



# Humans Combine the Optic Flow with Static Depth Cues for Robust Perception of Heading

A. V. VAN DEN BERG,\*† E. BRENNER\*

Received 27 August 1993; in revised form 13 December 1993

The retinal flow during normal locomotion contains components due to rotation and translation of the observer. The translatory part of the flow-pattern is informative of heading, because it radiates outward from the direction of heading. However, it is not directly accessible from the retinal flow. Nevertheless, humans can perceive their direction of heading from the compound retinal flow without need for extra-retinal signals that indicate the rotation. Two classes of models have been proposed to explain the visual decomposition of the retinal flow into its constituent parts. One type relies on local operations to remove the rotational part of the flow field. The other type explicitly determines the direction and magnitude of the rotation from the global retinal flow, for subsequent removal. According to the former model, nearby points are most reliable for estimating one's heading. In the latter type of model the quality of the heading estimate depends on the accuracy with which the ego-rotation is determined and is therefore most reliable when based on the most distant points. We report that subjects underestimate the eccentricity of heading, relative to the fixated point in the ground plane, when the visible range of the ground plane is reduced. Moreover we find that in perception of heading, humans can tolerate more noise than the optimal observer (in the least squares sense) would do if only using optic flow. The latter finding argues against both schemes because ultimately both classes of model are limited in their noise tolerance to that of the optimal observer, which uses all information available in the optic flow. Apparently humans use more information than is present in the optic flow. Both aspects of human performance are consistent with the use of static depth information in addition to the optic flow to select the most distant points. Processing of the flow of these selected points provides the most reliable estimate of the ego-rotation. Subsequent estimates of the heading direction, obtained from the translatory component of the flow, are robust with respect to noise. In such a scheme heading estimates are subject to systematic errors, similar to those reported, if the most distant points are not much further away than the fixation point, because the ego-rotation is underestimated.

Optic flow   Heading perception   Noise   Heading bias   Optimal observer

## INTRODUCTION

In some cases we may have less trouble looking through another person's eye than sharing his point of view. In films we can perceive the motion of the actor, whose view the camera is showing us, as our own, although we remain aware of the flat screen we are looking at. The changing perspective of a moving vantage point (the optic flow) provides a compelling impression of depth and motion through space. Gibson (1966) was the first to present a theory of depth and ego-motion as perceived from the optic flow. He pointed out that for linear travel the optic flow takes a radial structure, emanating from the focus of outflow. In that case the focus of outflow coincides with the direction of heading. However, moving around, we usually look at an object in the environ-

ment, keeping its image on the fovea by rotation of our eyes, head and body. The retinal flow is therefore more complicated than for pure translation of the vantage point; the retinal motion vectors roughly converge on the fixation point. In order to perceive the direction of heading from the retinal flow, we should be able to disregard the component that is caused by ego-rotation. Indeed, humans are able to perceive their direction of motion across the ground with high precision (about 1.5 deg) when the retinal flow contains components due to ego-rotation and ego-translation (Warren & Hannon, 1988, 1990; van den Berg, 1992; Cutting, Springer, Braren & Johnson, 1992).

For self-generated rotations, the visual stimulus is accompanied by proprioceptive or efference copy signals. These extra-retinal signals are useful but not always necessary for discounting the rotational part of the flow-field. Perceived heading is equally accurate when subjects make a real eye movement to fixate a point in the ground plane, as when the eye rotation is simulated

\*Physiology I, Erasmus University Rotterdam, Faculty of Medicine,  
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

†To whom all correspondence should be addressed.

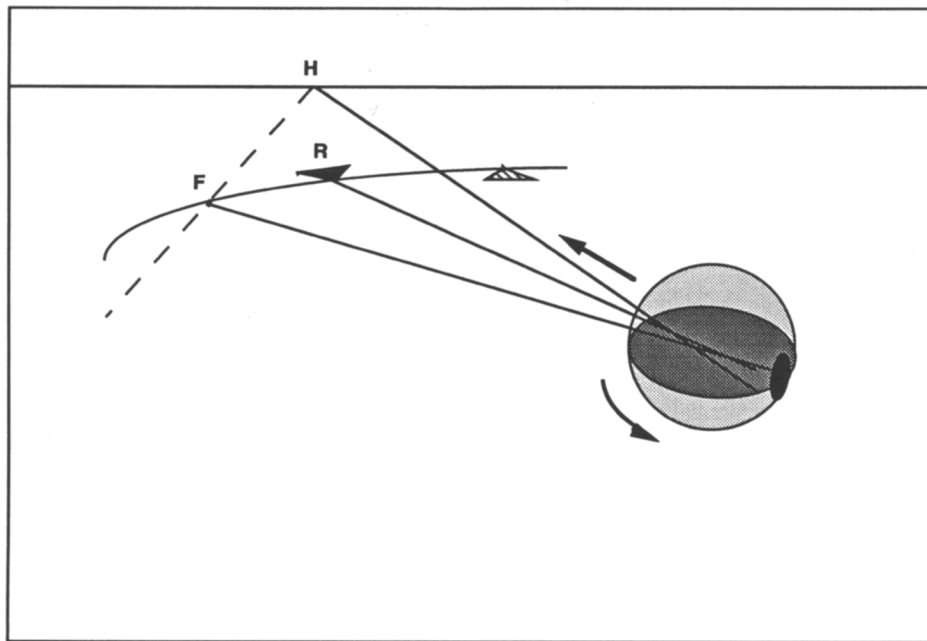


FIGURE 1. The simulated motion of the vantage point. The horizontal translation towards the point H is combined with a rotation to fixate the point F in the plane (arrows). At the end of the motion sequence two triangles are drawn in the ground plane at the same distance as the fixation point. One is fixed, indicating the reference direction (R) for the discrimination task. The other (shaded) is moved by the subject along a circle to point in the perceived direction of heading. The dashed line indicates a strip of locations in the ground plane at which the motion vectors are exactly aligned.

