PERGAMON

Vision Research 38 (1998) 3129–3145

Vision Research

Depth information and perceived self-motion during simulated gaze rotations

Sheryl M. Ehrlich, Diane M. Beck, James A. Crowell, Tom C.A. Freeman, Martin S. Banks *

Department of Psychology, School of Optometry, University of California, Berkeley, CA 94720–2020, USA

Received 27 August 1996; received in revised form 23 April 1997; accepted 21 November 1997

Abstract

When presented with random-dot displays with little depth information, observers cannot determine their direction of self-motion accurately in the presence of rotational flow without appropriate extra-retinal information (Royden CS et al. Vis Res 1994;34:3197–214.). On theoretical grounds, one might expect improved performance when depth information is added to the display (van den Berg AV and Brenner E. Nature 1994;371:700–2). We examined this possibility by having observers indicate perceived self-motion paths when the amount of depth information was varied. When stereoscopic cues and a variety of monocular depth cues were added, observers still misperceived the depicted self-motion when the rotational flow in the display was not accompanied by an appropriate extra-retinal, eye-velocity signal. Specifically, they perceived curved self-motion paths with the curvature in the direction of the simulated eye rotation. The distance to the response marker was crucial to the objective measurement of this misperception. When the marker distance was small, the observers' settings were reasonably accurate despite the misperception of the depicted self-motion. When the marker distance was large, the settings exhibited the errors reported previously by Royden CS et al. Vis Res 1994;34:3197–3214. The path judgement errors observers make during simulated gaze rotations appear to be the result of misattributing path-independent rotation to self-motion along a circular path with path-dependent rotation. An analysis of the information an observer could use to avoid such errors reveals that the addition of depth information is of little use.

Keywords: Optic flow; Heading; Navigation; Self-motion; Stereopsis; Occlusion; Perspective

1. Introduction

A moving observer can determine the direction of self-motion from the pattern of retinal image motion. When the observer is moving on a straight path and the gaze is fixed relative to the direction of self-motion, there is a point in the retinal flow field from which all texture elements flow centrifugally. This focus of expansion corresponds in such cases to the direction of self-motion [3]. The estimation of the path of self-motion is more complicated when the direction of gaze changes relative to the direction of self-motion because of an eye and/or head rotation. In this case, the self-motion direction is not given by the position of a focus of expansion because such foci are eliminated or displaced by the added rotational flow [4]. Humans can nonetheless perceive their direction of self-motion reasonably accurately during eye rotations [1,5–7].

Two general methods of recovering the self-motion path during gaze rotation have been proposed. Retinal-image models estimate direction of self-motion in the presence of rotational flow directly from the retinal velocity field. Although there are many such algorithms [8–11], they all rely on the fact that flow due to translation (that is, displacement of the observing eye) and flow due to rotation have different properties. In particular, flow due to translation is depth-dependent while flow due to rotation is not. Extra-retinal models estimate the gaze rotation directly by means of extra-retinal signals. In the case of eye rotation, the signals could be provided by proprioceptive feedback from or efferent signals to the extra-ocular muscles [12,13]. In the case of head movements, the signals could be

* Corresponding author. Fax: +1 510 6435109; e-mail: marty@john.berkeley.edu.

0042-6989/98/$19.00
PII: S0042-6989(97)00427-6

provided by feedback from the neck muscles or by the vestibular system. Once the gaze rotation is estimated from those signals, it is relatively simple to subtract the indicated rotational flow from the retinal flow field [1] and then estimate the direction of self-motion perhaps by means of one of the algorithms mentioned above.

In actuality, there are a variety of ways in which the information contained in the retinal image could be combined with extra-retinal signals from the eye and neck muscles and from the vestibular system [14]. At issue in the work presented here is whether extra-retinal signals from the eye muscles influence the self-motion path people perceive during smooth pursuit eye movements. There have been several reports on this issue in recent years and as yet no firm conclusion has emerged [1,2,5–7,15–21]. In most of these reports, humans' ability to estimate self-motion paths has been examined during real and simulated eye movements. In the real eye movement condition, a translational flow field is displayed (that derives from a simulated scene of a ground plane, frontal plane(s), or 3D cloud) and the observers fixate a target that moves relative to the head. In the simulated eye movement condition, the observers fixate a stationary target and a flow field is presented with translational and rotational flow components such that the retinal image over time is the same as in the real movement condition. If humans do not use extra-retinal, eye-velocity signals in estimating self-motion paths, they should perform similarly in the real and simulated conditions because the retinal images are identical. If they do use such signals, their judgments should be significantly less accurate in the simulated condition because the rotational flow estimated from extra-retinal signals (i.e. that no eye movement has occurred) does not match the rotational flow in the display.

Banks and colleagues [1,5,15] reported that human observers could not judge direction of self-motion accurately for simulated rotations greater than 1°/s; the reported path errors were displaced in the direction of the simulated rotation by an amount proportional to the rotation rate (see also refs. [16,20])[1]. Observers reported perceiving curved self-motion paths rather than the simulated linear path. These findings imply that extra-retinal, eye-velocity signals are used in the estimation of self-motion paths. Because observers perceived curved rather than linear paths, we will refer to the psychophysical judgements as path judgements rather than heading judgements because the latter term implies either a linear path or an attempt to judge the instantaneous tangent to a curved path.

Adding depth information to optic flow displays (other than that contained in the flow field) could plausibly aid the estimation of self-motion paths during eye and head movements. Flow velocity due to linear translation is inversely proportional to the distance of a texture element from the observer's eye. Thus, any observed flow of texture that is very distant from the observer must be due primarily to gaze rotation and knowing which texture elements are more distant would allow one to estimate the rotation (and subsequently the observer's direction of self-motion) more accurately. Conversely, flow of texture elements close to the observer is determined primarily by the observer's translation, so one could also estimate the observer's direction of self-motion reasonably accurately by monitoring the flow of texture at close range.

Following this reasoning, van den Berg and Brenner [2] examined the ability to estimate direction of self-motion during simulated rotations when the amount of depth information in the display was varied[2]. They also manipulated task difficulty by varying the signal-to-noise ratio in their displays (this was done by adding a random component to the motion of each dot in the display). They reported that stereoscopic information aided performance when the signal-to-noise ratio was low and the stimulus was a 3D cloud. Specifically, the addition of binocular disparity yielded less variable responses with no obvious change in mean response error. Van den Berg and Brenner [2] conducted a similar experiment with a ground-plane stimulus and found little effect of adding stereoscopic information in that case. All of Banks and colleagues' [1,5,15] stimuli contained no noise and therefore are most comparable to van den Berg and Brenner's [2] high signal-to-noise ratio condition. Thus, the only incompatibility between the data of van den Berg and Brenner and Banks and colleagues is the reported response biases in the high signal-to-noise ratios. Banks and co-workers observed large biases at high simulated rotation rates and two of van den Berg and Brenner's three observers did not exhibit such biases. A secondary purpose of the work reported here is to examine the cause of this discrepancy.

More importantly, the results of van den Berg and Brenner [2] raise the intriguing possibility that adding depth information to optic flow displays aids the esti-

[1] Papers by van den Berg [19] and Van den Berg and Brenner [21] reported accurate heading estimation during stimulated gaze rotation. It is difficult to compare those data to the ones reported here for reasons that will become clearer below; they concern the response maker distance, the rigid attachment of the fixation point to the depicted scene, and the manner in which the data are plotted.

[2] Vishton et al. [35] also examined the influence of added relative and changing size depth information. They reported that observers estimated heading more accurately when relative size information was consistent rather than inconsistent with the display motion. However, they did not run a condition in which there was no relative size cue, so they could not determine whether consistent relative size actually aided performance or whether performance was only degraded by inconsistent relative size.

mation of self-motion paths during simulated eye rotations. The main purpose of the current paper is to examine the perception of self-motion during simulated eye rotations when the amount of depth information is varied.

2. General methods

Several observers, all with corrected visual acuities of 20/20 or better, participated. Two of them (MSB and SME) had considerable experience as psychophysical observers and were aware of the experimental hypotheses. DMB was also aware of the experimental hypotheses, but had little experience. The others had little or no experience and were unaware of the hypotheses. The experiments were conducted with the understanding and written consent of each observer.

The observers viewed displays of randomly-placed dots whose motions depicted linear self-motion or linear self-motion plus gaze rotation with respect to three types of scenes (which are described in the context of the individual experiments). They were instructed to fixate a small cross at eye level at a depicted distance of 1000 cm; the cross was stationary on the display screen (simulated eye movement condition) or moved (real eye movement condition). In all but one experiment, the fixation cross was not attached to the otherwise rigid scene and therefore moved independently of it. In Experiment 4, the fixation cross appeared on the horizontal midline and was rigidly attached to the scene.

The stimuli were generated on a Power Macintosh 9500/132 and displayed with an Electrohome ECP-4000 projector onto a 97×78.5 cm rear-projection screen. At the 70 cm viewing distance, the screen subtended $69 \times 59°$. We used a rectangular software clipping window to limit the angular subtense of the displayed dot field; thus, at any instant, the visible portion of the display subtended $60 \times 55°$. The software clipping window moved across the screen over time in the simulated rotation conditions to ensure that the exit of the dots from the display window was the same for the simulated and real rotation cases.

The dots themselves were composed of 2×2 clusters of pixels. We used an anti-aliasing routine to move the centroid of the 2×2 cluster smoothly over time. Dot positions were updated at the 75 Hz frame rate of the projector. The combination of fast refresh rate and anti-aliasing yielded an appearance of very smooth dot motion[3].

[3] Many of the functions used to generate the stimuli and collect responses are posted at the following Web Site: http://john.berkeley.edu/MatLab_ExperimentLibraries/Experiment/Libraries.html.

We used an anaglyphic technique to stimulate the two eyes separately. Green and red dots were viewed through green and red filters in front of the left and right eyes, respectively. The intensities of the green and red dots were adjusted to equate the luminances through the filters. There was no leakage of red and green between the eyes. There were three viewing conditions: monocular, synoptic, and dichoptic. In the monocular condition, observers viewed the display with one eye. In the synoptic condition, the dots were viewed with two eyes, but contained no binocular disparity; their positions on the screen were calculated from the perspective of a point midway between the two eyes. In the dichoptic condition, the dots were presented with appropriate binocular disparity given the interocular separation of each individual observer; disparities changed with the dots' motion in depth.

