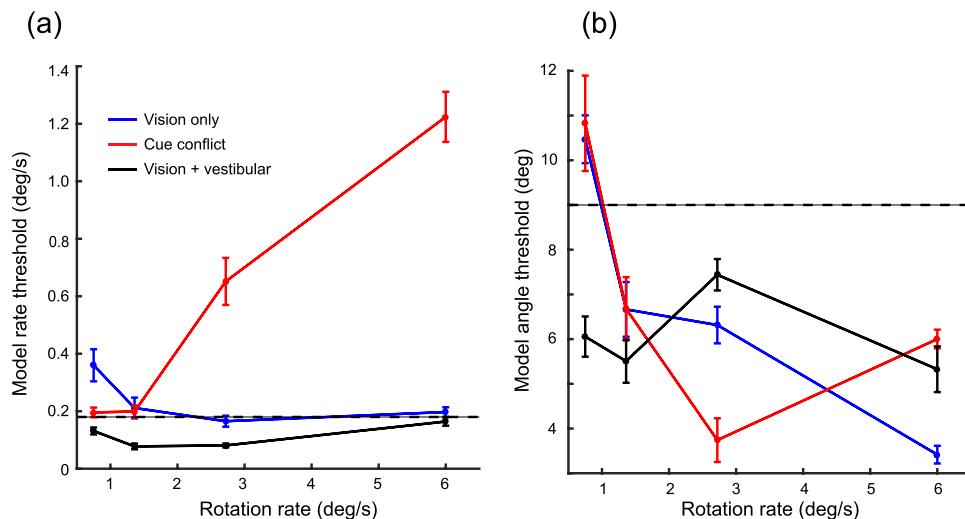


Figure 12. Curvilinear model threshold estimates. The error bars represent the standard deviation from 12 simulations. (a) Threshold for rotation-rate estimates  $R_{\text{est}}$ . Each curve shows data for different types of vestibular compensation. Horizontal dashed line is the threshold level required to account for human psychophysical heading data (see Figure 4a). (b) Threshold for rotation-angle estimates  $R_{\text{ang}}$ . Horizontal dashed line is precision level required to account for human heading data (see Figure 5c).

Note that the absolute radius of the path in the world being traveled cannot be determined from the output ( $R_{\text{est}}, \phi_{\text{est}}$ ) unless one has knowledge of the observer's forward speed  $T$ , since the radius =  $T/R$ . Similarly, obtaining the rotation relative to the world ( $\omega_Y, \omega_P$ ) requires a transformation from the retinal-coordinate-frame-based  $\phi_{\text{est}}$  value to a world coordinate frame, and this most likely requires a vestibular signal from the otoliths (Angelaki & Cullen, 2008). The retina-based  $\phi_{\text{est}}$  output is sufficient for the purposes of rotation compensation, however, because the operation occurs in a retinal coordinate system (Perrone & Krauzlis, 2008).

## Experiment 2

### Model threshold estimates

The new model was tested to see if it was able to achieve the criterion threshold estimates previously derived for successful rotation compensation during heading estimation (Figure 5). The thresholds for rotation rate and direction were determined using the same receiver-operating-characteristic technique adopted for heading (Figure 5).

### Methods

The same stimulus conditions (based on W. H. Warren et al., 1991) were used as described for Figure 5. For determining the model rotation-rate thresholds,

a set of rates (rate  $\pm$  50%, 25%, 12.5%, and 0%) were used to test the curvilinear model over 30 trials, with a different input field used each time. The generated modelometric functions (see Figure 5a) were then used to obtain the smallest change in the value of the rate that could be distinguished. For establishing the rotation-direction thresholds, a set of directions that spanned the test direction ( $180^\circ \pm 20^\circ, 10^\circ, 5^\circ$ , and  $0^\circ$ ) were inputted into the curvilinear model. For both the rate and direction thresholds, a range of rotation rates were tested ( $0.75^\circ/\text{s}, 1.36^\circ/\text{s}, 2.72^\circ/\text{s}$ , and  $6^\circ/\text{s}$ ), as well as three different types of model conditions: (a) purely visual with no vestibular rotation signal present; (b) cue conflict with the vestibular distribution centered on  $0^\circ/\text{s}$ ; (c) vision and vestibular (as per Figure 10d). For each condition and rotation rate, 12 trials were run, and the mean thresholds (and standard deviations) are plotted in Figure 12a. Although a purely vision-based test is unrealistic (except for observers with a dysfunctional vestibular system), it is included in this and the following tests in order to demonstrate how the addition of vestibular inputs alters the model's output.

### Results

For the vision-only condition (blue curve in Figure 12a), the precision of the curvilinear rate estimates is inadequate at very low rotation rates to account for the heading data of W. H. Warren et al. (1991). It was established earlier (Figure 5) that the threshold needed to be equal to or below  $0.18^\circ/\text{s}$  (dashed line in Figure 12a). Curvilinear estimation using just vision meets the

criterion at intermediate rates but is just above it at the highest rate used in the tests ( $6^\circ/\text{s}$ ).

For the situation that replicates most human psychophysical heading experiments, where a cue-conflict situation exists (the vestibular signal indicates  $0^\circ/\text{s}$ ), the model's threshold estimates (red curve in Figure 12a) are close to the 0.18 criterion line at low rates of rotation and so are consistent with the W. H. Warren et al. data; the model's curvilinear estimation system is able to provide curvilinear rate estimates with sufficient precision to account for the human heading data. At higher levels of actual rotation rate (smaller radii of curvature) the thresholds rise rapidly, and this would lead to higher thresholds and greater variability in the heading estimates derived from these rotation estimates. This predicted rise in thresholds with increasing rotation rate is consistent with the W. H. Warren et al. data (see their figure 9).

When both vision and a correct vestibular signal are incorporated into the model (black curve in Figure 12a), it generates threshold values below the 0.18 criterion level for all rotation rates tested. The means ( $SDs$ ) for each of the different rotation rates were 0.13 (0.01), 0.08 (0.01), 0.08 (0.01), and 0.16 (0.01). For the direction estimates (Figure 12b), threshold values from the model were all below the  $9^\circ$  limit required for accurate heading estimation (Figure 5b) except for the very lowest rotation rate tested ( $0.75^\circ/\text{s}$ ). The model is therefore able to generate a sufficiently precise signal to account for the  $1^\circ$  heading threshold figure required for safe navigation.

## Experiment 3

### Overall model performance: Vision only

The previous tests of the model span only a very limited range of possible input parameters; heading was always set such that  $(\text{Az}_H, \text{El}_H) = (0^\circ, 0^\circ)$ , and the rate and direction of the curvilinear rotation were limited to a few values to mimic the conditions used in the human psychophysical experiments. The field of view of the input flow fields was also quite narrow. In order to test the model more extensively, a wider range of input parameters was chosen.

### Method

The model was tested with 100 randomly generated input flow fields with parameters drawn from uniform distributions (with the range shown in square brackets) as follows: test rotation speed  $R$  [ $0^\circ/\text{s}, 10^\circ/\text{s}$ ]; test rotation angle  $\phi$  [ $0^\circ, 360^\circ$ ]; test azimuth heading  $\text{Az}_H$

[ $-20^\circ, 20^\circ$ ]; test elevation heading  $\text{El}_H$  [ $-20^\circ, 20^\circ$ ]. The field of view of the input flow fields was  $60^\circ \times 60^\circ$ , and the 3-D virtual world was made up of random dots occupying a range from 2 to 30 m so that around 120 dots were present in the field. The simulated observer's forward speed was 1.5 m/s. Tests were carried out with only the visual input (in which case the activity of the vestibular distribution in the model was set to 0) and for the situation where both vision and vestibular signals were available.

## Results

Figure 13 shows the results for the vision-only test. Figure 13a is a plot of actual test rotation rate versus the model estimates. The blue line represents perfect performance. Figure 13b is a histogram of the errors ( $R_{\text{test}} - R_{\text{est}}$ ). The root-mean-square error (RMS) was  $0.91^\circ/\text{s}$  in this case, and this will be the metric used to summarize the rotation-speed estimation performance of the model in the simulations that follow. For assessing direction error, a metric was developed based on that used in circular statistics to avoid wraparound problems:  $\text{error}_d = 1 - \cos(\phi_{\text{test}} - \phi_{\text{est}})$ . This ranges from 0 (no error) to 2.0 ( $180^\circ$  error). Figure 13c shows the direction error using this metric and plotted in the form of a scatterplot, with the mean shown in red. Figure 13d is a histogram showing the distribution of the errors across the 100 test trials. For the vision-only condition the mean error was 0.12, which corresponds to an angle error of  $28.4^\circ$ .

### Overall performance: Vision and vestibular

When a vestibular signal was incorporated into the model's mechanisms, the data shown in Figure 14 resulted. Figure 14a shows the rate-estimation errors in histogram form for one simulation run (100 trials). The RMS error for this run was  $0.3^\circ/\text{s}$ , and the direction mean error was 0.04 ( $16.3^\circ$ ). In order to gauge the robustness of these estimates (since they depend on the random flow-vector locations), 12 simulations were run, and the means and standard deviations of the 12 different RMS- and angle-error terms were found for the vision-only and the vision-plus-vestibular conditions. The data are shown in Figure 14c and 14d. The rate means ( $SDs$ ) for the vision and vision-plus-vestibular conditions were 0.94 (0.06) and 0.31 (0.02), respectively. For direction, the values were 0.13 (0.03) and 0.06 (0.02).