in the display and presented to the stationary eye. This has been shown for relatively slow ego-rotations [1–3 deg/sec (Warren & Hannon, 1990)] as well as for faster eye rotations [up to 6 deg/sec (van den Berg, 1992, 1993)]. Thus, humans can *visually disregard* the rotation and need not rely on extra-retinal signals. On the other hand, visual decomposition of the flow-field into translatory and rotatory parts has a number of limitations that are at least partially overcome when extra-retinal signals are present. Royden, Banks and Crowell (1992), examined perceived heading during smooth pursuit of a moving object that was not a part of the ground plane. They reported that subjects could not accurately per-

ceive heading for the faster eye rotations (2–5 deg/sec) unless a real eye movement was made. This might mean that for visual decomposition fixation of a point in the rigid environment is required (van den Berg, 1993). Visual decomposition is in some cases also difficult when points of the rigid environment are fixated. For example, when the scene contains limited depth, as in motion towards a fronto-parallel plane, subjects cannot distinguish the direction of heading from the direction in which they are looking (Regan & Beverly, 1982; Rieger & Toet, 1985; Warren & Hannon, 1990), unless a real eye movement is made (Warren & Hannon, 1990). Also, visual decomposition prefers an environment with dots

FIGURE 2 (*facing page*). (a) Data for one subject (RG). Each symbol in the graph represents the pointing response on one trial. Indicated in the figure are the perceived and simulated heading *relative to the fixation point*. The perceived heading was a linear function of the simulated heading. For a final depth range of 35 m, perceived heading was close to the simulated heading (solid triangles). The slope of the perceived heading was lower (circles) when the horizon was nearby. For a final depth range of 7 m, RG's perceived heading was biased towards the fixation point by about 15% of the simulated heading. (b) The slope of the linear relation between the perceived heading and the simulated heading for the nearby horizon (7 m) is plotted against the corresponding slope for the distant horizon (35 m) for each subject (AB, EB, RG, RR, HS). Horizontal error bars indicate the SE for the slope when the horizon is at 35 m. Vertical error bars indicate the corresponding error for the horizon at 7 m. If pointing were perfect, i.e. if perceived heading were equal to the simulated heading irrespective of the location of the horizon, all data should lie within the square drawn at (1, 1). If the perceived heading is not perfect, but does not depend on the distance of the horizon, the slopes for the two conditions should be identical and the points should lie on the diagonal line. This is the case for subject EB only. Data points are below the diagonal line in all other subjects. In these subjects the slopes decrease when the depth range decreases. (c) Comparison of the two measures of heading precision. Data of five subjects are shown. Each point is based on data collected in one subject for one depth range. The two depth ranges are represented by different symbols. The discrimination threshold obtained from the staircase procedure is indicated on the ordinate. The standard deviation of the residual errors is indicated on the abscissa. The residual error is computed as the difference between the perceived heading in a trial and the perceived heading predicted on the basis of the best fitting linear function to all the pointing responses collected for the same depth range. The staircase finds the 75% discrimination level whereas the residual error corresponds to the 66% discrimination level. Also the bias has some effect on the former but not on the latter estimate. Nevertheless, the difference between the measures is small.