For the issues under examination in this paper, it is important that environmental features, such as the edge of the display screen, be made invisible. To accomplish this, the room was made completely dark except for the display. In addition, a bright uniform field was presented between trials to maintain light adaptation. With this setup, no environmental features could be seen during the stimulus presentations.

Before each trial, the first frame of the forthcoming motion sequence appeared until the observer initiated the sequence with a button press. The duration of the motion sequence was always 1500 ms. Depicted translation speed was 200 cm/s. Three types of scenes were presented in different experiments. The simulated scenes extended well beyond the limits of the clipping window.

At the end of a motion sequence, a bright dot appeared and the observer adjusted the azimuth of this response marker until it lay on the perceived future self-motion path. The simulated motion of the marker was constrained such that it moved in a circle with the observer at its centre. When the simulated scene contained a ground plane, the marker appeared on the ground at the appropriate distance. In the dichoptic condition, the marker distance was also specified by binocular disparity. No feedback was given during practice before actual data collection nor during the data collection phase. This is an important feature of the experimental design because feedback could train the observers to respond in ways that they would otherwise not respond.

Observers sometimes perceive an illusory motion of the fixation point during simulated eye rotations [6]. Van den Berg [20] reported recently that observers can use this illusory motion (which often contains a depth component) to judge the direction of self-motion. Theoretically, this cue could be useful when the fixation point is rigidly attached to the scene, but it is generally

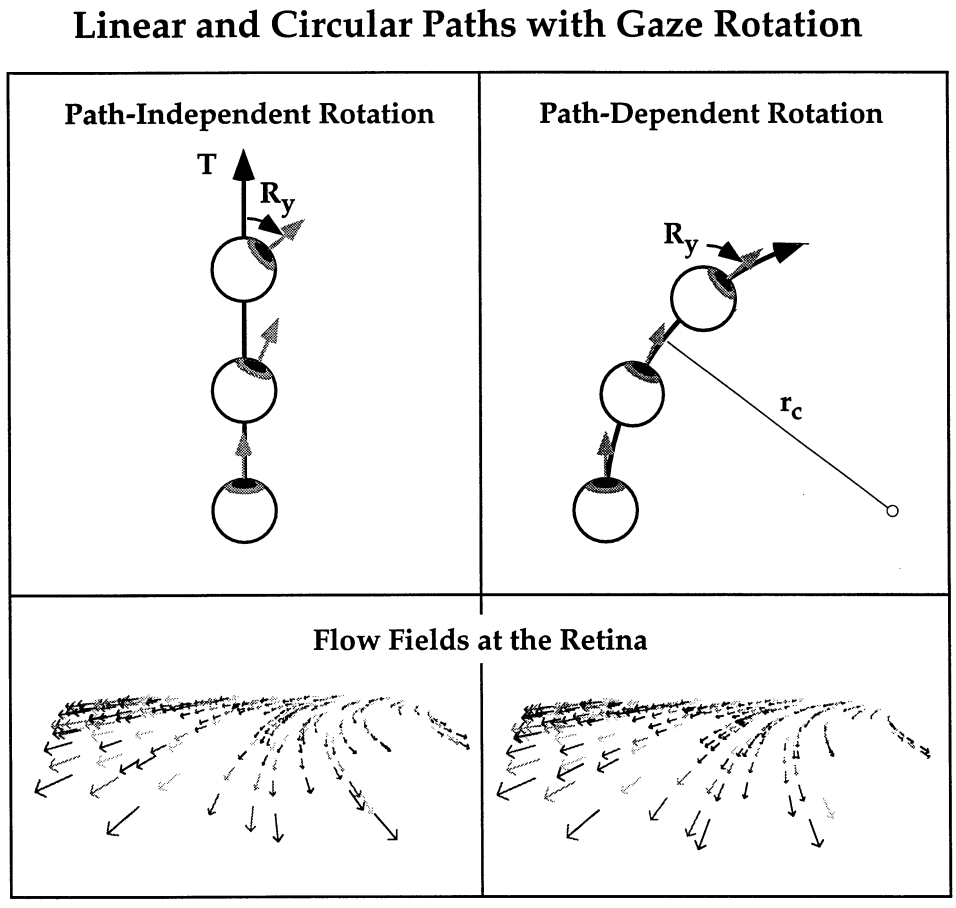

Fig. 1. Linear and circular self-motion paths with gaze rotation. The upper panels schematise two forms of self-motion. Upper left: Linear translation plus gaze rotation; the gaze rotation is path independent. Upper right: Circular-path translation (with radius of curvature r_C); the gaze rotation is path dependent. In both cases, the dark arrow represents the displacement of the observer over time and the gray arrows represent the gaze direction over time. The lower panels display locations and velocities of texture elements at the retina that occur for linear-path translation plus gaze rotation and for circular-path translation plus gaze rotation; the initial positions of the texture elements were randomly selected, so they are not the same in the left and right panels. Lower left: Initial direction of translation $=0°$, translation speed $=2$ m/s, rotation rate $=5$ deg/s, duration $=2$ s, and eye height $=1.6$ m. These parameters are similar to the conditions of some of the experiments reported here. Lower right: Initial direction of translation $=0°$, translation speed $=2$ m/s, rotation rate $=5$ deg/s, duration $=2$ s, radius of curvature $=22.9$ m, and eye height $=1.6$ m. The motion in the upper panels is over a much longer duration than the motions that produced the flow fields in the lower panels.

not useful when the fixation point is not attached[4]. At any rate, our observers were unaware of the potential usefulness of such a strategy and presumably did not use it unwittingly because they exhibited larger biases than van den Berg's [20] observers.

[4] During linear translation, the motion of a fixation point that is attached to the rigid scene is constrained to be away from the direction of self-motion; van den Berg's [20] cue can hence be used. However, when the fixation point is unattached its motion is independent of the translation, so its perceived motion is uninformative. This can be easily shown by considering the following situation. The translation is straight ahead ($T_X = 0$), the fixation point is off to the right ($X > 0$), and the rotation is to the left ($R_Y < 0$). In this case, during a simulated gaze rotation, the fixation point appears to move leftward and toward the observer which, according to van den Berg [20], would specify a self-motion path to the right of the fixation point!

3. Experiment 1: response marker distance and perceived path errors

Royden et al. [1,5], Banks et al. [15], and van den Berg [20] reported that observers perceive curved paths during simulated gaze rotations even though the display depicts linear translation plus gaze rotation. At each instant in a display simulating linear self-motion plus gaze rotation, there is a circular self-motion that would give rise to the same retinal velocity field (provided that the direction of gaze was always fixed with respect to the direction of self-motion) [1,17,20,22]. Let us examine those two cases, which are schematized in Fig. 1, more closely. In the case of linear translation plus rotation (left panel), the direction of gaze rotates independently from the translation; the rotation is path independent. Therefore, the direction of translation

changes over time in oculocentric coordinates, but does not change with respect to the world. In the case of curvilinear translation (right panel), the gaze direction is determined by the curvature of the self-motion path; the rotation is path dependent[5]. Therefore, the direction of translation remains fixed in oculocentric coordinates, but varies with respect to the world. The two situations diverge over time: The retinal velocity fields at different instants are consistent with different circular paths so the trajectories of features in the retinal image become less consistent with a unique circular-path interpretation as duration increases. This divergence, however, is small over short time intervals for the experimental conditions in the literature [17]. The lower panels depict the flow fields at the retina for linear translation plus gaze rotation (left panel) and for circular-path translation (right panel). The depicted flow fields represent 2 s of self-motion. The vectors represent both the locations and velocities of texture elements at different times; the base of each vector shows the position and the arrow indicates the instantaneous velocity. The translation speeds, rotation rates, and durations are given in the figure caption.

Most experiments on perceived self-motion employ one of two psychophysical tasks. In one, the observer is asked whether the perceived self-motion path would carry him or her to the left or right of a landmark in the scene [6]. In the other, the observer is asked to place a marker on the perceived path at a specified distance from the final station point. Whenever a curved path is perceived in a display depicting linear self-motion, the error between the depicted and perceived paths should depend on distance [17]. Fig. 2 illustrates this point. The perceived path is curved to the right. When the observer places the response marker on the perceived path, the azimuth of the placement should depend on the marker's perceived distance. For example, if the marker distance appeared to be large, the observer would probably place it toward the right edge of the screen (angle α); if the marker appeared to be at close range, the observer would probably place it near the middle of the screen (β). Thus, whenever a curved path is perceived in these displays, the apparent distance to the marker becomes significant in interpreting the data.

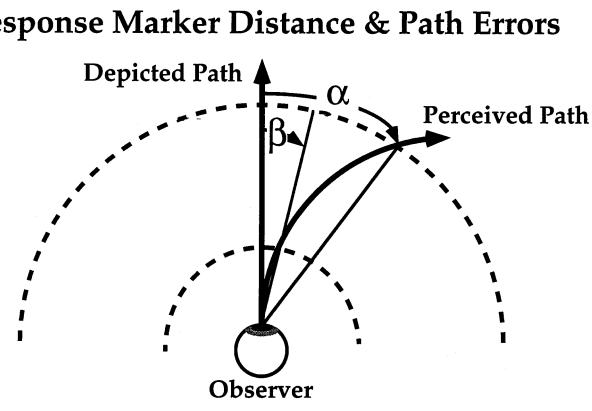

Fig. 2. The relationship between response marker distance and self-motion path judgement. The straight line from the observer represents the direction of depicted translation in our displays; the curved line represents the perceived path (for a rightward simulated eye rotation). The dashed semicircles represent two response marker distances, one near and one far. If the observer perceived a curved self-motion path, he/she would place the marker near the direction of depicted linear translation for near markers (indicated by the angle β), but would place the marker far to the right of the depicted direction for distant markers (indicated by α). The error in indicating self-motion direction, therefore, ought to be roughly proportional to response marker distance when circular paths are perceived.