Rotation-rate thresholds depend on the rate of rotation (Mallery, Olomu, Uchanski, Militchin, & Hullar, 2010), but over the small range of rates tested here ( $0^\circ/\text{s}–10^\circ/\text{s}$ ) it is in the region close to the  $1.45^\circ/\text{s}$

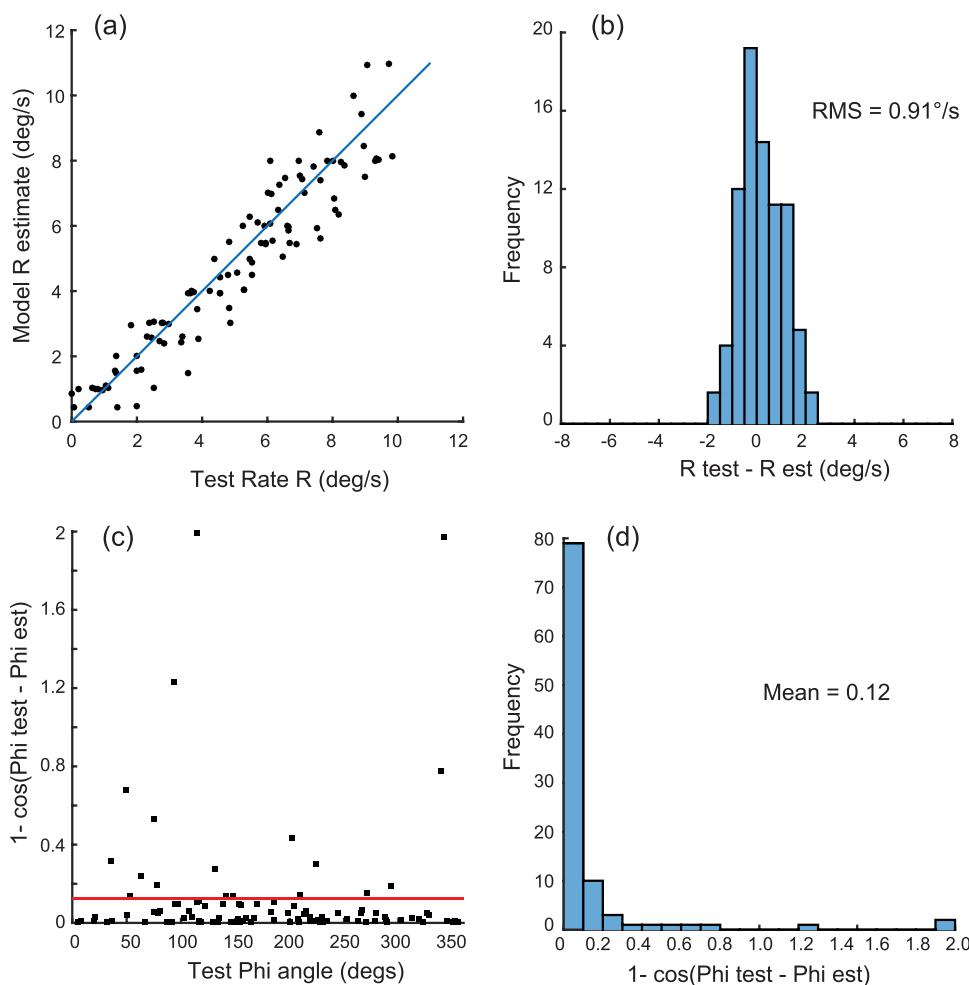


Figure 13. Test of the model when only the visual input is used. (a) Scatterplot of test rate  $R$  versus the model estimate for  $R$ . Data from 100 trials with randomly generated input flow fields. (b) Histogram showing frequency of rate errors. (c) Direction error as a function of actual test rotation angle. (d) Frequency of direction errors.

value reported by MacNeilage, Turner, & Angelaki (2010). If we assume that this threshold value is the same across a range of directions and heading values, we can use the standard deviation of the underlying signal distribution in the MacNeilage, Turner, & Angelaki signal-detection study ( $1.45/\sqrt{2} = 1^\circ/\text{s}$ ) as an indication of the RMS value that would be obtained using just a vestibular signal. This is shown in Figure 14c.

The performance of the model when only the vision module is used is close to this estimate. However, when both vision and vestibular signals are used, the model performance improves by over 50%; the model curvilinear rate estimates when both vision and vestibular signals are used are better than what either of them can generate on their own. For the rotation-direction estimates, there are no data equivalent to the  $1^\circ/\text{s}$  rate threshold from MacNeilage, Turner, & Angelaki, but the incorporation of a vestibular signal produces a mean RMS error that is 46% of the vision-only error rate (Figure 14d).

## Experiment 4

### Model performance under different input and parameter conditions

The previous tests were carried out with a particular input stimulus (random-dot cloud) with many vectors. To test how robust the model is and to demonstrate that the performance is not contingent on just the specific stimulus type adopted for these tests, another series of simulations were carried out in which the number of vectors was manipulated ( $n=4, 8, 16, 32, 64$ , and  $128$ ) as well as the configuration of the simulated 3-D world (cloud, single vertical plane, or ground plane). In addition, the free parameters in the model were modified to demonstrate how sensitive it is to a particular choice of parameter values and to the effect of pincushion distortion introduced via planar-projection surfaces.

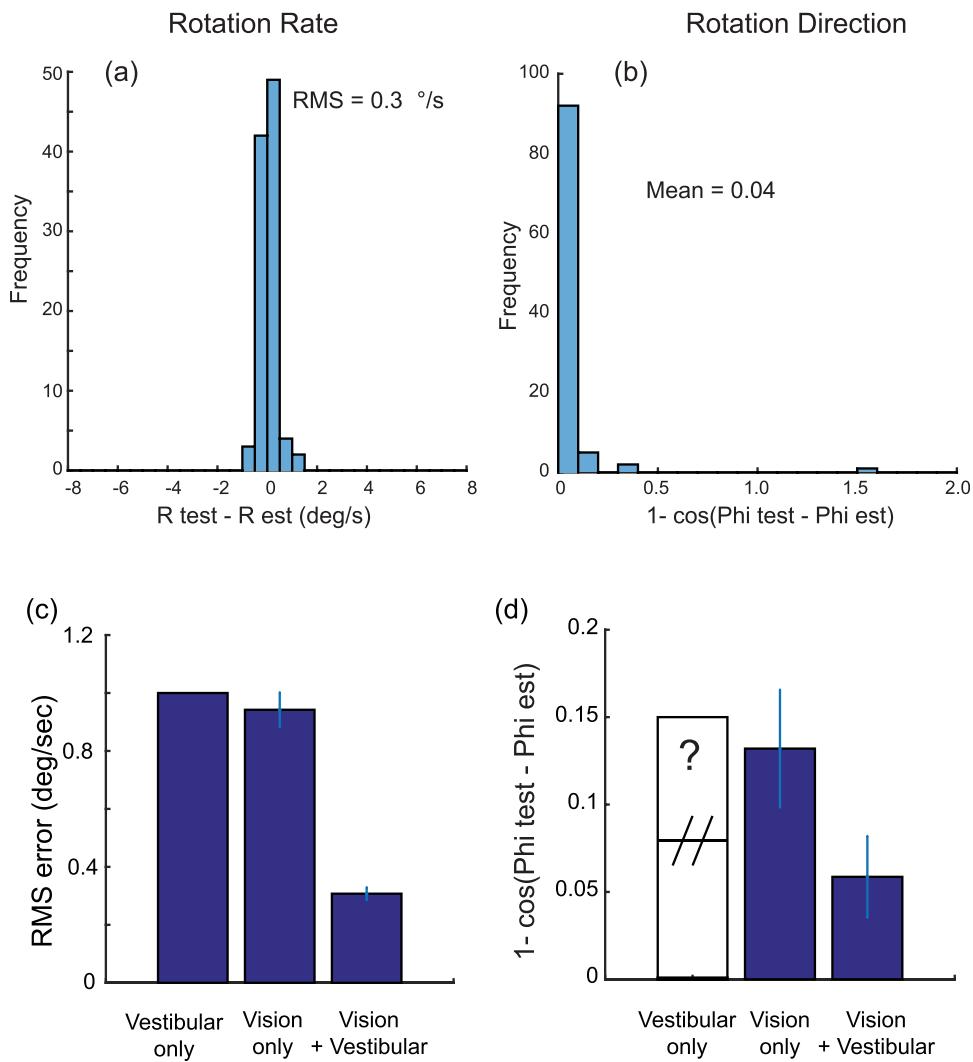


Figure 14. Test of model when both visual and vestibular information is used. (a) Frequency distribution of rate errors. (b) Frequency distribution of direction errors. (c) Root-mean-square rate errors for vision-only and vision-plus-vestibular conditions. The vestibular-only condition is an estimate based on data from MacNeilage, Turner, & Angelaki (2010). Error bars represent one standard deviation of the root-mean-square values obtained from 12 simulations of the model. (d). Comparison of direction errors across the two different conditions. There are currently no data for the expected error from a vestibular-only condition.