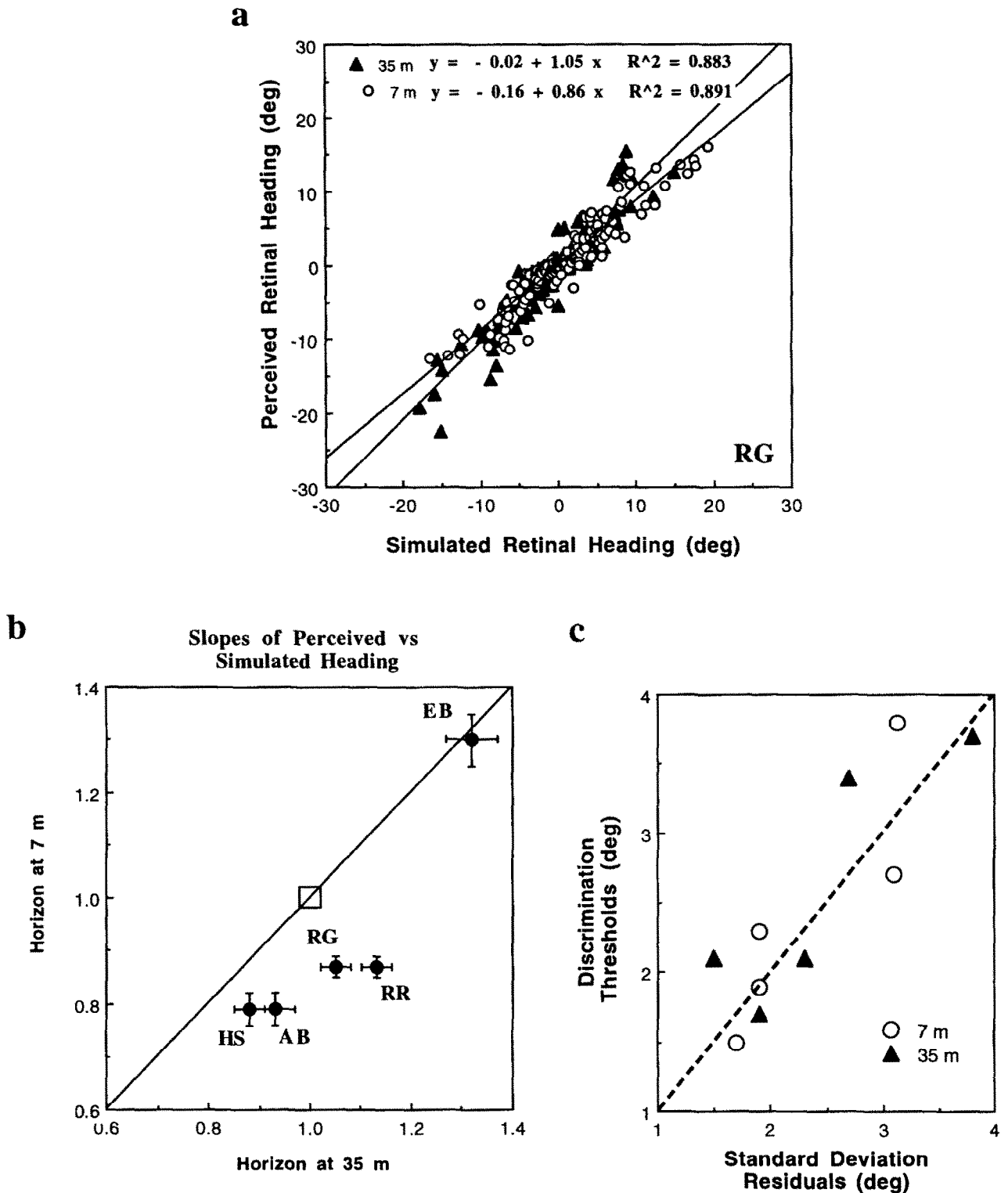


FIGURE 2. Caption opposite.

on continuous surfaces rather than in a random structure; when fixating a point of a cloud of randomly positioned dots heading accuracy is low for moderate to high simulated eye rotation (van den Berg, 1992; Royden *et al.*, 1993) although accurate performance has been reported for slow simulated eye rotation (Warren & Hannon, 1990). In contrast, when real eye movements are made, heading perception is accurate when moving

through a cloud of dots (Royden *et al.*, 1993). Finally, van den Berg (1992) showed that heading perception is more robust against the presence of noise when real eye movements are made than when the same eye rotation is simulated.

All these observations suggest that under natural conditions visual and extra-retinal mechanisms work in concert to perceive heading accurately and robustly.

However, the fact that humans can perceive their direction of heading when eye rotations are simulated in the display shows that extra-retinal information is not essential and that purely visual heading perception is possible within the limitations mentioned above. This study aims to further elucidate the purely visual contribution to heading perception.

Two types of model have been proposed for the visual decomposition of the flow-field into its translatory and rotatory parts. Rotation results in retinal slip that is independent of distance. Therefore, the difference in retinal velocity between two objects in adjacent viewing directions, but at different depth, depends only on the translatory component of the flow. The scheme by Longuet-Higgins and Prazdny (1981) uses such changes in local velocity at depth edges to obtain the radial structure that belongs to the translatory component in the retinal flow. Semi-local schemes (Rieger & Lawton, 1985; Hildreth, 1992) do not require depth edges, but determine the axis in which the spatial differences between neighbouring motion vectors is largest. That axis is directed approximately towards the focus of outflow. These schemes, then, do not determine the rotational component but eliminate it by local subtraction.

Perrone (1992) proposed a scheme inspired by neurophysiology. It computes the direction and magnitude of the eye rotation that best fits the retinal flow in the entire visual field. Subsequently, it subtracts for each location the estimated rotational component to obtain the translational flow. This *two-stage* scheme is based on a global estimate of ego-rotation, which could be obtained by a set of "templates" or visual channels that are sensitive to

the rotational part of the flow [as found in area MST of the monkey (Tanaka, Hikosaka, Saito, Yukiie, Fukada & Iwai, 1986; Duffy & Wurtz, 1991)].

Van den Berg (1992) argued that the presence of a horizon is important for discounting the ego-rotation. It could be an advantage to selectively use points near the horizon to estimate the direction and magnitude of the rotation, because for these points the translational component is minimal. This suggestion was inspired by two observations: (1) heading perception was much less robust against noise when dots were presented at random positions within a cloud, than when they formed a ground plane with a visible horizon, and (2) the point life-time required for determining one's heading direction corresponded with the processing time (120 msec) of local motion detectors that were tuned to the speed of the horizon (0.15 deg/sec), although the local speed was larger for most points in the display and detected by faster mechanisms.

Essentially, this hypothesis suggests that humans combine static depth information (from the nearness to the horizon, the texture gradient, perspective or any combination of these cues) with optic flow to obtain the heading direction. We have investigated two further implications of this hypothesis. First, if true, humans should make systematic errors in perceived heading when the distance to the furthest points in the display (the horizon) is reduced. This follows, because for nearby points the velocity is more contaminated by the translational component in the flow than for distant points. Secondly, combining independent sources of information can make heading perception more robust against noise. Thus, if the visual stimulus contains independent information on depth, humans may perform better than the optimal observer that uses only the optic flow. Our data confirmed both implications suggesting that humans combine static depth cues (and possibly knowledge that the visible landmarks are confined to a continuous planar surface) with optic flow to obtain a precise and robust heading percept.

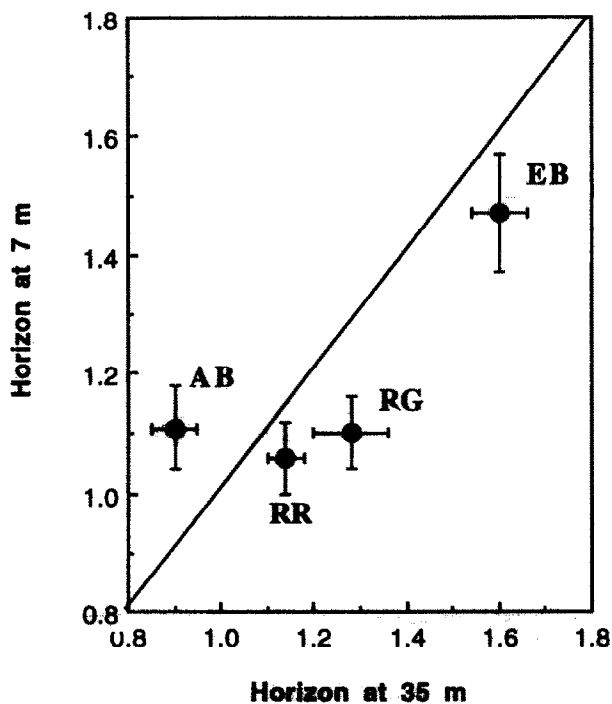


FIGURE 3. Slopes of the linear relation between the perceived and the simulated retinal heading direction when a real eye movement was made to fixate the target in the ground plane. For explanation see legend of Fig. 2(b).