[5] Of course, there is an infinite set of combinations of curved paths plus gaze rotations that are equally consistent with the instantaneous retinal velocity fields. We have chosen only two simple cases, one in which the translation is linear and all the rotation is made independently from the translation and one in which the translation is circular and none of the rotation is made independently from the translation.

In Experiment 1, we examined the effect of the response marker distance on judgements of perceived self-motion paths.

3.1. Method

The stimuli depicted linear translation across a ground plane or through a 3D cloud. In the case of the ground plane, translation was parallel to and 160 cm above the plane. At the beginning of each motion sequence, the plane had simulated depths of 300–2000 cm and 250 dots were visible on average. The dots were randomly positioned in the plane according to a uniform distribution, so a geometrically-correct density gradient was present in the 2D projection. In the case of the 3D cloud, translation was in the horizontal plane through the horizontal meridian of the display. At the beginning of each sequence, the dot distances were 300–2000 cm and 250 dots were visible on average. The dots were positioned according to a uniform distribution in the simulated 3D cloud. The ground-plane and cloud stimuli were viewed binocularly and the dots were presented with geometrically-correct binocular disparities.

Three translation directions were presented: straight ahead (0°), to the left ($-10°$), and to the right (10°). Only simulated eye rotations were presented. The rotation axis was vertical and the magnitudes varied from 0–5°/s in both directions. During simulated eye rotations, the fixation marker was stationary and appropriate rotational flow was added to the motions of the dots. The fixation point was not rigidly attached to the depicted 3D scene. The distance to the response marker was varied systematically: 250, 500, 1000, 1500, and 2000 cm. The marker distance was specified in all cases by binocular disparity; in the case of the ground plane,

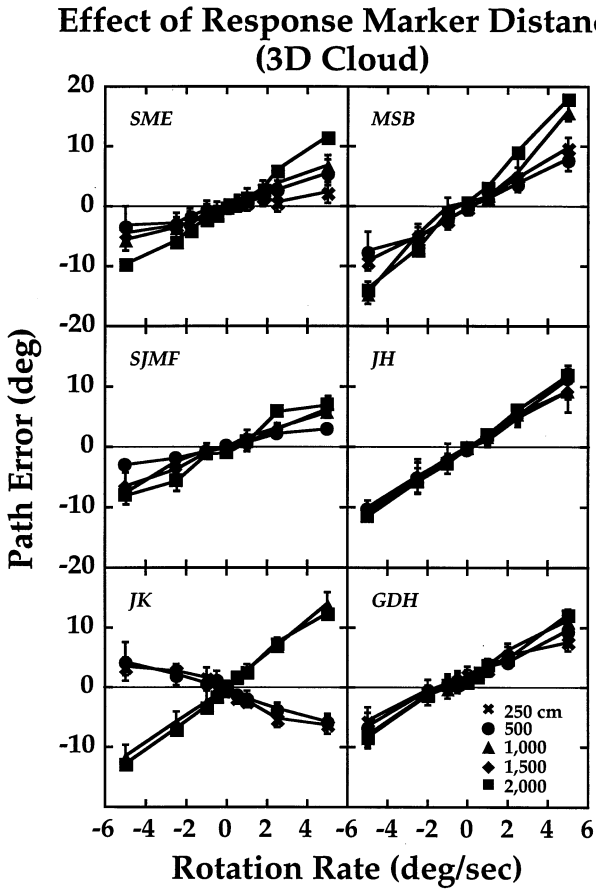

Fig. 3. Experiment 1: Effect of response marker distance during simulated eye rotations. Path judgement errors are plotted as a function of rotation rate. The dichoptic stimulus depicted linear translation through a 3D cloud of dots. The thin horizontal lines indicate errors of zero. Different symbols represent different response marker distances: X's represent 250 cm, circles 500 cm, triangles 1000 cm, diamonds 1500 cm, and squares 2000 cm. Each data point represents the mean of 18–30 judgements. Error bars represent ±1 S.D.

it was also specified by the vertical position of its base. Four observers participated. They made 9–30 judgements (3–10 at each of three self-motion directions) for each rotation rate.

3.2. Results and discussion

The results for the cloud and ground plane are displayed in Fig. 3 and Fig. 4, respectively. The angle between the direction of depicted translation at trial end and the average marker setting is plotted as a function of rotation rate. Veridical responses would lie on the horizontal line at 0°. Specified distances to the response marker are indicated by different symbols. Observers perceived paths curved in the direction of the rotation and, accordingly, marker settings deviated consistently from the depicted translation path.

With the exception of JK's data[6], there was a systematic, but small effect of response marker distance on path error with the 3D-cloud stimulus (Fig. 3). To assess the relationship between response errors and response marker distance, we fit each observer's data at each marker distance with a regression line and then computed the correlations between marker distance and the slopes of the regression lines. These correlations were 0.9–1.0 for observers SME, MSB, SJMF, and GDH; they were 0.3 and 0.1 for JH and JK, respectively. Although there was a clear effect of marker distance in four of the six observers, it was smaller than expected by the hypothesis depicted in Fig. 2. Observers reported that disparity alone in this stimulus did not yield a clear sensation of response probe distance; this lack of perceptual salience may explain the small effect of specified response probe distance.

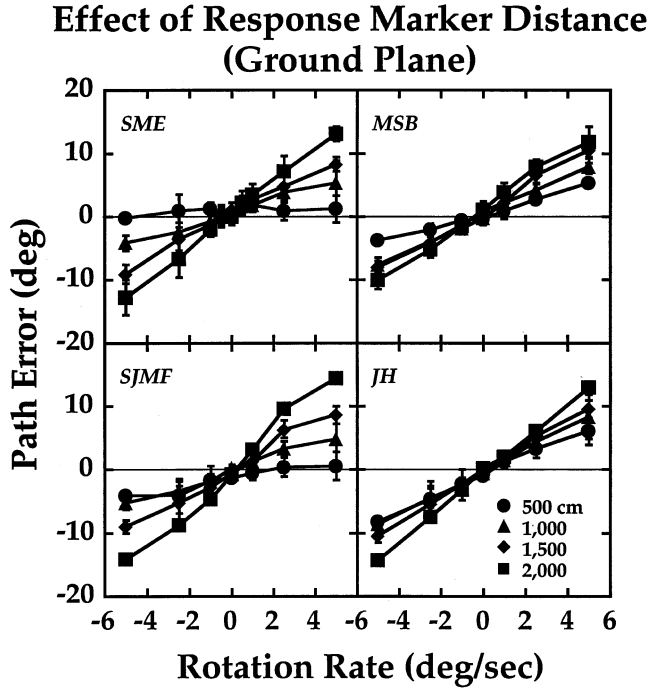

Fig. 4. Experiment 1: Effect of response marker distance during simulated eye rotations when the stimulus is a ground plane. Path judgement errors—the angular differences between the depicted translation and the observers' settings—are plotted as a function of rotation rate. The dichoptic stimulus depicted linear translation across a ground plane. The thin horizontal lines indicate errors of zero. Different symbols represent different response marker distances: Circles represent 500 cm, triangles 1000 cm, diamonds 1500 cm, and squares 2000 cm. Each data point represents the mean of 9–30 judgements. Error bars represent ±1 S.D.

[6] When the response marker distance was 1000 or 2000 cm, JK's responses were quite similar to the other observers'; however, when the marker distance was 250 or 500 cm, her responses exhibited a bias in a direction opposite to that of the other observers. Some observers in van den Berg and Brenner [21] exhibited similar behavior. We have no explanation for reverse bias.

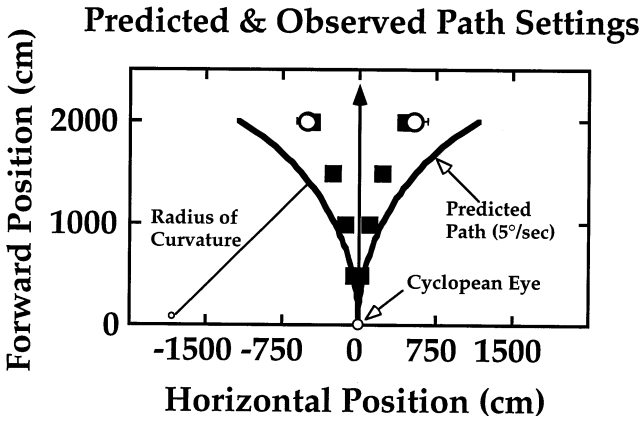

Fig. 5. Predicted and observed path settings for ground-plane data of Experiment 1. Predicted and observed path judgement errors are plotted in the *X-Z* plane in front of the observer. The predictions and data are for a rotation rate of $\pm 5°$/s. The solid curves represent the predicted paths in the *X-Z* plane according to the circular-path hypothesis of Royden [17]. The filled symbols represent the average settings across four observers when the stimulus depicted a linear path plus simulated gaze rotation. The unfilled symbols represent the average settings across the same four observers when the stimulus depicted a circular path.

With the ground-plane stimulus, the vertical placement of the marker on the projection of the ground surface provided additional distance information: The four observers whose data also appear in Fig. 3 reported a clearer distance percept. As one might expect, there was a much more systematic relationship between marker distance and path error (Fig. 4): As the specified marker distance increased, the magnitude of the response errors increased as well. Correlations for all four observers were 0.95–1.0.

We make the assumption that observers perceive circular self-motion paths with no path-independent rotation instead of the linear self-motion with path-independent rotation that is actually depicted. If this assumption is correct, we should be able to predict their path errors. Following Royden [17], we assume that the curvature of the perceived path can be calculated from the following equation:

$$r_C = T/R_Y \quad (1)$$

where r_C is the radius of curvature of the perceived circular path, T is the translation speed, and R_Y is the rotation rate (in radians/s) about the vertical axis[7]. These quantities are labelled in Fig. 1. We used Eq. (1) to calculate the expected responses at the end of the motion sequence at different marker distances for the ground-plane version of Experiment 1. Observers were better able to perceive the specified response marker distance in the ground plane than in the cloud, so we did not do these calculations for the cloud data. Predicted and observed responses are shown in Fig. 5 for rotation rates of $\pm 5°$/s. The figure is a plan view of the ground plane in front of the observer. The simulated linear translation is represented by the vertical arrow and the circular paths predicted from Eq. (1) are represented by the solid curves. The filled symbols represent the average of four observers' placements of the response marker (same data replotted from Fig. 4). The observed and predicted settings exhibit increasing deviations from the depicted linear path with increasing response marker distance. However, at the longer distances, the observed errors are smaller than predicted by the circular-path hypothesis.