## Methods

The cloud is the same one used in the previous simulations, and the vertical plane was located at 12 m from the moving eye and contained approximately 200 points. The ground plane was located at 1.6 m below eye level and extended out to 45 m. The ego-speed was set at 1 m/s. All model parameter values were set as per the description given for the Figure 10 test. The results are shown in Figure 15.

## Results

For the number-of-vectors test (Figure 15a), the RMS errors for rate started high for the lowest number

of vectors but quickly reached an asymptote down to a performance level of around 0.3°/s. As long as approximately 16 or more vectors are available, the model rate estimates have very low RMS values. For the different world types, the means (*SDs*) for the three cases (cloud, single plane, ground plane) were 0.27 (0.02), 0.38 (0.03), 0.26 (0.02) for rate and 0.046 (0.024), 0.059 (0.025), 0.05 (0.024) for direction.

Performance was significantly different across the three different world types for rate-error RMS,  $F(2, 33) = 157.6, p < 0.001$ , but not for direction,  $F(2, 33) = 0.81, p > 0.05$ . Post hoc tests showed that the single plane was worse than both the cloud and the ground plane for rate, but the cloud and ground plane did not differ significantly from each other. The poorer performance with a single vertical plane mirrors what

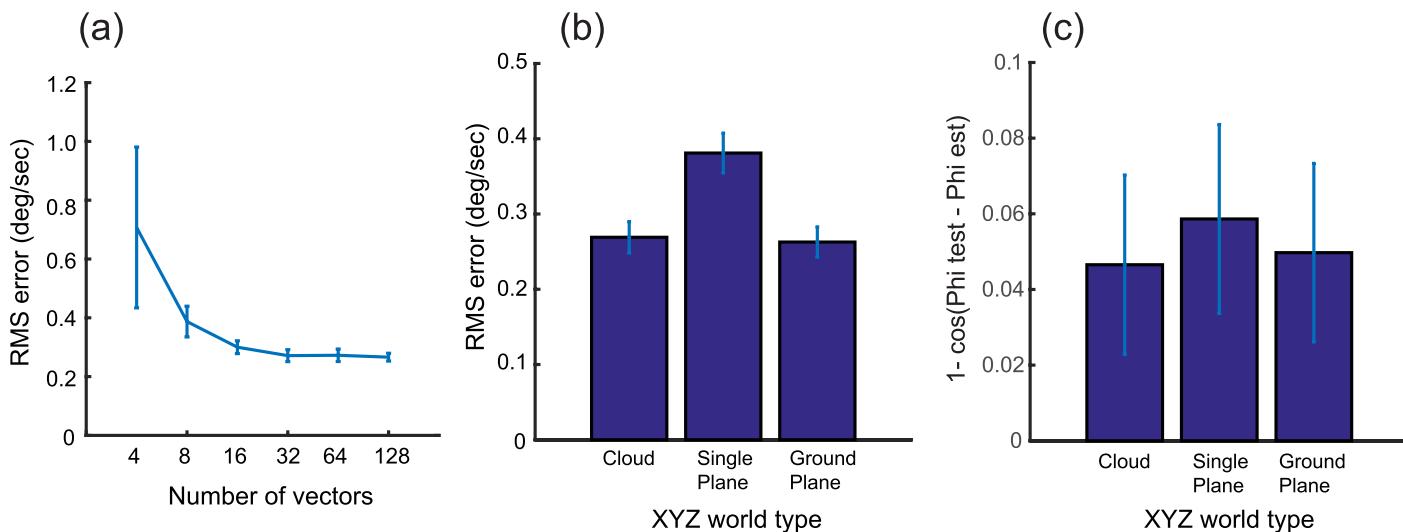


Figure 15. Tests demonstrating robustness of model. (a) Effect of number of vectors in the input flow field upon the rate root-mean-square values. Error bars represent the standard deviation across 12 simulations. (b) Effect of world type on rate root-mean-square. (c) Effect of world type on direction root-mean-square.

has been found in a number of human psychophysical heading studies (W. H. Warren, 2003). Even in the worst-case wall condition, the performance is still very good, though, and such low levels of error would support accurate heading estimation. The simulation shows that the model is relatively robust to the type of environment in which the curvilinear rotation is occurring and does not break down for cases such as the single vertical plane, which can cause problems for rotation-extraction techniques based on vector differencing (Rieger & Lawton, 1985; Royden, 1997).

#### Model free-parameter tests

There are three free parameters in the curvilinear model: the assumed spread of the vestibular distribution in the rate direction,  $\sigma_R$ ; the spread in the angle direction,  $\sigma_d$ ; and the threshold used for estimating the centroid,  $R_{\text{thresh}}$ . For all of the tests, these were set to 1.0°/s, 30°, and 55% of maximum, respectively. The value of 1.0°/s for  $\sigma_R$  is based on human psychophysical data from MacNeilage et al. (2010). The other values were set to optimize performance of the model. The free parameters were decreased and increased by a certain amount to measure their impact on the model output. The same test used in Experiment 4 was used and the percentage change in the RMS value (as a result of the parameter size variation) was measured.

To assess the overall drop in performance of the model, the RMS values for rate and direction were combined by adding them to produce an overall model performance measure  $\text{RMS}_{\text{rd}}$ . For  $\sigma_R$  and  $\sigma_d$  the value was scaled by 0.5 and 1.5 (50% and 150% change) and the percentage change in performance was assessed using the mean of the 50%  $\text{RMS}_{\text{rd}}$  change and the 150%

$\text{RMS}_{\text{rd}}$  change. For  $\sigma_R$  the change was only 8.8%, despite the large change in this free parameter's value. For  $\sigma_d$  it was 11.8%, also modest given the range of  $\sigma_d$ . Given that the threshold parameter  $R_{\text{thresh}}$  cannot exceed 1.0, the extended range tested was 75% and 125% (relative to 0.55), and this resulted in a mean change of just 8.5%. The model free parameters are therefore very tolerant to shifts in their range, and the model performance values reported in this article are not contingent on the exact values chosen.

#### Pincushion distortion and the planar image-projection assumption

As discussed earlier under Model design, the model equations and mechanism assume that the vector component of the image motion created by the curvilinear rotation is identical across the image. This is true for the curved human retina but not exactly so when a planar image projection is used to test the model; a certain amount of pincushion distortion is introduced depending on the field of view of the optic flow field being analyzed. In order to provide guidance for researchers adopting the curvilinear model to analyze the flow fields from planar imaging systems, I tested the model using the same conditions as in Experiment 4 but with pincushion compensation either on or off. Across all of the 100 randomly selected test conditions, the  $R_{\text{est}}$  and  $\phi_{\text{est}}$  values with no pincushion compensation were compared to those with compensation and a repeated-measures  $t$  test carried out to see if the performance without compensation was significantly worse than with compensation. Performance was assessed via an error measure ( $|R - R_{\text{est}}|$  and  $|\phi - \phi_{\text{est}}|$ ) for both  $R$  and  $\phi$  separately, because the pincushion

distortion introduces different types of perturbation to the vectors. A range of field-of-view values were tested:  $60^\circ \times 60^\circ$ ,  $90^\circ \times 90^\circ$ ,  $120^\circ \times 120^\circ$ , and  $130^\circ \times 130^\circ$ .

For the standard model (with parameter values set as already named) there was no significant difference in the performance of the model across all of the field-of-view values tested. The model turns out to be extremely robust to the types of flow-field distortion introduced by the use of a planar-projection imaging surface. At extremely wide angles, the pincushion distortion results in the pure rotation vectors in the periphery of the field having a higher magnitude than the true value  $R$  at the center of the field. However, the high threshold  $R_{\text{thresh}}$  used in the model centroid mechanism means that these larger, but minority, values are excluded from the  $R_{\text{est}}$  centroid calculation and have minimal impact on the model outputs. Similarly, the perturbation of the vector angles caused by the pincushion distortion tends to be symmetrical, and the vector-summation mechanism used to derive  $\phi_{\text{est}}$  (Figure 11b) is barely affected.

If other schemes or parameter values are adopted to determine the  $R_{\text{est}}$  and  $\phi_{\text{est}}$  values from the activity distribution (Figure 11), then care should be taken to allow for the pincushion distortion at wide field-of-view angles. For example, when the value of  $R_{\text{thresh}}$  was dropped from 0.55 to 0.25, statistically significant drops in performance were noted when the field of view reached  $90^\circ \times 90^\circ$ . For  $R_{\text{est}}$ ,  $t(99) = -5.39$ ,  $p < 0.001$ . The mean  $R_{\text{est}}$  error was still very small ( $0.27^\circ/\text{s}$ ), but it does indicate that if maximum model performance is the goal, then the pincushion distortion introduced through the use of a planar projection surface needs to be taken into consideration if alternative model parameters are adopted.