#### GENERAL METHODS

Seven male subjects (the authors and five naive subjects), participated in the experiments. Two subjects were emmetropic (AF and HS). The five myopic subjects wore their normal corrective spectacles. The right eye of one of the authors (AB) had been surgically removed about 6 months prior to the experiments. All measurements were monocular with the left eye viewing. Subjects were seated in a dark room in front of a transparent projection screen (distance 2 m; horizontal  $\times$  vertical dimensions 60  $\times$  50 deg). Measurements started after about 10 min of dark adaptation. A chinrest and forehead support were used to fixate the subject's head. The motion sequence was shown by rear projection. Displays consisted of a field of 256 randomly positioned bright dots on a dark background. The simulated dot positions were defined by three-dimensional vectors constrained to lie in a horizontal plane. Observer motion was simulated

by changing the position of the vantage point which was located 1.3 m above the simulated ground plane. This height corresponded to the eye height of the subject above the floor when seated in front of the projection screen. Thus, the screen served as a window on an infinite plane that was continuous with the ground plane. Special graphics hardware enabled the computer (Silicon Graphics GTX-210) to perform the computations required for the perspective projection of these vectors, the anti-aliasing of the projected points and the clipping by the borders of the screen on a frame-to-frame basis (refresh rate 60 Hz). To check the projection, two vertical reference lines were drawn (1 m apart horizontally) in two fronto-parallel planes at simulated distances of 4 and 20 m from the observer. Their projected positions on the screen corresponded to the expected values to an accuracy of better than 1%.

In all experiments motion parallel to the horizontal plane was shown. Forward motion of the observer (3 m/sec) and the eye rotation that is required to fixate one red point in the plane, were simulated in the display. Subjects had to indicate the perceived horizontal direction of ego-motion relative to the plane. We did not distinguish between perception of self-motion and perception of objects on a horizontal plane moving below the feet.

Each trial started with presentation for 1 sec of the stationary scene followed by 1.67 sec of motion (100 frames). Subsequently, a green probe appeared, which could be moved by the subject. It remained on the simulated ground plane, at a fixed simulated distance from the observer: the final distance of the red fixation point. However, the precise position of the probe on this circle was determined by the subject so as to indicate the perceived direction of heading (Fig. 1). Thus, we only measured the horizontal component of perceived heading. Concurrently with these heading judgements we determined discrimination thresholds for the perceived heading direction using a method introduced by Warren, Morris and Kalish (1988). To this end, a red probe appeared on the same circle, indicating a reference direction. The subject's pointing response with the green probe was directly used to evaluate whether the subject had perceived his motion as right- or leftward with respect to the reference direction. The deviation of the simulated ego-motion was randomly left- or rightward with respect to the reference. This deviation or "heading offset", was changed in the course of the experiment according to a staircase method. On correct discrimination the heading offset was reduced by 0.25 deg. The bias angle was increased by 0.75 deg when the subject made an error. This procedure converges to the threshold angular difference at which 75% correct responses are given (Finney, 1965). Threshold was computed as the mean of the last 11 (out of 14 collected) reversal levels of the bias angle. At the outset of a staircase the heading-offset angle was set to 4 deg. Three thresholds were obtained for each condition. For some conditions the reversal levels did not reach a stable level but showed a steady increase. In

those cases thresholds were set at an arbitrary ceiling level (15 deg).

During the simulation a distance of 5 m was covered. Thus, depending on the direction of heading relative to the fixation point, the depth range was reduced by 4.7–5 m during the simulation. Also, the angle between the fixation point and the heading direction increased during the simulation. The corresponding average eye rotation varied in proportion to the eccentricity of the heading and ranged from about 0.1 to over 6 deg/sec. In the figures depth ranges and heading angles relative to the fixation direction (retinal heading) will refer to the final values of these quantities, because the pointing task was done at the end of the trial. We should remark that specification of the depth range (in m) by itself is meaningless. The optic flow only specifies distances relative to the speed of the ego-motion. Because the ego-speed was not varied in this study we preferred to specify depth ranges in simulated distances.

To eliminate the global rotation in the display as a cue to the direction of heading, the fixation direction was randomly chosen from a range that was not symmetrical with respect to the reference direction. In case of a symmetrical range, the fixation point would be more likely to occur leftward of the heading direction for a rightward heading-offset (i.e. reference occurs to the left of the direction of heading). Thus, for rightward heading-offset the display would be more likely to contain a global rotation of the image to the right. To remove this cue, the range of fixation directions was chosen to be 5 times larger than the heading-offset for that trial and this range was symmetrical with respect to the direction midway between the reference and the heading direction. This arrangement ensured that the global rotation of the image was only predictive of the heading-offset in 60% of the trials irrespective of the magnitude of the heading-offset. Thus the 75% correct threshold could only be reached on the basis of heading perception.

To rule out any possibility of discrimination on the basis of the motion of recognizable dot configurations the locations of the points were randomized on each new trial.