Fig. 6 plots predicted and observed path errors for all conditions in the ground-plane version of Experiment 1. The symbols represent the average horizontal path errors of individual observers and the lines represent the errors predicted by Eq. (1). The equation predicts monotonically increasing path errors with increasing rotation rate and all four observers exhibited such errors. However, the observed errors are again smaller than predicted by the circular-path hypothesis.

Because the observed errors are smaller than predicted, these data seem inconsistent with the circular-path hypothesis that path-independent rotation, not accompanied by an actual eye (or head) rotation, is

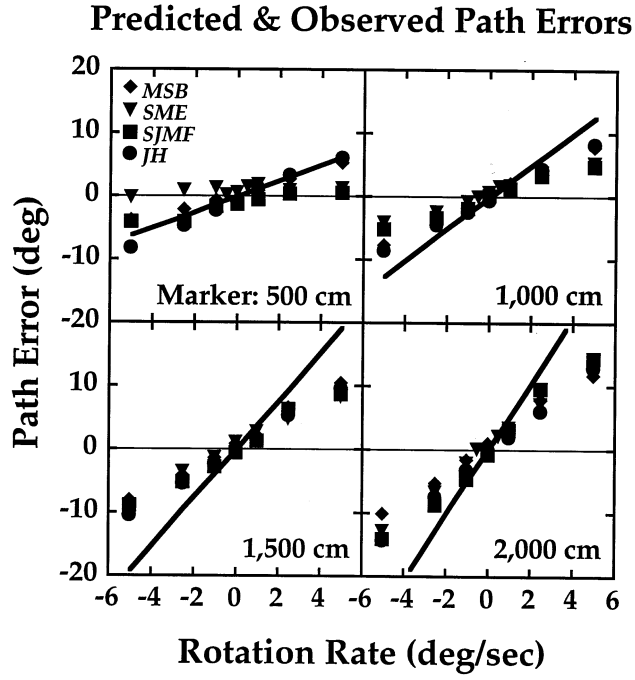

Fig. 6. Predicted and observed path errors for ground-plane data of Experiment 1. Path judgement errors are plotted as a function of rotation rate. The lines represent the errors predicted by the model of Royden [17]. The symbols represent the actual judgement errors for the four observers participating in the ground-plane version of Experiment 1.

[7] Royden's equation assumed a constant Z value to specify the distance to the response marker whereas our experiments used a constant radius from the cyclopean eye. We compensated for this small difference in calculating the values presented in Figs. 5–7.

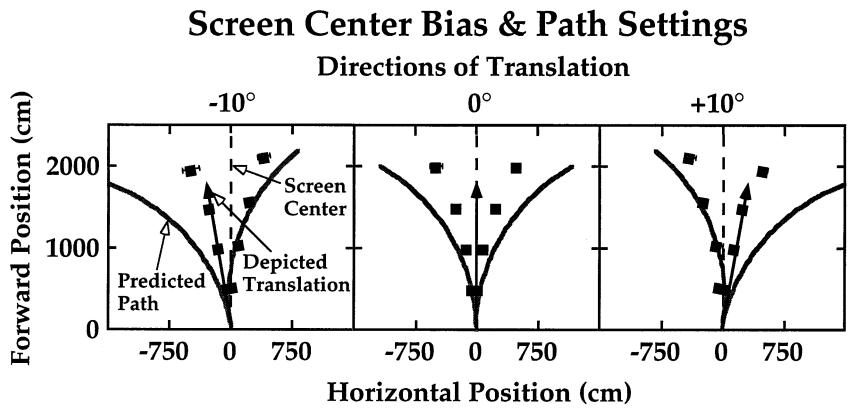

Fig. 7. Path settings for ground-plane data of Experiment 1 for the three translation directions. Each panel shows the observed settings and the ones predicted by the circular-path hypothesis in screen coordinates. The symbols represent the settings and the curves the predictions. The predictions are for the temporal midpoint of the motion sequence. The depicted translations are represented by the solid arrows and they are the directions in screen coordinates at the temporal midpoint of the motion sequence. The dashed lines represent the direction of the centre of the screen (and of the fixation cross).

completely attributed to path-dependent rotation [15,17]. We cannot reject this hypothesis, however, without knowing how observers respond when presented displays depicting a truly circular self-motion path with path-dependent rotation (see Fig. 1). To determine this, we presented optic flow displays depicting circular paths across a ground plane. The specific paths were those predicted by [17] for the end of the motion sequences depicting linear paths plus simulated gaze rotation. The path-dependent rotation rate was $\pm 5°$/s; the path-dependent rotation was simulated, so the fixation marker remained fixed in screen coordinates. The depicted distance to the response marker was 2000 cm. Four observers indicated whether their perceived path would carry them to the left or right of the marker. A 1-down/1-up staircase procedure was used to adjust the azimuth of the marker. The results are represented by the unfilled symbols in Fig. 5. Notice that the settings for the circular-path displays were very similar to the settings for the displays depicting linear paths plus simulated gaze rotation. Thus, it appears that two types of stimuli yield similar path percepts although the perceived path may not be precisely circular [17].

Why do observers' responses fall short of the values predicted by the circular-path hypothesis? The shortfall may be the consequence of a bias to see the self-motion path closer to the centre of the screen (or to the fovea) than it really is[8]. To examine this possibility, we reanalysed the data in Fig. 5. Specifically, we analysed separately trials in which the predicted circular path was toward the centre of the screen (and the fovea) and trials in which the predicted path was away from screen centre (and the fovea). Fig. 7 displays the outcome. The three panels show the depicted paths, predicted circular paths, and marker settings in the format of Fig. 5. The left, middle, and right panels show the paths and settings when the depicted directions of translation were respectively 10° to the left of screen centre, toward screen centre, and 10° to the right of screen centre. During simulated gaze rotations, the directions of translation rotate over time in screen coordinates; we have displayed the directions (and predicted circular paths) corresponding to the midpoint of the motion sequence. The rotation rates were $\pm 5°$/s, so two predicted paths-one curving to the left and the other to the right-are displayed in each panel. Notice that observers' settings were in fact biased toward the centre of the screen (or toward the fovea). Consider, for example, the two cases in which the circular-path hypothesis predicts perceived paths that are far from screen centre: Direction equals $-10°$ (left of screen centre) with a leftward rotation (depicted in left panel) and direction equals 10° (right of centre) with a rightward rotation (right panel). In the former case, the circular-path hypothesis predicts a perceived path near the left edge of the display screen and, in the latter, it predicts a perceived path near the right edge. Notice that observers' settings were closer to screen centre than predicted for both cases. Now consider the two cases in which the circular-path hypothesis predicts paths near the screen centre: Direction equals $-10°$ (left panel) with a rightward rotation and direction equals 10° (right panel) with a leftward rotation. Notice that observers' settings were in both cases very close to the predictions of the circular-path hypothesis. These data support the idea that observers' settings were affected by bias toward screen centre (or fovea).

[8] We cannot distinguish between biases toward the centre of the screen as opposed to toward the fovea because the line of sight was always directed toward the screen centre during simulated gaze rotations.

To sum up, the most important observation is that settings were very similar for displays depicting linear translation plus simulated gaze rotation and displays depicting circular translation whether affected by centre-of-screen biases or not[9]. Such a similarity is predicted by the main underlying assumption of the circular-path hypothesis: Path-independent rotation, not accompanied by an appropriate extra-retinal, eye-velocity signal, is attributed to path curvature and to path-dependent rotation (see Fig. 1).

4. Experiment 2: eye movements during real and simulated rotation conditions

The interpretation of data from experiments using simulated and real eye movements hinges on the assumption that the retinal images in the two conditions are identical over time [1,2,5–7,15–21]. Any variation from the intended pursuit velocity in the real eye movement condition or from zero eye velocity in the simulated rotation condition would produce different temporal characteristics in the retinal images than intended.

4.1. Method

To determine whether rotational flow during simulated and real eye rotations was in fact equivalent for our stimulus conditions, we measured eye movements in two observers-an author (SME) and a naive subject (SJMF)-during a replication of the ground-plane version of Experiment 1. To do so, we used a limbus eye tracker (Eye Trac Model 210, Applied Sciences Laboratories). The observers' heads were fixed by use of a bite bar; the tracker was mounted to the same apparatus. Eye position was sampled at 300 Hz. In order to calibrate the eye-movement recorder, observers fixated in sequence nine points at horizontal intervals of 2.5° and the mapping between position and voltage in the instrument was thereby determined.

Once the instrument was calibrated, we began the experimental session. The ground-plane stimuli were viewed monocularly. Observers fixated the fixation marker and performed the experimental task as they had done in Experiment 1. The fixation marker was presented at a distance of 1000 cm and the response marker at 2000 cm. There were three translation directions (0 and $\pm 10°$) and five rotation rates (0, ± 2.5, and $\pm 5°$). Trial duration was 1.5 s. The path errors exhibited by the observers were indistinguishable from those shown in Fig. 4.

[9] We could not perform a similar analysis for settings with the circular-path displays because the software used to run that experiment did not store data separately for the three initial translation directions.

Average eye velocity was calculated off-line in MATLAB. To do so, position records were first low-pass filtered and then the time derivative was taken. Saccades were removed using a velocity amplitude criterion of 10°/s. The average speed was computed over the remaining record excluding the initial and final 75 ms.