## Experiment 5

### Model evidence from human heading psychophysical experiments

Once the rotation rate and direction are established, the heading direction can be found using the compensation mechanism outlined earlier (Figure 3). Thus the true instantaneous heading direction is recovered as part of the output of the model along with the rate  $R$  and direction of rotation  $\phi$  that the simulated observer is experiencing. Therefore, the model can be used to simulate the results from a number of early experiments on heading estimation where rotation was added to the translation of the observer.

There is a relatively small data set available from human or nonhuman primate psychophysical experiments on curvilinear path estimation in the presence of

both visual and vestibular stimuli. As mentioned previously, the specialized equipment that is required prevents many researchers from undertaking this form of research. There are a series of studies that have used mobile robots in conjunction with virtual-reality displays that have provided valuable data indicating a number of intriguing visual–vestibular interactions during motion along curved paths (Bertin et al., 2000; Bertin & Berthoz, 2004; Bertin & Israel, 2005; Ivanenko et al., 1997). However, these studies looked at trajectories over relatively long time courses (approximately 8 s), and their observers reproduced their trajectories and angular-velocity percepts by sketching on a tablet. It is therefore difficult to relate their results to the model output, which originates from a very brief (66 ms) instance of visual motion and represents an instantaneous percept of heading and curvilinear rotation.

There are also a number of human curvilinear path studies using static observers that could be replicated using the model (e.g., J. C. Cheng & Li, 2012; Li & Cheng, 2011), but these studies have used textured scenes as stimuli. To do justice to these studies, a velocity-detecting front end needs to be used that can work with images (e.g., Perrone, 2012) rather than theoretical flow vectors. For now, I will limit myself to studies that used random-dot flow stimuli and that provide data bearing directly on the inner workings of the new model.

Except for a small set of studies mentioned earlier (Bertin et al., 2000; Bertin & Berthoz, 2004; Bertin & Israel, 2005; Ivanenko et al., 1997; Nooij et al., 2016; Telford et al., 1995), the majority of previous research on human heading perception in the presence of rotation has involved visual–vestibular conflict; the vestibular signal carries information that the observer is stationary and not undergoing rotation, while the visual information indicates both forward motion and a rotation component. However, the new curvilinear self-motion estimation model enables us to make predictions as to what should occur in these circumstances.

Many experiments have been carried out to determine the impact of added rotation upon heading-estimation performance, and these typically involve added rotation while the observer is on a simulated straight-path heading. However, over short time intervals the flow field is identical to that occurring during motion along a curved path, and participants often report that they are experiencing curvilinear self-motion (Banks et al., 1996; Bertin & Berthoz, 2004; Li & Cheng, 2011; W. H. Warren, 2003; W. H. Warren et al., 1991). One of the more detailed analyses of heading estimation in the presence of rotation was carried out by Banks et al. (1996), who systematically examined the

heading errors that occurred as more and more rotation was added.

## Methods

I replicated the flow-field stimuli used by Banks et al. (1996) as closely as possible and used a random cloud of points extending from 2 to 14 m and an ego-speed of 1.65 m/s. The field of view was  $30^\circ \times 30^\circ$  with approximately 64 dots visible. The heading was always  $(0^\circ, 0^\circ)$ , and following Banks et al., the simulated rotation was  $0^\circ/\text{s}$ ,  $0.6^\circ/\text{s}$ ,  $1.25^\circ/\text{s}$ ,  $5^\circ/\text{s}$ , and  $7.5^\circ/\text{s}$ . The rotation was always around a vertical axis such that true  $\phi = 180^\circ$ . Banks et al. also used different proportions of simulated and actual rotation rates, but here I report on the data for cases in which the simulated rotation proportion was 1.0 (i.e., there were no actual eye movements made during the trial).

The curvilinear self-motion estimation model was first run over the vector field to obtain estimates of the rate and direction of the rotation ( $R_{\text{est}}$ ,  $\phi_{\text{est}}$ ). The flow field was also passed through the heading template model (Figure 1) with templates tuned to  $-80^\circ$  to  $80^\circ$  in  $5^\circ$  steps for both azimuth and elevation. The activity from the estimated rotation was subtracted from the template activity distribution (Perrone & Krauzlis, 2008).

The model's heading estimate was found from the resulting distribution by setting a high threshold ( $0.95 \times \text{peak}$ ), converting each direction- and heading-detector activity into a 3-D vector and finding the direction of the resultant vector (Georgopoulos et al., 1986). This method for estimating the heading from the heading map's activity distribution map replaces the winner-take-all or maximum rule that was used in the original template-model articles (Perrone, 1992; Perrone & Stone, 1994). It follows the centroid method used to estimate the rotation rate in the curvilinear model (Figure 11), with the same advantage in that it enables interpolation and allows sparser sampling of the heading-detector array. The threshold value is not critical and was fixed at 95% of the peak for all of the simulations in this article that involved heading estimation.

Only the azimuth value was used (the elevation estimates were close to zero) to find the heading error. A total of 100 trials were run, with a new vector flow field generated each time. Simulations were performed for the case in which no rotation compensation was applied (no-compensation condition), when no vestibular contribution was included in the model (vision-only condition), and when vestibular signals were included but signaled that no rotation of the observer was occurring (vision-plus-static-vestibular condition). This matches the situation for the Banks et al. (1996)

observers, who remained with their heads fixed relative to the screen.

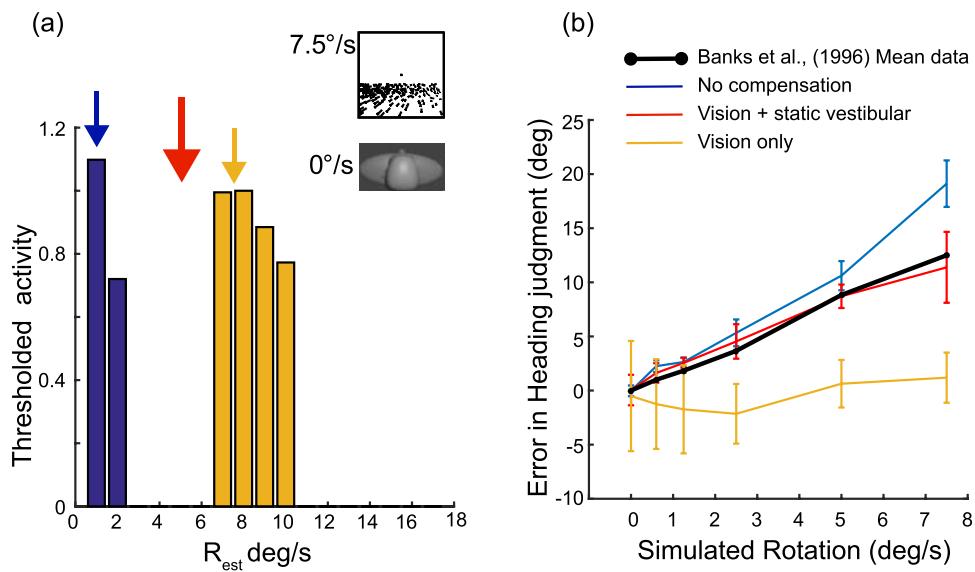
## Results

Figure 16a shows the distribution of the curvilinear model-stage  $R_{\text{est}}$  values for the  $\phi = 180^\circ$  direction from a typical trial in which the actual rotation rate was  $7.5^\circ/\text{s}$ . The histogram values on the left of the plot represent the model's vestibular-signal distribution (a Gaussian with  $\sigma_r = 1.0^\circ/\text{s}$ ). If the human observers in the Banks et al. (1996) study had not taken into account any vestibular information, their visual estimate would be around  $7.1^\circ/\text{s}$  for this trial (yellow arrow in Figure 16a). This is a slight underestimate, and if this degree of rotation compensation were applied to the heading model stage, the heading estimate would be close to  $0^\circ$  azimuth. The model simulation data (means and standard deviations as error bars) for the vision-only condition are shown as the yellow curve in Figure 16b.

If the Banks et al. observers' perceptual systems took into account the vestibular information signaling  $0^\circ/\text{s}$  rotation (blue arrow) and used some sort of winner-take-all or maximum rule to determine the rotation rate, the rate output of the curvilinear rotation stage would be  $0^\circ/\text{s}$  and no heading compensation would be expected to occur. The model heading-error predictions for this situation are shown as the blue curve in Figure 16b. Note that this is also the prediction for a system that does not include any vestibular signals. The prediction is for large heading errors in the direction of the rotation that increase as the added rotation rate increases (blue curve in Figure 16b).

If (as the model predicts) the human heading perceptual system takes into account the vestibular signals indicating  $0^\circ/\text{s}$  rotation as well as the visual rotation signal, combines them additively, and then uses a centroid-estimation stage to determine the rotation rate (red arrow in Figure 16a), the model predicts errors that fall along the red curve in Figure 16b. When average data from the three observers in the Banks et al. experiment 1 are added to this graph, they fall almost directly over the model (vision-plus-vestibular) prediction curve. This is strong evidence for the vestibular combination rule used in the model as well as the centroid stage. A winner-take-all readout of the rotation rate predicts the blue or yellow curves depending on whether or not a vestibular signal is assumed to be present or not.