## THE EFFECT OF THE DEPTH RANGE ON THE PERCEIVED HEADING DIRECTION

### Methods

We varied the simulated depth range of the plane. Depth range was 12 or 40 m. The 256 bright points were randomly distributed in the rectangular (12 × 40 m or 40 × 40 m) ground area in front of the observer. Because only part of this area was visible on the screen, about 40 (small depth range) or 120 points (large depth range) were visible at each instant. At the outset of the trial the simulated horizontal distance of the red fixation point was 8 m. In a part of this experiment pure translation and no eye rotation was simulated. The subject then pursued the red point with an eye movement. For simulated eye rotation, the red point was presented at

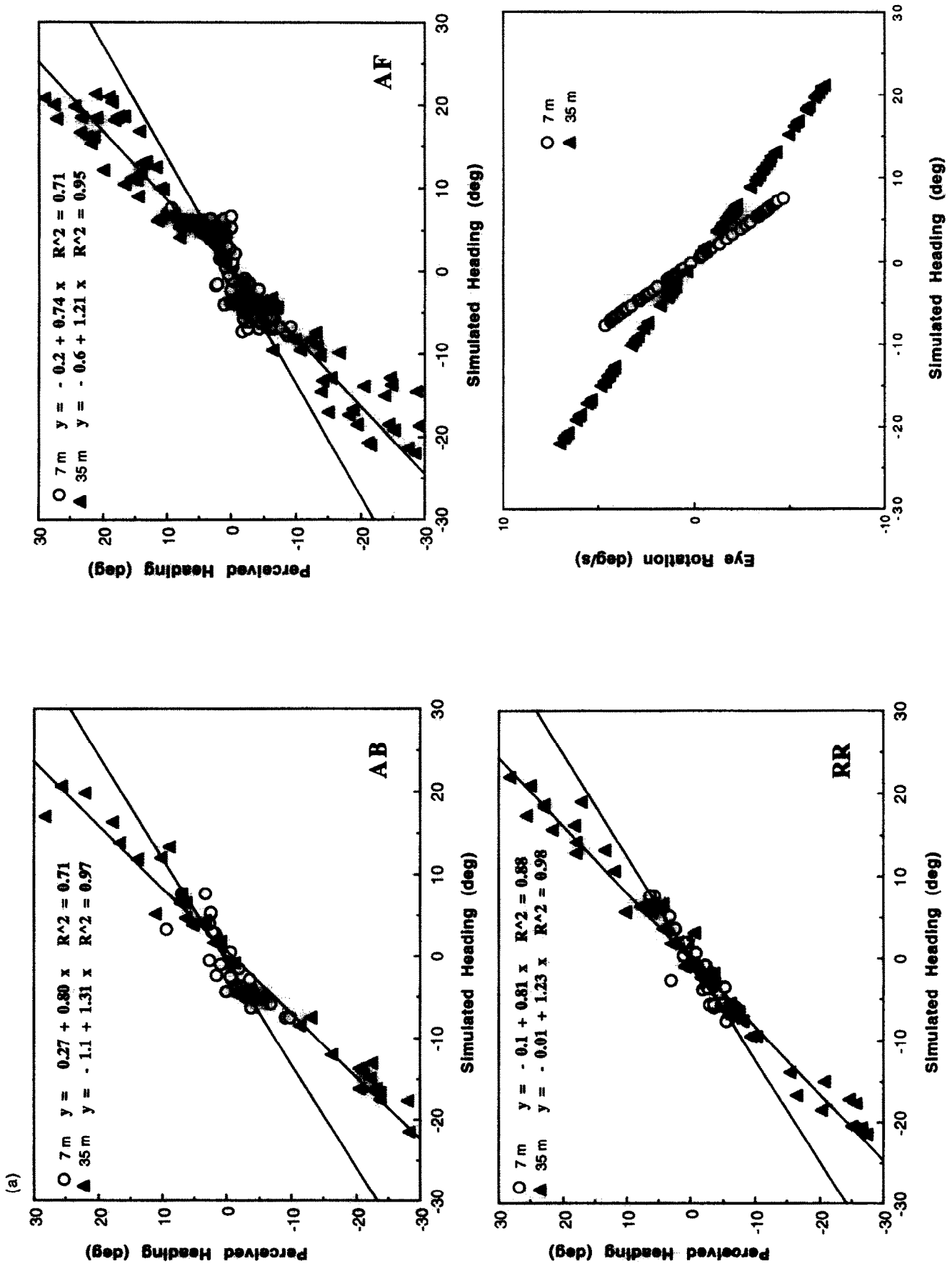


FIGURE 4(a). Caption on facing page.

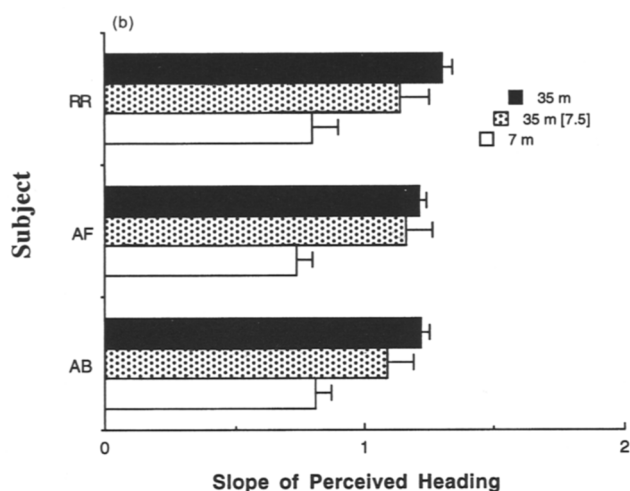


FIGURE 4. Perceived heading as a function of the simulated heading in subject AB, AF and RR for two depth ranges (circles, 7 m; solid triangles, 35 m). For explanation see Fig. 2(a). The fourth panel shows the average horizontal (simulated) eye rotation as a function of the retinal heading. Data points represent the actual simulated eye rotations that were presented to AF. (b) The slope of the relation between the perceived and the simulated retinal heading for the three subjects. Depth range was 35 m (solid bars) or 7 M (open bars). Error bars indicate 1 SE of the slope. The shaded bars indicate the slope for a depth range of 35 m, taking into account only those trials in which the simulated retinal heading was  $< 7.5$  deg.

the centre of the screen and was fixated by the subject with a stationary eye.