4.2. Results and discussion

Fig. 8 shows average eye velocity plotted as a function of rotation rate for the real and simulated eye movement conditions. Each data point represents the mean of nine tracking records. The circles and squares represent the data from the real and simulated movement conditions, respectively. If tracking were completely accurate in the real movement condition, the data points (circles) would lie on the diagonal lines. The average pursuit velocity was closely matched to the velocity of the fixation marker in this condition; the mean gains averaged across rotation rates were 0.8 and 1.2 for observers SME and SJMF, respectively. There were also relatively few saccades (as defined above) in

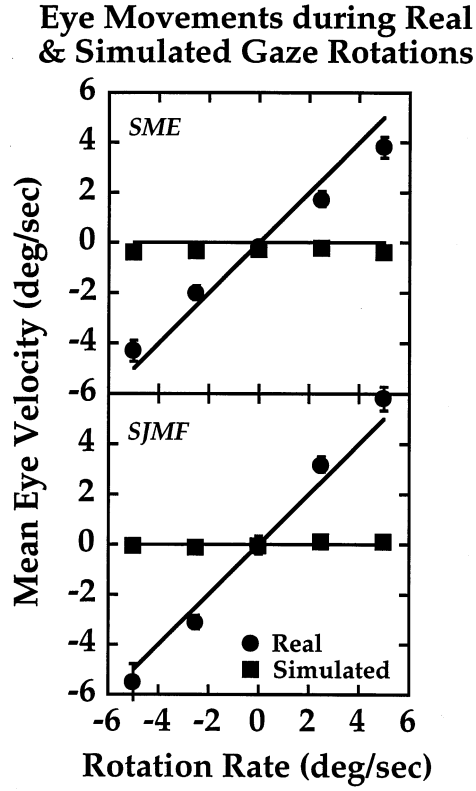

Fig. 8. Eye velocity during real and simulated eye movements. The average eye velocity is plotted as a function of rotation rate. The diagonal lines represent the expected eye velocity if tracking were perfectly accurate in the real movement condition; the circles represent the data for that condition. The horizontal lines represent the expected eye velocity if fixation were accurate in the simulated movement condition; the squares represent the data for that condition. Each data point represents the mean of nine recordings. Error bars represent ± 1 S.D.

this condition: On average, there were 1.1 and 0.5 detected saccades per trial for SME and SJMF, respectively. If fixation were completely accurate in the simulated rotation condition, the data from that condition (squares) would lie on the horizontal lines. The average eye velocities were close to zero; average velocities were $-0.3°$/s and $-0.0°$/s for observers SME and SJMF, respectively. In addition, there were only 0.4 and 0.1 detected saccades per trial for SME and SJMF, respectively. Thus, fixation was indeed accurate during the simulated eye movement condition.

In conclusion, eye movements were generally quite accurate in the real and simulated movement conditions, so the assumption of equivalent retinal images over time in the two conditions is valid.

5. Experiment 3: stereoscopic depth cues and perceived paths during simulated rotations

The results of Experiment 1 make the important point that observers' path errors decrease monotonically as the distance of the response marker is reduced. In van den Berg and Brenner's [2] experiment, the marker distance was only 575 cm, so it is difficult to determine from their result whether observers perceived linear or curved paths. Thus, we decided to re-examine the issue of whether binocular disparity aids performance during simulated gaze rotations. We used a large marker distance of 1000 cm in order to increase the likelihood of differentiating linear from curved path percepts.

5.1. Method

The stimuli and procedure were identical to those in Experiment 1 with three exceptions: (1) Three viewing conditions were presented: monocular, synoptic, and dichoptic, (2) real and simulated eye rotations were presented, and (3) the distance to the response marker was always 1000 cm. Once again, the stimuli depicted translation across a ground plane or through a 3D cloud. Four observers participated. They made 18–30 judgements (6–10 at each of three self-motion directions) for each combination of rotation and experimental condition.

5.2. Results and discussion

The results are shown in Figs. 9–11. Fig. 9 and Fig. 10 display the results when the stimulus was a 3D cloud and Fig. 11 displays the results when it was a ground plane. Fig. 9 displays results for the real eye movement condition and Fig. 10 and Fig. 11 those for the simulated movement condition. In each figure, error in the judged self-motion path (averaged across the three translation directions) is plotted as a function of the rotation rate.

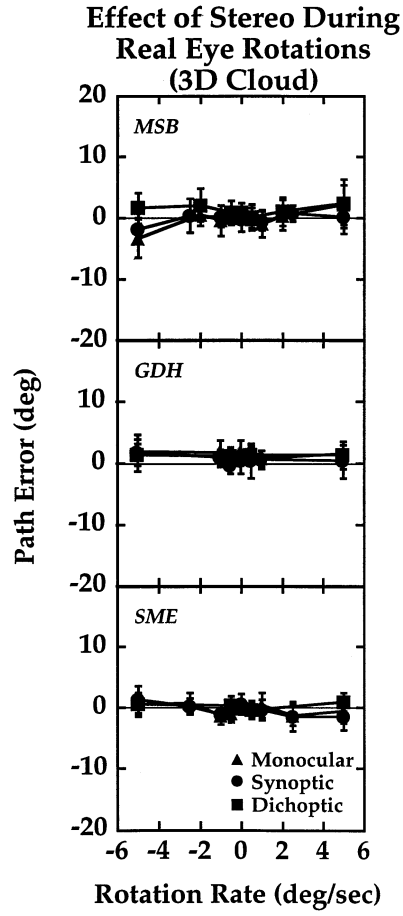

Fig. 9. Experiment 3: Perceived self-motion path errors during real eye rotations with monocular, synoptic, and dichoptic viewing. Path judgement errors are plotted as a function of rotation rate. The stimulus depicted linear translation through a 3D cloud of dots. The thin horizontal lines indicate errors of zero. Results from the monocular condition are represented by triangles, those from the synoptic condition by circles, and those from the dichoptic condition by squares. Each data point represents the mean of 18–30 judgements. Error bars represent ± 1 S.D.

Fig. 9 reveals that all three observers tested in the real eye rotation condition were able to estimate the direction of translation quite accurately at all rotation rates and for monocular, synoptic, and dichoptic viewing. This result is consistent with the existing literature [1,5–7,15,20].

Fig. 10 shows that the same three observers responded quite differently in the simulated eye movement condition. The perceived self-motion path was curved in the direction of and by an amount proportional to the simulated eye rotation. Interestingly, the path errors were not systematically smaller in the dichoptic than in the monocular and synoptic viewing conditions. The variability of the responses, indexed by S.D., was also similar across the three conditions. These results are both consistent and inconsistent with van

den Berg and Brenner's [2]. They are consistent in the sense that we find no improvement in performance with the addition of depth information via binocular disparity when the signal-to-noise ratio in the display is high (they find improved performance at low signal-to-noise ratios only). They are inconsistent in the sense that we observe large and consistent path judgement biases during simulated eye rotations and two of van den Berg and Brenner's three observers (EB and RG) exhibited rather small biases. However, the distance to their response marker was only 575 cm, so even if their observers perceived illusory curved paths, their responses would be similar to those for the depicted linear paths.

Fig. 11 displays the results when the stimulus depicted motion across a ground plane and the eye rotation was simulated. Again there was no systematic effect of viewing condition: The perceived path was curved in the direction of and by an amount proportional to the simulated eye rotation for monocular, synoptic, and dichoptic viewing conditions. These results are again both consistent and inconsistent with van den Berg and Brenner [2]. They are consistent in the sense that they too observed little effect of viewing condition when the stimulus was a ground plane and are inconsistent in the sense that they reported accurate path estimation under such conditions and our observers exhibited large and systematic biases. As noted above, however, van den Berg and Brenner's [2] use of a near response marker may have obscured path errors.

It is interesting to note that response variability did not decrease with addition of stereoscopic cues. The average S.D. across observers and rotation rates for the data of Fig. 10 (3D cloud) were 1.97, 1.99, and 1.80 for the monocular, synoptic, and dichoptic conditions, respectively. The average S.D. for the data of Fig. 11 (ground plane) were 1.53, 1.41, and 1.94 for the monocular, synoptic and dichoptic conditions, respectively. Thus, response variability was not reduced by the addition of stereoscopic depth cues.

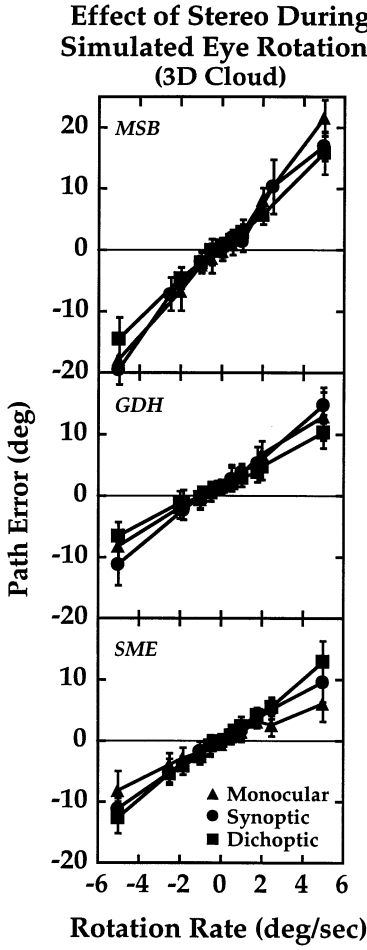

Fig. 10. Experiment 3: Perceived self-motion path errors during simulated eye rotations with monocular, synoptic, and dichoptic viewing. Path judgement errors are plotted as a function of rotation rate. The stimulus depicted linear translation through a 3D cloud of dots. The thin horizontal lines indicate errors of zero. Symbol conventions are the same as in Fig. 9. Each data point represents the mean of 18–30 judgements. Error bars represent ±1 S.D.

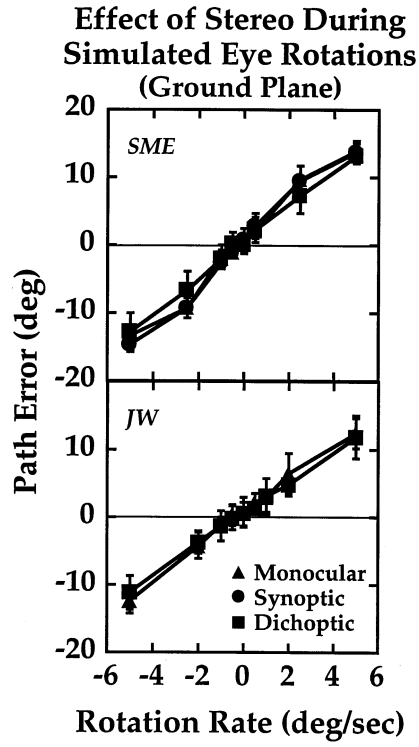

Fig. 11. Experiment 3: Perceived self-motion path errors during simulated eye rotations with monocular, synoptic, and dichoptic viewing when the stimulus is a ground plane. Path judgement errors are plotted as a function of rotation rate. The stimulus depicted linear translation across a ground plane. The thin horizontal lines indicate errors of zero. Symbol conventions are the same as in Fig. 9 and Fig. 10. Each data point represents the mean of 18–30 judgements. Error bars represent ±1 S.D.