In order to get the very close fit to the human data, the threshold used in the curvilinear rotation-detection stage (set to  $0.55 \times \text{peak}$  in all the previous simulations) was increased to  $0.65 \times \text{peak}$ . The same overall trends result when the threshold is 0.55, but the 0.65 value for this parameter produced the almost-perfect superim-

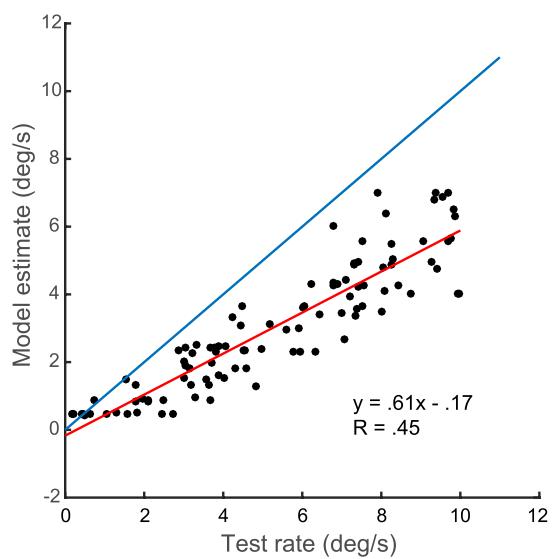


**Figure 16.** Static observers viewing flow fields with rotation: Comparison of model predictions and Banks et al. (1996) data. (a) Single-trial output of curvilinear rotation-rate estimates for case when actual rotation rate  $R = 7.5^\circ/\text{s}$  but static observer vestibular system indicates  $0^\circ/\text{s}$  (inset). The activity shown for the different candidate  $R$  values is shown as vertical blue bars for the vestibular part of the signal and yellow bars for the visual part of the signal. All values below the threshold ( $0.55 \times \text{peak}$ ) have been set to 0. This is for the case where  $\phi = 180^\circ$  (the actual rotation direction). The distribution of values peaking at  $0^\circ/\text{s}$  (blue) represents the vestibular signal present when the observer is not rotating. The centroid estimate for  $R$  with only vision is shown as the yellow arrow. For vestibular only or a winner-take-all (maximum) output scheme, the output is  $0^\circ/\text{s}$ . The large red arrow at  $5.0^\circ/\text{s}$  represents the centroid estimate when both vision and vestibular signals are used. (b) Predictions of the model over a range of simulated rotation rates. Error bars represent 1 standard deviation from 100 simulations. Black curve points are data from Banks et al. (1996), averaged from three participants in their experiment 1.

position of the model estimates with the Banks et al. mean data. This indicates that specifically designed heading experiments could be used to narrow in on the best values for these model parameters and to unearth the shape of the underlying visual–vestibular curvilinear-activity distribution (Figure 16a).

I would argue that the heading errors that occur in combined translation–rotation experiments for the condition without actual eye movements (Banks et al., 1996; Li & Cheng, 2011; W. H. Warren et al., 1991) or head movements (Crowell et al., 1998) are not as bad as one would expect from the complete absence of a compensatory mechanism (e.g., blue line in Figure 16b). These data have typically been regarded as vision-only results, but I suggest that they actually represent the situation where the perceptual system is estimating a curvilinear rotation signal on the basis of visual information and a vestibular signal indicating zero rotation. This leads to an estimate of the rotation rate that is less than the simulated rotation, and hence the compensation is incomplete.

In order to show that the rotation underestimation is not unique to the stimulus conditions used in the Banks et al. study, I tested the model over a greater range of rotation rates and multiple heading directions. Figure 17 shows simulations of the curvilinear rotation model for a range of simulated test rotation rates and the same



**Figure 17.** Model rotation-rate underestimation when the observer is static and the vestibular system signals  $0^\circ/\text{s}$  rotation. The input test rates (x-axis) were randomly selected from a uniform distribution extending from  $0^\circ/\text{s}$  to  $10^\circ/\text{s}$ . The red line is the fitted linear regression line, and the slope is less than 1.0, indicating rotation-rate underestimation.

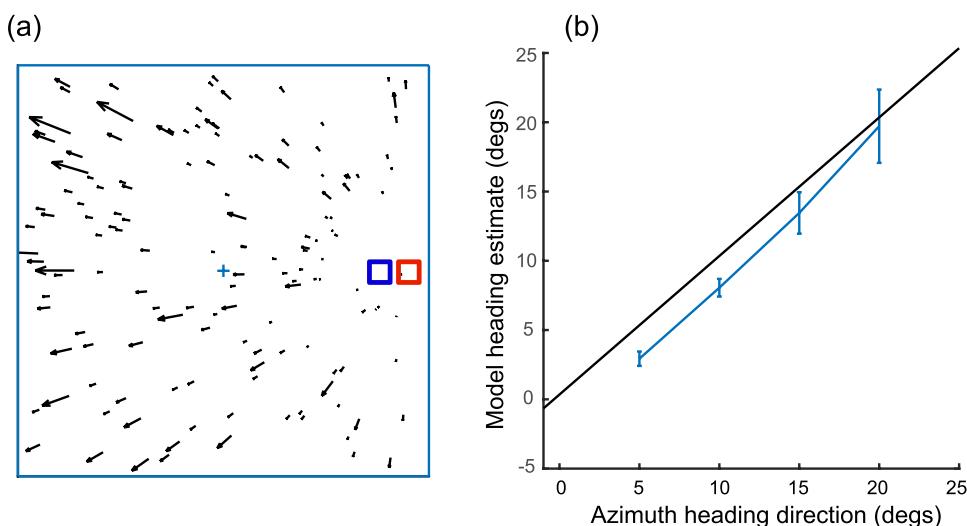


Figure 18. Model simulation of the center-screen-bias effect. (a) Actual heading is in the direction of the red square ( $15^\circ$ ) to the right. Observers typically judge their heading to be closer to the center of the display (blue square), by approximately  $2^\circ$  (W. H. Warren, 2003). (b) Model simulation data exhibiting the center-screen-bias effect. Error bars are the standard deviations from 100 simulation trials for each test heading direction.

stimulus conditions used in the Figure 16 plots. The vestibular-signal distribution was centered on  $0^\circ/\text{s}$  to represent the static observer. This plot is for 100 trials, with the rotation rate randomly drawn from a uniform distribution extending from  $0^\circ/\text{s}$  to  $10^\circ/\text{s}$ . The fitted regression line has a slope of 0.61, indicating a significant amount of rotation-rate underestimation. Note, however, that the rotation estimate is not  $0^\circ/\text{s}$  as is assumed in the human psychophysical experiments when no actual eye or head rotation occurs. The model shows that rotation information is available visually in these simulated-rotation conditions but the rate is reduced because of the fact that observers are not moving their heads. This reduced amount of compensation can explain the location of the simulated-rotation line in many human psychophysical heading data sets (e.g., Banks et al., 1996), as well as why curved paths are often seen to be less curved than they actually are (e.g., Bertin & Berthoz, 2004; Li & Cheng, 2011).

## Experiment 6

### The center-screen-bias effect

Further evidence for the mechanisms of the model and the centroid-estimation stage comes from a common phenomenon often noted in heading-estimation experiments referred to as the center-screen bias (W. H. Warren, 2003). When observers are asked to estimate their heading direction for the case of pure translation flow fields that have an eccentric heading direction (see red square in Figure 18a), they tend to

judge the heading to be biased a few degrees toward the center of the screen (blue square, Figure 18a). This effect was noted in some of the earliest experiments on heading perception (Johnston, White, & Cumming, 1973; Llewellyn, 1971) and can be seen in many of the studies that followed (Banks et al., 1996; Li & Cheng, 2011). The reason for the bias has yet to be explained adequately (W. H. Warren, 2003).

In terms of the new curvilinear rotation-detection model, a static observer viewing a flow field on a screen generates a vestibular-signal distribution centered on  $0^\circ/\text{s}$  as well as a small signal created by the lamellar flow that is present in flow fields with an eccentric FOE location. Some of the “uniform” motion at the center of the displays has almost the same direction; this is detected by the curvilinear rotation detectors and generates a signal indicating a small amount of rotation depending on the structure of the flow field and the heading direction. Combined with the vestibular distribution at  $0^\circ/\text{s}$  that has a spread that extends into nonzero regions (see blue histogram to the left of Figure 16a graph), the estimate of the rotation rate occurring for these pure translation flow fields can end up being nonzero. If this were the case, the rotation would be subtracted from the heading map (see Figure 2) to produce a percept of heading that is closer to the center of the screen than the actual position.

### Method

In order to test this hypothesis and explanation for the center-screen-bias effect, I ran the curvilinear rotation model and heading model on the type of

stimuli used in experiments that have typically displayed the bias. The stimulus was the same as that used in the Banks et al. (1996) simulation (Figure 16), except the ego-speed was set at 1.5 m/s and there was no rotation added to the forward translation. Heading directions of 5°, 10°, 15°, and 20° were tested (all in the azimuth direction only).