### Results and discussion

Most subjects (when asked or spontaneously) reported perceived motion of the fixation point, although the point was stationary on the screen throughout the simulation. This suggests that they perceived the translatory component of the flow field and disregarded (at least partially) the rotational part of the flow field.

The simulated distance to the horizon did not affect the precision of the heading estimates but it changed heading accuracy. Because this bias was proportional to the retinal eccentricity of the heading we chose to describe the perceived and the simulated heading direction in the retinal reference frame; these directions will be denoted as perceived and simulated "retinal heading". Even for the distant horizon some bias was apparent. Some subjects tended to underestimate the retinal eccentricity (AB, HS) of the heading direction and others overestimated the eccentricity (EB, RR, RG). When the horizon was located at 7 m, the perceived heading was biased towards the fixation point in four out of five subjects.

This bias was not constant but rather increased linearly with the simulated retinal eccentricity of heading. For example, the naive and unexperienced subject RG pointed almost in the direction of simulated heading, irrespective of its retinal eccentricity, when the horizon was located at 35 m [Fig. 2(a), circles]. For the horizon at 7 m, perceived heading was biased towards the fixation point by about 15% of the simulated

heading relative to the fovea. This is visible as the change of the slope of the regression line relating perceived to simulated retinal heading. Whereas this slope is slightly larger than one for the extended depth range, it is about 0.86 for the reduced depth range [Fig. 2(a), triangles]. Figure 2(b) shows that all subjects, except one (EB), showed a similar reduction in the slope of the regression lines (10–20%) for the reduced depth range. Subject EB was exceptional because he overestimated the retinal eccentricity of the heading irrespective of the simulated depth range or the retinal heading. We do not know the reason for his bias. In recent experiments we have observed that for the distant horizon his bias is much reduced when the simulated ego-motion is presented stereoscopically.

The thresholds for heading perception ranged from 1.4 to 3.8 deg depending on the subject and the depth range. The residual errors of the perceived heading relative to the best fitting slope were computed for each subject and depth range. This gives an independent estimate of the precision of heading. The discrimination thresholds and the SD of the residual errors corresponded to one another with a margin of about 1 deg as shown in Fig. 2(c). Warren *et al.* (1988) suggested that poor accuracy of heading perception in older reports was partly caused by the use of a pointing task. Measuring the precision of heading with a two-alternative forced-choice (2AFC) technique as introduced by Warren *et al.* (1988) sidesteps possible confounding effects of slack pointing. The close correspondence that we find between the standard deviation of the pointing responses and the simultaneously determined heading thresholds suggests that the gain in precision by using the 2AFC technique may be less than previously thought.

Interestingly, when the subjects made a real eye movement to fixate the red target, with other conditions similar to those in the first experiment, we found in three of the four subjects a shift of 5–15% of the perceived heading direction towards the fixation point. A shift in the opposite direction was found in the fourth subject (AB, Fig. 3).

We found a clear bias of the perceived heading towards the fixation point (for simulated eye rotation) when the simulated distance to the horizon was reduced. Was this really an effect of the removal of the flow vectors of the distant points? It might be argued that other factors than the reduction of the horizon's distance to the vantage point, caused the bias. For example, the vertical extent of the pattern of dots on the screen was smaller when the depth range was reduced. The difference amounted to about 4.5 deg at the start of the trial and increased to about 8.7 deg at the end of the trial. Also, the number of visible points was reduced from about 120 to 40 when the depth range was reduced. It is difficult to imagine a mechanism by which these differences would have caused the bias. Nevertheless, it seemed prudent to investigate whether the bias remains when these factors are removed.

## ON THE ORIGIN OF THE BIAS IN THE HEADING DIRECTION

### Methods

To control for these factors we repeated the experiment with simulated eye rotation, but with a reduced number of points for the large depth range (80 dots on the ground plane; about 40 visible on the screen). To remove the difference in vertical extent between the two conditions one requires that the vertical extent of the stimulus above the fixation point is identical (the fixation point is always presented at the centre of the screen). Thus, we used a fixation distance of 15 m for the distant horizon and the same fixation distance (8 m) as in the previous experiment for the nearby horizon. In both cases the vertical extent of the stimulus above the fixation point was now 3 deg. Note, however, that for the same retinal heading direction the average horizontal eye rotation was about two times slower when the distant horizon was presented. We compensated for this difference by presenting stimuli with a larger range of retinal heading directions for the distant horizon (for the horizon at 40 m the initial retinal heading ranged from  $-15$  to  $15$  deg; for the horizon at 12 m the initial retinal heading ranged from  $-3$  to  $3$  deg). Contrary to the previous experiment we did not collect threshold data, but only analysed the pointing responses. Therefore, the red reference target was not presented at the end of the trial. Also, the direction of heading was randomly chosen from either of the above mentioned intervals.

We noted that whereas the number of visible points on the screen was now comparable in the two conditions, that the distribution of the points on the screen was not. For the large depth range most projected points were located close to the projected horizon. For the small depth range the points were more evenly distributed on the screen. To minimize this difference we distributed the points non-randomly in the plane for the large depth range. Using the procedure described in Appendix A we obtained a distribution on the screen that closely mimicked the distribution obtained for the small depth range. Other conditions were as in the first experiment.

### Results and discussion

Data were collected for two subjects (AB and RR) that had also participated in the original experiment and one new naive subject (AF). In all subjects, the slope of the relation of the perceived vs simulated heading was smaller when the depth range was reduced [Fig. 4(a)]. The range of average horizontal (simulated) eye rotation was comparable in the two conditions [lower-right panel of Fig. 4(a)]. The slopes for the distant horizon were somewhat higher than in the original experiment. This may have been caused in part by the use of large eccentricities of heading for the distant horizon. When only the data in which the retinal eccentricity of the heading was less than  $7.5$  deg [Fig. 4(b), shaded bars] were analysed (i.e. identical to the range for the nearby horizon) the slopes were closer to 1, indicating that subjects pointed more accurately in these cases (this

analysis included about 30% of the data for the distant horizon). Yet, even for this small range of retinal heading directions, pointing was not biased towards the fixation point in the manner found for the nearby horizon.