6. Experiment 4: replication of van den Berg and Brenner [2]

Unlike van den Berg and Brenner [2], we find large and systematic path judgement biases during simulated eye rotations even when binocular disparity is present in the displays. It should be noted that van den Berg and Brenner's experimental goal was different than ours: They examined whether the addition of disparity decreases the variability and bias in judgements of self-motion paths in noisy displays and we examined whether adding disparity to noise-free displays decreases the bias in path judgement errors reported by Royden, Banks, Crowell, and colleagues [1,5,15]. Accordingly, van den Berg and Brenner varied the signal-to-noise ratio in their displays and we always presented noise-free displays. Despite these differences in approach, the results still appear incompatible because our observers' path judgements were systematically biased when there was no noise in the display and van den Berg and Brenner's observers' judgements were apparently more accurate than ours when the display contained noise.

Besides the manipulation of signal-to-noise ratio, there are three other differences between our Experiments 1 and 3 and those of van den Berg and Brenner [2]. First, as mentioned above, van den Berg and Brenner placed their response marker at a simulated distance of 575 cm; as we showed in Experiment 1, path judgement errors tend to be small at such short marker distances. Second, the fixation point in van den Berg and Brenner's displays was rigidly attached to the scene and in ours it was not; this may be important because the computational problem of estimating translation is simplified with an attached fixation point [10,11][10]. Third, the translational flow speeds were higher in van den Berg and Brenner's experiments than in ours. They used a slower depicted translation speed of 150 cm/s (vs. 200 cm/s), but in the ground-plane experiments, their simulated observation height was 65 cm (vs. 160 cm) and in their 3D-cloud experiments the nearest points were 100 cm (vs. 300 cm). Therefore, for a given rotation rate, the ratio of translational flow speed divided by rotational flow speed in their ground-plane experiments was about twice the ratio in our experiments[11]. One would expect more accurate judgements of self-motion paths in their experiment than in ours because the rotational flow in theirs was a smaller fraction of the translational flow.

Given these dissimilarities, it seemed worthwhile to conduct a replication of their experiment. Experiment 4 is designed accordingly.

6.1. Method

The stimuli and procedure were identical to the 3D-cloud version of Experiment 1 with two exceptions. First, the specified distance of the response marker was 575 cm as in van den Berg and Brenner [2]; its distance was specified by binocular disparity. Second, the fixation point was attached to the cloud at an initial distance of 8 m. Because the fixation marker was attached to the cloud, its simulated distance decreased over the duration of the motion sequence and, consequently, the simulated rotation rate increased over time.

As in Experiment 1, only the dichoptic viewing condition was presented. Four observers participated; observer SJMF was naive to the experimental hypotheses. Observers made nine judgements (three at each of three self-motion directions) for each rotation rate.

6.2. Results and discussion

The results are displayed in Fig. 12. Path judgement errors are plotted as a function of the time-average rotation rate. The reader should note that van den Berg and Brenner [2] plotted the azimuths of the responses as a function of the azimuths of the simulated directions of self-motion. Fig. 12 uses our conventional format of plotting response errors (in degrees) as a function of rotation rate. The upper panel shows individual data from our four observers. The lower panel shows the range of data from our four observers and the range of data from van den Berg and Brenner [2]. The three observers in van den Berg and Brenner behaved differently, but this means of plotting the data shows that the errors exhibited by our observers are not substantially different than the ones exhibited by theirs. We compared the slopes of the data averaged across observers in the two experiments. Data from the 5°/s rotation rate

[10] When the fixation point is attached to the scene, the axis and direction of rotation are completely determined and the eye must rotate directly away from the heading. Thus, the number of degrees of freedom in the estimation of the translation direction is reduced.

[11] Points in the scene have the coordinates (X, Y, Z) and points in the retinal image have the coordinates (x, y) where $x = -X/Z$ and $y = -Y/Z$. Consider a situation in which the direction of translation is straight ahead (e.g. the only motion component is in Z) and the rotation is about a vertical axis. The translational flow speeds in directions x and y are xT_Z/Z and yT_Z/Z, respectively. Now consider a retinal image point with the coordinates (0, 1). In the ground-plane experiments, the point in the scene corresponding to that point in the retina is $(0, -65, 65)$ for van den Berg and Brenner and $(0, -160, 160)$ for our experiments. The translation flow speeds at that point would be 2.31 cm/s for the van den Berg and Brenner ground plane and 1.25 cm/s for ours, a ratio of 1.85:1. In the 3D-cloud experiments, the nearest point in the scene corresponding to (0, 1) in the image is $(0, -100, 100)$ for van den Berg and Brenner and $(0, -300, 300)$ for our experiments. The translational flow speeds at that point would be 1.5 cm/s for the van den Berg and Brenner 3D cloud and 0.67 cm/s for ours, a ratio of 2.25:1. However, very few of these nearest points are visible.

were excluded because van den Berg and Brenner's experiment did not include this rate. The averaged path errors from the two experiments did not differ significantly ($F(1, 5) = 1.063$, $P > 0.1$). Thus, when the experimental displays, the task, and the manner of plotting the data are made similar, the path errors we observe are generally small as reported by van den Berg and Brenner [2]. The key point is that the same observers exhibited large and consistent errors during simulated eye rotations when the response marker distance was large (see Figs. 3–7, 10 and 11). Thus, the observation of small errors when the response marker is near and the translational/rotational flow ratio is large does not necessarily imply that observers do not rely on extra-retinal, eye-velocity signals while estimating direction of self-motion during eye rotations.

7. Experiment 5: monocular depth cues and perceived paths during simulated rotations

We have shown that the addition of binocular disparity does not yield significant improvement in the ability to estimate the path of self-motion during simulated eye rotations. There is, of course, other information in the display that specifies a flat surface; for example, the sizes of the dots do not change over time as they would in the real world. Perhaps the displays simply do not contain enough depth information to specify the 3D structure of the depicted scene. We tested this possibility by adding monocular depth information. Specifically, we added depth information from occlusion, dynamic occlusion, relative and changing size, and linear perspective.

7.1. Methods

The methods in this experiment were identical to the ground-plane version of Experiment 1 with the following exceptions. The ground plane contained 400 dots. Eight vertical walls of equal objective size were positioned in a regular 4×2 grid on the ground plane; they were rigidly attached to the ground plane. The walls were specified by dots alone (that is, their border was defined by the occlusion relationship with the background dots and not by visible contours). Each wall contained 70 dots and the dot motions were geometrically correct for the depicted motion of the observer with respect to the wall. The initial depicted distances to the walls were 700 and 1200 cm. The walls were transparent in one condition and provided linear perspective and relative and changing size cues to depth. They were opaque in another condition and therefore provided occlusion and dynamic occlusion cues as well. The walls had the appropriate sizes and binocular disparities for their simulated depths and they grew as they approached the observer (although, as before, the dots defining the walls were fixed in size). In order to assess the effect of adding the walls, we also ran a condition without walls. Four observers participated. Only the dichoptic viewing condition was run. The simulated distance to the response marker was 2000 cm.

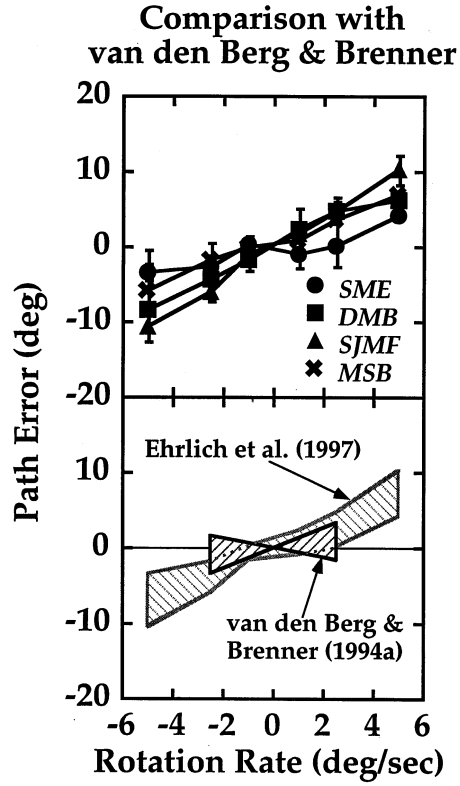

Fig. 12. Perceived path errors in van den Berg and Brenner [2] and in Experiment 4 for the 3D-cloud stimulus. Path judgement errors are plotted as a function of time-average rotation rate. The upper panel shows the individual data from Experiment 4; each data point represents the mean of nine judgements and error bars represent ± 1 S.D. The lower panel shows the range of data from Experiment 4 and the range of data from the three observers of van den Berg and Brenner [2]; the lines for the van den Berg and Brenner data were derived from the slopes of the regression lines they reported.

7.2. Results and discussion

Fig. 13 shows the data for displays with opaque walls, transparent walls, and no walls. There were no discernible differences in path errors among these three conditions: Path judgements were again displaced in the direction of and by an amount proportional to the simulated rotation in all conditions. Therefore, the addition of monocular depth information from occlusion, dynamic occlusion, relative and changing size, and linear perspective had no systematic effect on the ability to estimate self-motion paths during simulated eye rotations.