## Results

The model heading estimates are shown in Figure 18b. They were all below the true values, which indicates the presence of a center-screen bias; the estimates are all closer to the center of the screen than the actual values. The mean estimates ( $SDs$ ) for the 5°, 10°, 15°, and 20° headings were 2.9° (0.52°), 8.1° (0.64°), 13.5° (1.5°), and 19.7° (2.7°), respectively. The errors were therefore 2.1° (0.52°), 1.9° (0.64°), 1.5° (1.5°), and 0.3° (2.6°), with a mean of 1.5°—which is close to the “few degrees” reported for the size of the center-screen bias (W. H. Warren, 2003).

The bias in the model can be traced directly to the “erroneous” detection of curvilinear rotation in the pure translation flow fields. For the four different heading directions, the values for  $R_{est}$  were 0.44°/s, 0.5°/s, 0.77°/s, and 1.1°/s. This rotation signal mainly originates from the vestibular distribution in the model. If the system used a maximum or winner-take-all rule for establishing the rotation rate, then the estimate for  $R_{est}$  would be 0°/s in all heading cases and no bias would have occurred. If, as I suggest, the rotation estimate is derived from a centroid mechanism, then the spread of the vestibular distribution ( $\sigma_r = 1.0^\circ/s$ ) means that the centroid is shifted to nonzero values. The size of the shift depends on the spread of the vestibular distribution and the threshold level used to truncate the values used in the centroid-estimation mechanism (0.55 × peak used in model). A more systematic testing of the heading center-screen bias could be used to estimate what values for  $\sigma_r$  and the threshold are appropriate for human observers. The model also predicts that the bias should increase when displays with a larger field of view are used to test heading, because the inclusion of faster moving vectors in the periphery of the display leads to higher values of  $R_{est}$ .

This hypothesis for the origin of the center-screen bias is based on the idea that curvilinear rotation is being detected in cases where no actual rotation is occurring. There is another possible explanation for the bias that does not rely on the detection of rotation and stems instead from the heading template model itself. For eccentric heading directions, the distribution of activity across the heading detectors (see Figure 2c) is skewed. The centroid-estimation mechanism used to determine heading from the distribution can lead to a

shift away from the peak and could also produce a center-screen bias. Evidence for biases in both visual and vestibular heading has been reported, and the cause attributed to skewed distributions and population-vector decoding (Cuturi & MacNeilage, 2013). If the bias is arising from the mechanism described earlier (Figure 18), anything that removes the compensation signal should eliminate the heading bias—but it should not affect it if the bias is arising from the centroid stage. Further experiments are required to determine which of these two explanations is the most viable.

## Discussion

I have presented a mechanism that could be used by the human brain to detect and measure the rotation that occurs as we move along curved paths or trajectories. Without this information, the visual system is unable to recover the heading direction from the visual flow field using some form of compensation mechanism (e.g., Figure 2) and hence determine the depth of points in the world. I demonstrated that in order to achieve the heading-discrimination abilities found from human psychophysical experiments on curvilinear self-motion (W. H. Warren et al., 1991), the rotation-rate precision needs to be at least 0.2°/s. This is far lower than the equivalent type of threshold (1.0°/s) reported for purely vestibular estimation of rotation rate (MacNeilage, Turner, & Angelaki, 2010). I developed a new system whereby information in the visual flow field could be used to estimate the direction and rate of rotation occurring during curvilinear self-motion. However, it too failed to achieve the 0.2°/s requirement, and simulations showed that a threshold of around 1.0°/s is the most one could expect from a system that uses only this vision-based technique. I then showed that by combining the relatively imprecise vestibular signal with the visually derived signal, it is possible to achieve curvilinear rotation estimates with a precision that supports heading thresholds of around 1°–1.5°.

The visual solution I developed for extracting the rotation from a combined translation-plus-rotation flow field relies on the fact that rotation introduces a common component that occurs at every vector in the field, and this can be detected by summing the magnitudes of the vectors weighted by an amount dependent on the angle of the vector and its location (Equation 2). This mechanism works for motion along any curved trajectory in 3-D space and while the observer is looking in any direction relative to the heading vector—that is, over a range of FOE locations on the retinocentric image plane. This process is not foolproof, however, and it can generate spurious

solutions, especially in the case where there are a small number of vectors available. I suggest that a rotation signal (rate and direction) from the vestibular system acts as a constraint on some of these erroneous solutions and only reinforces those solutions that fall within the vestibular system's relatively broad precision boundaries.

It is also possible to impose further constraints on the rotation solution by incorporating a vestibular signal from the otoliths indicating the approximate direction of heading. This would eliminate some of the incorrect solutions for the values of the rotation rate and direction generated by the visual estimation process (see the multiple curves in Figure 8a created from the inclusion of many FOE locations). There is also some existing psychophysical evidence pointing to changes in curvilinear path estimation when a brief pure forward-translation vestibular signal occurs (Bertin & Berthoz, 2004), and electrophysiological evidence that curvilinear-sensitive neurons in the primate brain respond differently when forward motion occurs concurrently with the rotation (Z. Cheng & Gu, 2016). I have tested the model with and without this additional heading constraint and found that it does not have a large effect on the performance characteristics for the simulations I have reported in this article. Therefore, to keep the model as simple as possible, I have left this stage out. However, it is acknowledged that a forward-motion signal may be useful in the future when a greater range of headings is considered and when information about one's forward speed is required (currently not available from the visual flow).

The new model highlights the impact that visual-vestibular conflict can have on heading-estimation performance. Early heading experiments tended to use a static observer viewing flow fields on a screen. When rotation was added to the simulated forward translation of observers without them actually moving their eyes, large heading errors resulted, and often the participants reported that they were moving along curvilinear paths (Banks et al., 1996; W. H. Warren, 2003). My model simulations show that this could be the result of a correct visual estimate of the rotation being "dragged down" by a vestibular signal indicating 0°/s of rotation. There is some compensation for the rotation occurring, but it is not sufficient to eliminate all of the heading error. I would argue that if we were not able to recover some rotation from these visual-vestibular conflict situations, large heading errors would occur, and viewing movies or playing first-person video games would produce larger distortions in self-motion perception than are currently reported. It is possible of course that we learn to ignore the incorrect vestibular signal after sufficient exposure to the cue-conflict situation.

Most experiments on heading that introduce rotation have used a  $\phi$  value set to 0° or 180°. We do not currently know if heading performance will be the same if the plane of rotation is not 0° or 180°, or what the vestibular threshold is for detecting a trajectory that is slightly different from 0°. I used a value for the spread ( $\sigma_d$ ) of the vestibular distribution (Figure 10c) equal to 30° because this produced the best performance in the model. In the model I assumed that the visual and vestibular distributions align along the rotation-direction axis  $\phi$  (Figure 10b and 10c), but how far can the vestibular estimate of  $\phi$  be from the correct value and the visual estimate before the heading errors get too large?

There are a number of experiments that now must be carried out in order to refine aspects of the model and gain support for its basic underlying mechanisms. For example, there are currently no data indicating the range of curved-path angles over which humans can accurately judge their heading direction. This is the  $\phi$  angle in Equation 2, and I arbitrarily sampled it using a range of angles from 0° to 330° in 30° steps. Can humans judge their trajectory along a curved path when the plane of the trajectory is inclined at, say, 30° to the horizontal? For terrestrial animals,  $\phi$  is most commonly experienced at 0° and 180°. Humans experience  $\phi$  values around 90° and 270° when moving from flat ground to an upward or downward slope. Values of  $\phi$  away from the cardinal directions are experienced when driving around an upward- or downward-sloping bend. Presumably we can recover the rotation correctly, and heading estimation is not compromised in these cases, but the experimental data addressing this are currently lacking. Tree-dwelling primates that swing from branches experience a much greater range of curvilinear paths (and a greater  $\phi$  range), as do pilots; is their precision for detecting changes in  $\phi$  greater than that of humans who experience curvilinear paths only in the vertical or horizontal planes?

As mentioned previously, the  $\phi_{\text{est}}$  value that is outputted from the curvilinear model is in a retinal coordinate frame, and it is not clear how this could ultimately be converted to the angle of the curved path in the world—that is, the yaw and pitch rates of the eye-head-body relative to the world ( $\omega_Y, \omega_P$ ). The transformation of vestibular signals from head-centric to world coordinates requires the incorporation of signals from the otoliths to disambiguate alternative solutions as to the rotation direction relative to gravity (Angelaki & Cullen, 2008). Therefore, one would expect differences in the ability to determine ( $\omega_Y, \omega_P$ ) from  $\phi_{\text{est}}$  to depend on whether or not the observer is moving forward during the rotation. This test, along with the determination of the true value for  $\sigma_d$  and measuring how accurate the vestibular  $\phi$  estimate needs

to be, is a relatively simple experiment for anybody with a moving-base simulator.

I have presented some evidence for my centroid-based mechanism used for determining the rotation rate, but this was based on past experiments that were designed to measure heading performance. A custom-designed experiment that manipulated the location of the vestibular distribution in my putative linear visual-vestibular summation mechanism (Figure 10d) by adjusting the amount of vestibular rotation would go a long way to supporting the basic premise in the model. The visual flow field could indicate motion along a curved path with, say,  $4^{\circ}/s$  of rotation while the actual rotation of the observer is titrated from  $0^{\circ}/s$  (or even negative values) to values beyond  $4^{\circ}/s$ . The model makes specific predictions regarding the rotation rate and heading direction that will be perceived in each case. The data would soon reveal if the human brain is using a centroid mechanism or some other means of combining the visual and vestibular signals.