These experiments were aimed to test if the most distant points in the display are used for estimation of the ego-rotation. Such a strategy would predict, that the perceived heading would show a bias towards the fixation point when the depth range is reduced. The flow velocity of the points contains a component due to translation, that partially compensates the rotational component. Because this translatory component is inversely proportional to the distance, the motions of the most distant points in the display are least "contaminated" by it. An estimate of the ego-rotation based on the motion of these points is the most reliable. It will, however, be too low, unless the points are located at infinity. As the "contamination" increases when the depth range decreases, the horizontal component of the ego-rotation will be increasingly underestimated when the depth range is reduced. Consequently, humans should underestimate the horizontal eccentricity of heading.

We think that the bias which we report argues against local schemes (Rieger & Lawton, 1985; Hildreth, 1992) of heading discrimination. Such schemes favour the nearby points (Hildreth, 1992) because the major velocity gradients are found at close range. We expect minor effects of removal of the distant points, if any, for such schemes, because it removes the information that is least reliable anyway. We stress that these predictions are based on our estimates concerning these models because, to our knowledge, the effects of the depth range on heading bias in these models has not been documented.

In line with our hypothesis we found that the perceived heading was biased towards the fixation point when the depth range was reduced. Confounding factors like the vertical extent of the display, or the number of points and their distribution on the screen, do not seem to have caused the pointing bias. One may object, though, that the effect that we report is quite small and does not occur in one subject [EB, Fig. 2(b)]. Also one may note, that in order to remove the confounding effects of the number of points and viewing area we had to introduce another confounding difference between the stimuli: the fixation distance. Consequently, for the same eccentricity of heading the simulated eye rotation was higher for the small depth range. We made the speed range of the horizontal component of the eye rotation similar in both conditions by increasing the range of simulated eccentricity of heading together with the depth range. The vertical component of the eye rotation remained higher for the small depth range. The large simulated vertical eye rotation by itself cannot explain the bias, because subjects showed no such bias for the same simulated eye rotation in the first experiment, when the large depth range was presented. This suggests, then, that the heading bias is caused by the removal of the most distant points. Nevertheless, we sought further



evidence for our hypothesis that heading perception is not exclusively based on optic flow, but uses additional information, such as depth from texture gradients, when available.

### THE EFFECT OF PERTURBATIONS OF THE LOCAL MOTION VECTORS ON HEADING THRESHOLDS

A *local* scheme for discounting the rotation component cannot profit from the noise reduction that is possible by spatial integration of the flow-field. Rieger and Lawton found that their algorithm was very susceptible to noise. Perturbation of all flow-vectors by 10% resulted in heading errors of about 10 deg. This level of perturbation is close to what may be expected for the internal noise of the visual system, because speed discrimination studies indicate a precision of velocity measurements of about 5% at best (McKee, 1981; DeBruyn & Orban, 1988). Using an improved algorithm Hildreth (1992) found better noise resistance when simulating motion across the ground plane.

The two-stage scheme derives the component of ego-rotation by combining the local flow-vectors across a considerable part of the visual field. This reduces its sensitivity to local disturbances of the flow-field. Moreover, if combined with static depth cues it allows for selection of the most reliable parts of the flow-field. The two schemes may possess quite different levels of noise tolerance. In this section we concentrate on the sensitivity of human heading perception to perturbations of the local-motion vectors of the flow-field.

In this experiment only simulated eye rotations were investigated. At the start of the trial the simulated distance to the fixation point was 16 m and the depth range was 40 m. Thresholds for discrimination of the heading direction were collected with the same methods as used in the first experiment. The authors (AB, EB) and three colleagues (RR, JK and AF), naive with respect to the purpose of the experiment but experienced in this task, served as subjects.

For the computation of the perturbations, the flow-field was treated as a collection of directions in which landmarks were located and associated flow-vectors (angular velocities) that specified the instantaneous retinal velocity of these points. Each flow vector was located in a plane perpendicular to the associated viewing direction. Thus, they specified the flow as projected on a spherical surface concentric with the nodal point of the eye.

We aimed to use noise that would not cause larger flow vectors to become more reliable for heading perception than smaller ones or vice versa. Each local flow-vector ( $\mathbf{p}$ ) was perturbed in proportion to its magnitude ( $|\mathbf{p}|$ ). This enabled us to characterize the noise level by a single number (the signal-to-noise ratio for speed) "SNR<sub>v</sub>",

$$\text{SNR}_v = |\mathbf{p}|/|\epsilon|$$

where  $|\epsilon|$  denotes the magnitude of the error component of the flow-vector ( $\epsilon$ ). For each local flow vector, SNR<sub>v</sub>

determined the error term's magnitude completely, but not its direction. Therefore we randomly chose the direction of the error term (without altering its magnitude) such that the error term was located in the plane perpendicular to the viewing direction towards that point (Appendix B). Prior to a trial the error terms were computed for all points in the plane and for all frames. Fixing the *relative* magnitude of the noise term makes the deviation in the direction of the perturbed flow vector (relative to the true flow vector) independent of the unperturbed flow-vector's magnitude. Also, the deviation in magnitude expressed as a constant multiple of the speed discrimination threshold is independent of the unperturbed flow-vector's magnitude. We think this is a more realistic measure for the difference in speed as experienced by the visual system than the physical speed difference.

SNR<sub>v</sub> was varied in octave steps ranging from 0.25 to 4. Data were collected in three sessions. Staircases for five different noise levels were run simultaneously, with the trials randomly interleaved.