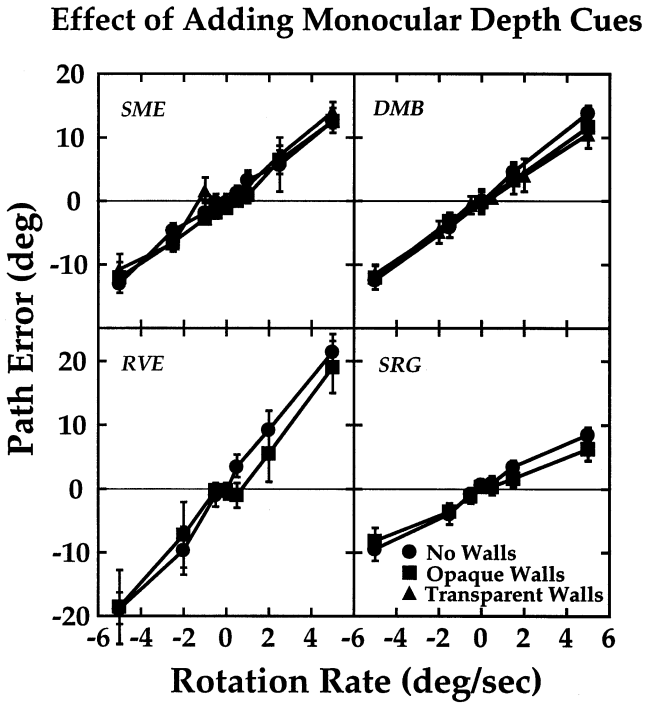

Fig. 13. Experiment 5: Effect of additional depth cues during simulated eye rotations. Path judgement errors are plotted as a function of rotation rate. The dichoptic stimulus depicted linear translation across a ground plane. The thin horizontal lines indicate errors of zero. The triangles represent the path judgements when transparent walls were present, the squares represent judgements when opaque walls were present, and the circles represent judgements when no walls were present. Each data point represents the mean of 18–30 judgements. Error bars represent ± 1 S.D.

8. Discussion and conclusion

Adding binocular and monocular depth information did not yield a significant improvement in the estimation of the self-motion path during simulated gaze rotations. Specifically, we found in Experiment 3 that adding binocular information for depth did not reduce the bias in path judgements and we found in Experiment 5 that adding a variety of monocular depth cues also did not reduce bias.

8.1. Methodological issues

The results of Experiments 1 and 4 illustrate the importance of two methodological parameters for experiments on perceived self-motion. First, whenever the observer perceives self-motion along a curved path, the distance to the response marker becomes critically important. In particular, the use of a marker at short distance can mask the presence of errors that are readily apparent at long marker distances. Second, from a computational standpoint, the ratio of translational flow divided by rotational flow ought to affect the ability to extract both the translation and the rotation from the flow field [23], both of which are required by a curved-path computation [17]. Stone and Peronne [24] reported systematic effects of the translation-rotation ratio on self-motion perception; we did not explicitly test this here. Notice that this ratio of translational/rotational flow speed is affected by a number of parameters including depicted translation speed and direction, rotation speed and direction, and the 3D layout of the scene. Given these complexities, it might be best in the future to assess human performance with respect to an efficient translation estimator [21,23,25].

8.2. Comparison with previous work by Royden, Banks, Crowell, and colleagues

Previous experiments by Royden, Banks, Crowell, and colleagues on perceived self-motion during gaze rotation revealed large errors when the rotation was not accompanied by an appropriate extra-retinal, eye-velocity signal [1,5,15]. Observers perceived self-motion paths curving in the direction of the rotation and so their responses were displaced in the direction of the simulated gaze rotation. The distance to the response marker was not specified in these experiments because the marker was viewed monocularly and was presented at the horizontal midline of the display.

We showed in Experiment 1 of this paper that the perceived distance of the response marker strongly affects the path judgement. The older monocular experiments revealed large judgement biases during simulated gaze rotations, so observers presumably perceived the response markers as rather distant. One can estimate this distance by reference to the results of Experiment 1. Consider, for example, Experiment 4 of Royden et al. [1]. The displays in this experiment mimicked linear translation plus gaze rotation across a ground plane (translation speed = 190 cm/s; observation height = 160 cm). At rotation rates of $\pm 5°$/s, translation errors were roughly $\pm 17°$ for observer MSB. This same observer yielded translation judgement errors of about $\pm 11°$ when the marker distance was 2000 cm. Therefore, this observer was behaving in the previous experiments as if the marker was at a distance greater than 2000 cm; this is consistent with the fact that the marker was positioned above the horizon in the display and, therefore, appeared quite distant. Naturally, this perceived distance may vary significantly from one observer to the next when the distance is not clearly specified much as the specific distance tendency [26] varies among individuals.

8.3. Should adding depth information improve performance?

In the introduction, we argued that depth information (other than that contained in the flow field) could in principle aid the estimation of self-motion paths

during gaze rotations from retinal-image information alone. In particular, an observer could use depth information to locate the most distant texture elements in the display; the motions of those elements are determined primarily by the gaze rotation, so from them one can estimate the rotation and subtract it from the observed flow in order to estimate the translation more accurately [8]. However, motion of the distant elements is also consistent with a curved path. Therefore, if no assumption is made about the linearity of the path, the stimulus is ambiguous (see discussion below). Alternatively, an observer could use depth information to locate the nearest texture elements; their motions are determined primarily by the observer's translation, so from the motions one could estimate the direction of self-motion. This argument is reasonable for the estimation of self-motion direction in noisy displays. To be specific, it would be useful in noisy displays to locate the texture elements whose motions are determined primarily by the observer's translation and to use the motion of those elements to estimate the direction of translation. In fact, van den Berg and Brenner [2] demonstrated that the addition of stereoscopic depth cues (with 3D-cloud stimuli for which element depths are not specified by position on the screen) enabled less variable direction estimates at low signal-to-noise ratios.

In our experiments with noise-free displays, however, observers' responses during simulated gaze rotations were characterised by bias, not variability; specifically, observers perceived curved self-motion paths rather than the depicted linear paths and responded accordingly. Apparently, in the absence of an appropriate extra-retinal signal, linear paths with path-independent rotation are mistaken for curved paths with path-dependent rotation as schematised in Fig. 1 [17].

What kind of estimation error would yield such a misperception? At each instant, there is a circular self-motion path (with gaze direction fixed with respect to the path) that gives rise to the same velocity field at the retina as a linear self-motion path plus gaze rotation [1,17,20,22]. The velocity fields arising from the two types of self-motion diverge over time; the field associated with linear self-motion plus gaze rotation is consistent with different circular paths at different instants, so the trajectories of features in the retinal image become less consistent with a unique circular path as duration increases. To put it another way, linear and circular paths differ in their acceleration and higher-order components; thus, in the absence of extra-retinal signals, they can only be discriminated on the basis of higher-order temporal derivatives of the retinal velocity field [22,27]. If observers failed to utilise the changes in velocity fields over time, the two types of self-motion might well be mistaken for one another. There is in fact considerable experimental evidence that human observers do not measure higher-order temporal derivatives of velocity accurately [28]. The ability to discriminate accelerations-temporal changes in speed and direction-is poor [29] and information in accelerations is not used efficiently in structure-from-motion tasks [30]. Thus, the bias to interpret linear translation with simulated gaze rotation as a translation along a curved path may result from an inability to integrate optic flow information over the longer time intervals necessary for accurate estimation of higher-order temporal motion derivatives.

To examine the temporal evolution of the velocity fields for linear paths with path-independent rotation and circular paths with path-dependent rotations, we calculated the image velocities for the two sorts of self-motion at various times. The stimulus and motion parameters were those of the ground-plane versions of Experiments 1–3. A point in the scene projects through the optical center onto an image plane with coordinates (x, y) where $x = -X/Z$ and $y = 1.6/Z$. The motion of the eye for the situations considered here is described by the instantaneous motion parameters T_X, T_Z, and R_Y.

Consider the situation in which the observer is moving forward on a linear self-motion path while making a horizontal eye movement (left panels of Fig. 1). This motion is described by two parameters—T_{Z_0} and R_Y—where T_{Z_0} is the initial translation. Following Royden [17], the coordinates of a point relative to the origin are:

$$\begin{aligned} X(t) &= X_0 \cdot \cos(R_Y \cdot t) + (t \cdot T_{Z_0} - Z_0) \cdot \sin(R_Y \cdot t) \\ Y(t) &= -1.6 \\ Z(t) &= (Z_0 - t \cdot T_{Z_0}) \cdot \cos(R_Y \cdot t) + X_0 \sin(R_Y \cdot t) \end{aligned} \tag{2}$$

where X_0 and Z_0 are initial coordinates. Projecting the point onto the image plane and then differentiating with respect to time yields the horizontal and vertical image velocities:

$$\begin{aligned} V_{x(t)} &= \\ &\frac{T_{Z_0} \cdot \sin(R_Y \cdot t) - R_Y \cdot Z(t) + x(t) \cdot [R_Y \cdot X(t) - T_{Z_0} \cdot \cos(R_y \cdot t)]}{Z(t)} \\ V_{y(t)} &= \frac{y(t) \cdot [R_Y \cdot X(t) - T_{Z_0} \cdot \cos(R_Y \cdot t)]}{Z(t)} \end{aligned} \tag{3}$$

Now consider the situation in which the observer is moving on a circular path while holding the direction fixed with respect to the path. This motion is described by two parameters: T_Z and R_Y. We do not use the subscripted version of the translation parameter because the translation is fixed with respect to the coordinates for this type of self-motion. Following Royden [17], the coordinates of a point in the scene are:

$$\begin{aligned} X(t) &= \left[X_0 - \frac{T_Z}{R_Y} \right] \cdot \cos(R_Y \cdot t) - Z_0 \cdot \sin(R_Y \cdot t) + \frac{T_Z}{R_Y} \\ Y(t) &= -1.6 \\ Z(t) &= Z_0 \cdot \cos(R_Y \cdot t) + \left[X_0 - \frac{T_Z}{R_Y} \right] \cdot \sin(R_Y \cdot t) \end{aligned} \tag{4}$$

Projecting the point onto the image plane and differentiating with respect to time yields:

$$V_{x(t)} = \frac{R_Y \cdot Z(t) + x(t) \cdot [T_Z - R_Y \cdot X(t)]}{Z(t)}$$

$$V_{y(t)} = \frac{y(t) \cdot [T_Z - R_Y \cdot X(t)]}{Z(t)} \quad (5)$$

From these equations, we can calculate the retinal velocity fields at different points in time for linear (Eq. (3)) and circular paths (Eq. (5)). The circular path chosen for the duration of the motion sequence was the one yielding the same velocity field as the linear path at the beginning of the motion sequence [17]; as we said earlier, the velocity fields associated with this particular circular path and the depicted linear path diverge over time. Image velocities were calculated at various times after the beginning of the motion sequence. The image velocity vectors associated with the same point in the scene were subtracted from one another (linear minus circular path). The magnitude of the resulting difference vector was then divided by the magnitude of the velocity vector from the linear-path condition in order to determine the proportional difference in velocity between the two depicted self-motions.