The model is currently not explicit about where the particular mechanisms occur in the visual system. I have highlighted possible mechanisms for extracting curvilinear rotation information from both visual and vestibular sources, but there are multiple ways of implementing the basic concept once the visual source of information has been identified (Equation 2). The electrophysiological evidence so far indicates that the process may well be distributed over a number of cortical areas (Chen et al., 2011; Z. Cheng & Gu, 2016; Gu et al., 2006; Takahashi et al., 2007). The model does constrain the order in which the different operations need to occur, though. For example, the extraretinal cancellation of rotation (Perrone & Krauzlis, 2008) arising from eye-in-head and head-relative-to-body movements needs to occur prior to the curvilinear rotation-detection stage. The values of  $V_{T+R}$  and  $\theta$  in Equation 2 need to be free of any other rotation components for the model to work. Other than that, there are many options for how the curvilinear detectors (Figure 10) could connect to the motion sensors extracting image velocity from the retinal flow.

I consider the problem of curvilinear self-motion estimation as being the last obstacle in the quest to extract depth from a velocity-vector flow field. We have already shown that the impact of rotation arising from eye or head movements relative to the head or body can be removed (Perrone & Krauzlis, 2008), and now I have shown how rotation arising from curvilinear trajectories can be measured and removed. Hence, the pure translation field can be recovered from combined translation-and-rotation flow fields. Once we have the pure translation vector field, it is relatively easy to recover heading and consequently the relative depth of the points in the world. Knowing the heading direction and the vector magnitude and direction enables depth

to be recovered using basic trigonometry (Koenderink & van Doorn, 1975; Longuet-Higgins & Prazdny, 1980). If one also knows the forward speed of the observer, it is possible to obtain the absolute depth and location in the world of the point generating the vector; otherwise, the relative depth of the dot positions can be derived. This is all within a very brief time span (dependent on the speed of processing of the 2-D motion sensors). Over longer periods of time, it is possible to track the locations of the recovered 3-D point locations to gain further information regarding one's motion through the world and whether or not a collision is possible.

Therefore, the new curvilinear self-motion estimation model opens up many possibilities regarding the recovery of information from brief monocular video sequences. I have already proposed a system based on the properties of neurons in the primate visual system for extracting 2-D image motion from video sequences (Perrone, 2012), and this will now be used as a front end to the model I have developed here. Besides opening up a greater range of psychophysical data to test with the model (e.g., J. C. Cheng & Li, 2012), such an image-based system would have many applications in robotics and autonomous vehicles. These currently rely on multiple active sensors to learn about the environment in front of the vehicle. Humans are able to do this using just the motion information from a single eye. The model I have presented in this article is one suggestion for how the brain achieves this amazing ability.

**Keywords:** heading estimation, path estimation, visual navigation, MSTd, vector flow fields, rotation estimation, visual–vestibular interactions, robotics, autonomous vehicles

## Acknowledgments

Thanks to Aden Garnett, Dorion Liston, Frans Verstraten, and Hamish MacDougall for feedback and helpful discussions regarding the model.

Commercial relationships: none.

Corresponding author: John A. Perrone.

Email: jpnz@waikato.ac.nz.

Address: School of Psychology, University of Waikato, Hamilton, New Zealand.

## References

- Angelaki, D. E., & Cullen, K. E. (2008). Vestibular system: The many facets of a multimodal sense.

- Annual Review of Neuroscience*, 31, 125–150, <https://doi.org/10.1146/annurev.neuro.31.060407.125555>.
- Banks, M. S., Ehrlich, S. M., Backus, B. T., & Crowell, J. A. (1996). Estimating heading during real and simulated eye movements. *Vision Research*, 36(3), 431–443.
- Beintema, J. A., & van den Berg, A. V. (1998). Heading detection using motion templates and eye velocity gain fields. *Vision Research*, 38(14), 2155–2179.
- Beintema, J. A., & van den Berg, A. V. (2001). Pursuit affects precision of perceived heading for small viewing apertures. *Vision Research*, 41, 2375–2391.
- Bertin, R. J., & Berthoz, A. (2004). Visuo-vestibular interaction in the reconstruction of travelled trajectories. *Experimental Brain Research*, 154(1), 11–21, <https://doi.org/10.1007/s00221-003-1524-3>.
- Bertin, R. J., & Israel, I. (2005). Optic-flow-based perception of two-dimensional trajectories and the effects of a single landmark. *Perception*, 34(4), 453–475, <https://doi.org/10.1088/p5292>.
- Bertin, R. J., Israel, I., & Lappe, M. (2000). Perception of two-dimensional, simulated ego-motion trajectories from optic flow. *Vision Research*, 40(21), 2951–2971.
- Bouguet, J. Y. (2015). *Camera Calibration Toolbox for Matlab*. Retrieved from [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/)
- Bradley, D. C., Maxwell, M., Andersen, R. A., Banks, M. S., & Shenoy, K. V. (1996, September 13). Mechanisms of heading perception in primate visual cortex. *Science*, 273(5281), 1544–1547.
- Britten, K. H. (2008). Mechanisms of self-motion perception. *Annual Review of Neuroscience*, 31, 389–410, <https://doi.org/10.1146/annurev.neuro.29.051605.112953>.
- Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, 12(12), 4745–4765.
- Britten, K. H., & van Wezel, R. J. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nature Neuroscience*, 1(1), 59–63.
- Butler, J. S., Smith, S. T., Campos, J. L., & Bulthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision*, 10(11):23, 1–13, <https://doi.org/10.1167/10.11.23>. [PubMed] [Article]
- Chen, A., DeAngelis, G. C., & Angelaki, D. E. (2011). Convergence of vestibular and visual self-motion signals in an area of the posterior sylvian fissure. *The Journal of Neuroscience*, 31(32), 11617–11627, <https://doi.org/10.1523/JNEUROSCI.1266-11>. 2011.
- Cheng, J. C., & Li, L. (2012). Effects of reference objects and extra-retinal information about pursuit eye movements on curvilinear path perception from retinal flow. *Journal of Vision*, 12(3):12, 1–21, <https://doi.org/10.1167/12.3.12>. [PubMed] [Article]
- Cheng, Z., & Gu, Y. (2016). Distributed representation of curvilinear self-motion in the macaque parietal cortex. *Cell Reports*, 15(5), 1013–1023, <https://doi.org/10.1016/j.celrep.2016.03.089>.
- Crane, B. T. (2014). Human visual and vestibular heading perception in the vertical planes. *Journal of the Association for Research in Otolaryngology*, 15(1), 87–102, <https://doi.org/10.1007/s10162-013-0423-y>.
- Crowell, J. A., Banks, M. S., Shenoy, K. V., & Andersen, R. A. (1998). Visual self-motion perception during head turns. *Nature Neuroscience*, 1, 732–737.
- Cutting, J. E. (1986). *Perception with an eye for motion*. Cambridge, MA: MIT Press.
- Cuturi, L. F., & MacNeilage, P. R. (2013). Systematic biases in human heading estimation. *PLoS One*, 8(2), e56862, <https://doi.org/10.1371/journal.pone.0056862>.
- Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli: I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65(6), 1329–1345.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433, <https://doi.org/10.1038/415429a>.
- Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2010). Visual-vestibular cue integration for heading perception: Applications of optimal cue integration theory. *European Journal of Neuroscience*, 31(10), 1721–1729, <https://doi.org/10.1111/j.1460-9568.2010.07207.x>.
- Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *The Journal of Neuroscience*, 29(49), 15601–15612, <https://doi.org/10.1523/JNEUROSCI.2574-09.2009>.
- Freeman, T. C., Champion, R. A., & Warren, P. A. (2010). A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Current Biology*, 20(8), 757–762, <https://doi.org/10.1016/j.cub.2010.02.059>.