We found that the magnitude of the largest velocity differences increases monotonically with increasing rotation rate and increasing time. The effect of time is the most interesting. For a simulated gaze rotation of 5 deg/s, the largest difference in the visual field is 8% at 100 ms, 12% at 200 ms, and 43% at 800 ms[12]. The spatial distribution of proportional differences is fairly uniform, so added depth information could not help the observer locate the largest differences.

Psychophysical measurements have shown that the just-discriminable change in velocity (acceleration) is roughly 30% of the base velocity for base velocities greater than ~ 1 deg/s [31]. None of the changes in velocity, therefore, reach suprathreshold levels during the first 400 ms in the stimuli considered here. Thus, if the cause of the bias we observe is the visual system's insensitivity to changes in velocity over such time intervals, it is unclear why the addition of independent depth information would help. For this reason, it is perhaps not unexpected nor surprising that the addition of stereoscopic and a variety of monocular depth cues does not decrease the bias in path judgements during simulated gaze rotations.

It is conceivable that adding depth information would reduce response variability, but, as described in Section 5.2, we did not observe such a reduction with dichoptic as compared with monocular viewing. Thus, we find no evidence for a contribution of independent depth information in noise-free displays.

We hasten to point out that added depth information has been shown to affect the perception of self-motion paths in other tasks and viewing situations. Of course, van den Berg and Brenner [2] have shown that the addition of stereoscopic cues aids path estimation from noisy displays. Additionally, Beusmans and Richards [32] and Beer et al. [33] have shown that misleading depth information can alter path percepts. Thus, our conclusion applies only to the particular perceptual bias that occurs with noise-free displays and during simulated gaze rotations.

8.4. Why do observers exhibit bias rather than uncertainty?

Observers have different percepts when presented displays depicting linear self-motion paths plus real versus simulated gaze rotations. During real gaze rotations (e.g. during a tracking eye movement), they generally perceive linear paths whereas during simulated rotations, they generally perceive curved paths. By the argument in the section above, the retinal images associated with the two sorts of self-motion are easily mistaken for one another. Why then do observers have different percepts during real and simulated gaze rotations (e.g. different biases in their judgements) rather than high uncertainty (e.g. high variability)?

We speculate that the two types of self-motion are distinguished in normal circumstances by the complex of retinal and extra-retinal signals that arise. For example, consider the driver of an automobile. When the automobile is on a linear path and the driver's gaze rotates to track a moving or stationary object off to the side, efferent or afferent signals associated with the extra-ocular and neck muscles inform the nervous system about the gaze rotation. If the extra-retinal signals suggest that the gaze rotation matches the rotational flow estimated from the retinal image [8], then that flow can be confidently attributed to a path-independent rotation. When the automobile is negotiating a curve and is on a circular path, drivers typically fixate the inside edge of the road ahead [34], so gaze rotates relative to the world, but not relative to the automobile, body, or head; the signals associated with the extra-ocular and neck muscles indicate that no path-independent rotation has occurred. We speculate that the nervous system interprets this complex of retinal and extra-retinal signals as the result of a curved self-motion path with path-dependent rotation; it does so because curved-path motion would in most everyday situations yield this particular complex of signals.

[12] Table 1 of Royden [17] presents the differences in final image point positions for linear and circular paths. These calculations were done for three types of scenes-a ground plane, two frontal planes, and a 3D cloud-and for two points that are initially straight ahead at different distances. Although Royden's calculations are relevant to the issues under examination here, we assume that the observer must discriminate changes in velocity and not changes in final position.

8.5. Summary

In summary, we find that human observers misperceive the specified self-motion path when presented a stimulus depicting linear translation plus simulated gaze rotation. Specifically, they attribute a path-independent rotation, unaccompanied by an appropriate extra-retinal, eye-velocity signal, to a path-dependent rotation (see Fig. 1) as predicted by the circular-path hypothesis [17,25]. Response errors are smaller than predicted, but this may be due to a foveal bias and was also observed with real circular path stimuli. Most importantly, we find that adding depth information to the display by including the cues of binocular disparity, relative and changing size, occlusion, dynamic occlusion, and linear perspective does not yield more accurate performance. Even with those cues present, observers still perceive curved self-motion paths during simulated gaze rotations and, therefore, exhibit biases in their responses.

Our findings are consistent with the hypothesis that the perceived self-motion path is strongly influenced by the extra-retinal signals normally associated with tracking eye movements. As Royden et al. [1,5] have argued before, the erroneous percepts reported here are probably uncommon during everyday perception because extra-retinal, gaze-velocity signals are generally appropriate for human movements during walking, running, driving, and flying.

Acknowledgements

We thank Bill Warren, Bert van den Berg, and Lee Stone for helpful discussion and Karsten Weber for assistance in programming the displays. This research was supported by AFOSR Research Grant 93NL366, NSF Research Grant DBS-9309820, and NSF Training Grant GER-9355034.

References

[1] Royden CS, Crowell JA, Banks MS. Estimating heading during eye movements. Vis Res 1994;34:3197–214.
[2] van den Berg AV, Brenner E. Why two eyes are better than one for judgements of heading. Nature 1994;371:700–2.
[3] Gibson JJ. The Senses Considered as Perceptual Systems. Boston: Houghton Mifflin, 1966.
[4] Regan D, Beverley KI. How do we avoid confusing the direction we are looking and the direction we are moving? Science 1982;215:194–6.
[5] Royden CS, Banks MS, Crowell JA. The perception of heading during eye movements. Nature 1992;360:583–5.
[6] Warren WH, Hannon DJ. Direction of self-motion is perceived from optical flow. Nature 1988;336:162–3.
[7] Warren WH, Hannon DJ. Eye movements and optical flow. J Opt Soc Am A 1990;7:160–9.
[8] Longuet-Higgins HC, Prazdny K. The interpretation of a moving retinal image. Proc R Soc Lon B 1980;208:385–97.
[9] Rieger JH, Lawton DT. Processing differential image motion. J Opt Soc Am A 1985;2:354–60.
[10] Hildreth EC. Recovering heading for visually-guided navigation. Vis Res 1992;32:1177–92.
[11] Perrone JA, Stone LS. A model of self-motion estimation within primate extrastriate visual cortex. Vis Res 1994;34:2917–38.
[12] von Holst E. Relations between the central nervous system and the peripheral organs. Br J Anim Behav 1954;2:89–94.
[13] Matin L. Visual Localization and Eye Movements. In: Wertheim AH, Wagenaar WA, Leibowitz HW, editors. Tutorials on Motion Perception. New York: Plenum Press, 1982.
[14] Crowell JA, Banks MS, Shenoy KV, Andersen RA. Perception of self-motion during head and body rotations. Nature 1997 (submitted).
[15] Banks MS, Ehrlich SM, Backus BT, Crowell JA. Estimating heading during real and simulated eye movements. Vis Res 1996;36:431–43.
[16] Rieger JH, Toet L. Human visual navigation in the presence of 3-D rotations. Biol Cybern 1985;52:377–81.
[17] Royden CS. Analysis of misperceived observer motion during simulated eye rotations. Vis Res 1994;34:3215–22.
[18] van den Berg AV. Robustness of perception of heading from optic flow. Vis Res 1992;32:1285–96.
[19] van den Berg AV. Perception of heading. Nature 1993;365:497–8.
[20] van den Berg AV. Judgements of heading. Vis Res 1996;36:2337–50.
[21] van den Berg AV, Brenner E. Humans combine the optic flow with static depth cues for robust perception of heading. Vis Res 1994b;34:2153–67.
[22] Warren WH, Mestre DR, Blackwell AW, Morris MW. Perception of circular heading from optical flow. J Exp Psychol: Hum Percept Perform 1991;17:28–43.
[23] Koenderink JJ, van Doorn AJ. Facts on optic flow. Biol Cybern 1987;56:247–54.
[24] Stone LS, Perrone JA. Translation and rotation trade off in human visual heading estimation. Invest Ophthalmol Vis Sci (Suppl) 1996;37:515.
[25] Crowell JA, Banks MS. Ideal observer for heading judgments with rotations. Invest Ophthalmol Vis Sci (Suppl) 1994;35:2000.
[26] Gogel WC. The sensing of retinal size. Vis Res 1969;9:1079–94.
[27] Rieger JH. Information in optical flows induced by curved paths of observation. J Opt Soc Am 1983;73(3):339–44.
[28] De Bruyn B, Orban GA. Human velocity and direction discrimination measured with random dot patterns. Vis Res 1988;28:1323–35.
[29] Werkhoven P, Snippe HP, Toet A. Visual processing of optic acceleration. Vis Res 1992;32:2313–29.
[30] Hogervorst MA. Limitations on the recovery of structure from motion. Thesis, Universiteit Utrecht, The Netherlands 1996.
[31] Snowden RJ, Braddick OJ. The temporal integration and resolution of velocity signals. Vis Res 1991;31:907–14.
[32] Beusmans JMH, Richards WA. Object-based perception of egomotion. Perception (suppl) 1994;23:19.
[33] Beer JMA, Gallaway RA, Previc FH. Scene cues can overcome image motion to influence locomotion control, not just passive heading. Perception (suppl) 1996;25:130.
[34] Land MF, Lee DN. Where do we look when we steer? Nature 1994;369:742–4.
[35] Vishton PM, Nijhawan R, Cutting JE. Moving observers utilize static depth cues in determining their direction of motion. Invest Ophthalmol Vis Sci (Suppl) 1994;35:2000.