- Gegenfurtner, K. R., Xing, D. J., Scott, B. H., & Hawken, M. J. (2003). A comparison of pursuit eye movement and perceptual performance in speed discrimination. *Journal of Vision*, 3(11):19, 865–876, <https://doi.org/10.1167/3.11.19>. [PubMed] [Article]
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986, September 26). Neuronal population coding of movement direction. *Science*, 233(4771), 1416–1419.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston, MA: Houghton Mifflin.
- Goldberg, J. M., Wilson, V. J., Cullen, C. E., Angelaki, D. E., Boussard, D. M., & Buettner-Ennever, J. A. (2012). *The vestibular system: A sixth sense* (Vol. 1). New York, NY: Oxford University Press.
- Grabherr, L., Nicoucar, K., Mast, F. W., & Merfeld, D. M. (2008). Vestibular thresholds for yaw rotation about an earth-vertical axis as a function of frequency. *Experimental Brain Research*, 186(4), 677–681, <https://doi.org/10.1007/s00221-008-1350-8>.
- Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: R. E. Krieger.
- Gu, Y., Watkins, P. V., Angelaki, D. E., & DeAngelis, G. C. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *The Journal of Neuroscience*, 26(1), 73–85, <https://doi.org/10.1523/JNEUROSCI.2356-05.2006>.
- Guedry, F. E., & Lauver, L. S. (1961). Vestibular reactions during prolonged constant angular acceleration. *Journal of Applied Physiology*, 16(2), 215–220.
- Howard, I. P. (1982). *Human visual orientation*. Chichester, UK: John Wiley & Sons.
- Imai, T., Moore, S. T., Raphan, T., & Cohen, B. (2001). Interaction of the body, head, and eyes during walking and turning. *Experimental Brain Research*, 136(1), 1–18.
- Ivanenko, Y., Grasso, R., Israel, I., & Berthoz, A. (1997). Spatial orientation in humans: Perception of angular whole-body displacements in two-dimensional trajectories. *Experimental Brain Research*, 117(3), 419–427.
- Johnston, I. R., White, G. R., & Cumming, R. W. (1973). The role of optical expansion patterns in locomotor control. *American Journal of Psychology*, 86(2), 311–324.
- Kim, H. R., Angelaki, D. E., & DeAngelis, G. C. (2015). A novel role for visual perspective cues in the neural computation of depth. *Nature Neuroscience*, 18(1), 129–137, <https://doi.org/10.1038/nn.3889>.
- Kim, N. G., & Turvey, M. T. (1999). Eye movements and a rule for perceiving direction of heading. *Ecological Psychology*, 11(3), 233–248, [https://doi.org/10.1207/s15326969eco1103\\_3](https://doi.org/10.1207/s15326969eco1103_3).
- Koenderink, J. J., & van Doorn, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22(9), 773–791.
- Kowler, E., & McKee, S. P. (1987). Sensitivity of smooth eye movement to small differences in target velocity. *Vision Research*, 27(6), 993–1015.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35(3), 389–412.
- Lappe, M. (1998). A model of the combination of optic flow and extraretinal eye movement signals in primate extrastriate visual cortex: Neural model of self-motion from optic flow and extraretinal cues. *Neural Networks*, 11, 397–414.
- Lappe, M., Bremmer, F., & van den Berg, A. V. (1999). Perception of self-motion from visual flow. *Trends in Cognitive Sciences*, 3(9), 329–336.
- Lee, D. N. (1974). Visual information during locomotion. In R. B. Macleod & H. Pick (Eds.), *Perception: Essays in honor of J.J. Gibson* (pp. 250–267). Ithaca, NY: Cornell University Press.
- Lee, D. N., & Lishman, R. (1977). Visual control of locomotion. *Scandinavian Journal of Psychology*, 18(3), 224–230.
- Li, L., & Cheng, J. C. (2011). Perceiving path from optic flow. *Journal of Vision*, 11(1):22, 1–15, <https://doi.org/10.1167/11.1.22>. [PubMed] [Article]
- Llewellyn, K. R. (1971). Visual guidance of locomotion. *Journal of Experimental Psychology*, 91(2), 245–261.
- Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society of London B*, B208, 385–387.
- MacNeilage, P. R., Banks, M. S., DeAngelis, G. C., & Angelaki, D. E. (2010). Vestibular heading discrimination and sensitivity to linear acceleration in head and world coordinates. *The Journal of Neuroscience*, 30(27), 9084–9094, <https://doi.org/10.1523/JNEUROSCI.1304-10.2010>.
- MacNeilage, P. R., Turner, A. H., & Angelaki, D. E. (2010). Canal-otolith interactions and detection thresholds of linear and angular components

- during curved-path self-motion. *Journal of Neurophysiology*, 104(2), 765–773, <https://doi.org/10.1152/jn.01067.2009>.
- Mallery, R. M., Olomu, O. U., Uchanski, R. M., Militchin, V. A., & Hullar, T. E. (2010). Human discrimination of rotational velocities. *Experimental Brain Research*, 204(1), 11–20, <https://doi.org/10.1007/s00221-010-2288-1>.
- Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception*, 3, 63–80.
- Niehorster, D. C., Cheng, J. C., & Li, L. (2010). Optimal combination of form and motion cues in human heading perception. *Journal of Vision*, 10(11):20, 1–15, <https://doi.org/10.1167/10.11.20>. [PubMed] [Article]
- Nooij, S. A., Nesti, A., Bulthoff, H. H., & Pretto, P. (2016). Perception of rotation, path, and heading in circular trajectories. *Experimental Brain Research*, 234(8), 2323–2337, <https://doi.org/10.1007/s00221-016-4638-0>.
- Ono, H., & Wade, N. J. (2005). Depth and motion in historical descriptions of motion parallax. *Perception*, 34(10), 1263–1273, <https://doi.org/10.1088/p5232>.
- Pack, C., Grossberg, S., & Mingolla, E. (2001). A neural model of smooth pursuit control and motion perception by cortical area MST. *Journal of Cognitive Neuroscience*, 13(1), 102–120.
- Perrone, J. A. (1987). Extracting 3-D egomotion information from a 2-D flow field: A biological solution? *Optical Society of America Technical Digest Series*, 22, 47.
- Perrone, J. A. (1992). Model for the computation of self-motion in biological systems. *Journal of the Optical Society of America*, 9, 177–194.
- Perrone, J. A. (2012). A neural-based code for computing image velocity from small sets of middle temporal (MT/V5) neuron inputs. *Journal of Vision*, 12(8):1, 1–31, <https://doi.org/10.1167/12.8.1>. [PubMed] [Article]
- Perrone, J. A., & Krauzlis, R. J. (2008). Vector subtraction using visual and extraretinal motion signals: A new look at efference copy and corollary discharge theories. *Journal of Vision*, 8(14):24, 1–14, <https://doi.org/10.1167/8.14.24>. [PubMed] [Article]
- Perrone, J. A., & Stone, L. S. (1994). A model of self-motion estimation within primate extrastriate visual cortex. *Vision Research*, 34, 2917–2938.
- Perrone, J. A., & Stone, L. S. (1998). Emulating the visual receptive field properties of MST neurons with a template model of heading estimation. *The Journal of Neuroscience*, 18, 5958–5975.
- Pouget, A., Deneve, S., & Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, 3(9), 741–747, <https://doi.org/10.1038/nrn914>.
- Regan, D., & Beverley, K. I. (1982, January 8). How do we avoid confounding the direction we are looking and the direction we are moving? *Science*, 215(8), 194–196.
- Rieger, J. H. (1983). Information in optical flows induced by curved paths of observation. *Journal of the Optical Society of America*, 73(3), 339–344.
- Rieger, J. H., & Lawton, D. T. (1985). Processing differential image motion. *Journal of the Optical Society of America A: Optics and Image Science*, 2(2), 354–360.
- Rohde, M., van Dam, L. C. J., & Ernst, M. O. (2016). Statistically optimal multisensory cue integration: A practical tutorial. *Multisensory Research*, 29(4–5), 279–317, <https://doi.org/10.1163/22134808-00002510>.
- Royden, C. S. (1997). Mathematical analysis of motion-opponent mechanisms used in the determination of heading and depth. *Journal of the Optical Society of America A*, 14(9), 2128–2143, <https://doi.org/10.1364/JOSAA.14.002128>.
- 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. *The Journal of Neuroscience*, 6(1), 145–157.
- Sommer, M. A., & Wurtz, R. H. (2008). Visual perception and corollary discharge. *Perception*, 37(3), 408–418, <https://doi.org/10.1088/p5873>.
- Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative Physiology and Psychology*, 43, 482–489.
- Stone, L. S., & Perrone, J. A. (1997). Human heading estimation during visually simulated curvilinear motion. *Vision Research*, 37(5), 573–590.
- Sunkara, A., DeAngelis, G. C., & Angelaki, D. E. (2016). Joint representation of translational and rotational components of optic flow in parietal cortex. *Proceedings of the National Academy of Sciences, USA*, 113(18), 5077–5082, <https://doi.org/10.1073/pnas.1604818113>.
- Takahashi, K., Gu, Y., May, P. J., Newlands, S. D., DeAngelis, G. C., & Angelaki, D. E. (2007). Multimodal coding of three-dimensional rotation

- and translation in area MSTd: Comparison of visual and vestibular selectivity. *The Journal of Neuroscience*, 27(36), 9742–9756, <https://doi.org/10.1523/JNEUROSCI.0817-07.2007>.
- 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. *Journal of Neurophysiology*, 62(3), 642–656.
- 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. *The Journal of Neuroscience*, 6(1), 134–144.
- Telford, L., Howard, I. P., & Ohmi, M. (1995). Heading judgments during active and passive self-motion. *Experimental Brain Research*, 104(3), 502–510.
- von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip. *Naturwissenschaften*, 37, 464–476.
- Wann, J. P., & Swapp, D. K. (2000). Why you should look where you are going. *Nature Neuroscience*, 3(7), 647–648, <https://doi.org/10.1038/76602>.
- Warren, R. (1976). The perception of egomotion. *Journal of Experimental Psychology: Human Perception and Performance*, 2(3), 448–456.
- Warren, W. H. (2003). Optic flow. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (Vol. 2, pp. 1247–1259). Cambridge, MA: Bradford.
- Warren, W. H., Jr., Mestre, D. R., Blackwell, A. W., & Morris, M. W. (1991). Perception of circular heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance*, 17(1), 28–43.