It is well known that motion detection requires some processing time. This time depends on the tuning velocity of the motion detector (van Doorn & Koenderink, 1982; van de Grind, van Doorn & Koenderink, 1986; Mikami, Newsome & Wurtz, 1986). Temporal filtering may reduce the effectiveness of the noise if the perturbation varies on a time scale much shorter than the processing time of the local motion detector. Thus we chose a refresh interval for the noise (160 msec) that was higher than the processing time of the slowest motion detectors [120 msec for motion detectors tuned to 0.1 deg/sec (van Doorn & Koenderink, 1982)]. This precaution may not be sufficient though. Recruitment phenomena (McKee & Welch, 1985; van Doorn & Koenderink, 1984; Williams, Phillips & Sekuler, 1986) suggest that local motion detectors with the same motion preference in adjacent regions of the visual field interact cooperatively. If correct, humans may rely on the perceived paths of the points sampled over a time period extending well beyond the processing time of the single motion detector. When this integration is such that the direction of the error term is randomly changed several times within an integration period the noise is only partially effective.

We therefore used a limited point lifetime that equalled the refresh interval for the error term. Thus, each time the error term was changed, a new distribution of points in the plane was shown. This makes it impossible for subjects to use the motion path sampled over a long time period to obtain a temporally smoothed flow-field and the robustness should now reveal the effect of spatial integration. Results for this experiment are shown in Fig. 5.

For comparison with the theoretical analysis to be found below, we plotted the data as a function of the total noise level (SNR<sub>t</sub>). In human observers two independent sources of noise determine the performance on the heading task. First, the limited precision with which the human visual system encodes the flow-vectors and

second, the perturbations imposed on the flow-vectors in the stimulus. Thus, we cannot simply compare the thresholds of our subjects with those of the optimal observer (to be defined below) as a function of  $SNR_t$ , because then the noise that the humans receive is underestimated.  $SNR_t$  denotes the signal-to-noise ratio taking into account the noise contributions of the limited measurement accuracy of the visual system and the perturbations of the flow vectors:

$$SNR_t^{-2} = SNR_v^{-2} + SNR_i^{-2}$$

where  $SNR_i$  denotes the signal-to-noise ratio due to the velocity measurement error of the visual system.  $SNR_v$  was estimated at 10 (corresponding to the inverse of the Weber fraction for motion discrimination: 10%). The formula reflects the assumption that the internal and the external noise sources are independent and thus that their variances add linearly. (The formula is not entirely correct for high external noise levels because the internal noise is proportional to the perturbed and not to the unperturbed flow-vector. This results in underestimating  $SNR_t$  for high noise levels by a factor of  $SNR_v^{-2}$  or 1%.) Note that the optimal observer has no internal noise, because its measurement error is zero by definition. Consequently,  $SNR_t$  equals  $SNR_v$  for the optimal observer.

In all subjects we found a sharp increase of the discrimination threshold ( $\Delta\phi$ ) when  $SNR_t$  decreased below a critical level (Fig. 5). No subject could perceive the direction of heading at the lowest  $SNR_v$  (0.25). Except for subject JK, none of the naive subjects obtained stable thresholds at the lowest two levels of  $SNR_v$  (0.25 and 0.5).

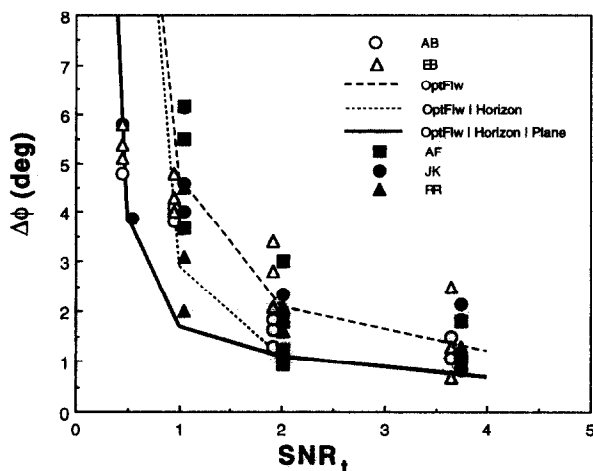


FIGURE 5. Heading thresholds ( $\Delta\phi$ ) of five experienced subjects as a function of the total signal-to-noise ratio ( $SNR_t$ ). For each condition and subject three thresholds are shown. For clarity, the data for the naive subjects (solid symbols) are shifted to the right by 0.05 units, whereas the data for the authors (open symbols) are shifted to the left by 0.05 units. For comparison, the thresholds of the optimal observer are also shown. The effect of the use of three sources of information by the optimal observer was investigated (optic flow, dashed line; optic flow and static depth cues and "shape knowledge", solid line).

## OPTIMAL OBSERVER ANALYSIS

To interpret our data, we compared human performance to that of the optimal observer. To this end we adopted the approach by Koenderink and van Doorn (1987). They analysed heading perception from the optic flow as a variational problem; that is, finding those ego-motion parameters (direction of heading and rotation) and distances to the points in the environment, that best fit the observed flow velocities in the least squares sense. Such an approach reveals the limits of information concerning depth structure and ego-motion that can be obtained from the optic flow. A simplified discontinuous description of the flow-field was used. It merely considered the directions and the motion vectors of the visible landmarks. This is appropriate, because in our experiment only the moving dots were visible, creating a discontinuous flow-field.

The unperturbed motion vectors are exactly known to the experimenter, because the simulated distances to the landmarks and the simulated ego-motion are exactly known. For the observer, the flow-vectors are known with finite precision due to the perturbations and measurement errors of the local velocity. For a given estimate of the ego-motion and lay-out of the environment, the unperturbed flow-vectors can be predicted. The optimal observer finds those distances and ego-motion parameters that minimize the sum of the squared differences between the *observed* motion vectors of the landmarks and the predicted motion vectors. The perturbations drive the solution away from the true solution because the optimal observer has no way of predicting the perturbations. The details of the algorithm can be found in Koenderink and van Doorn (1987). Briefly, initial estimates of the ego-rotation and ego-translation are obtained using the directions to the landmarks and the observed flow-vectors only. Using the estimated ego-motion, estimates of the distances are obtained from the observed flow-vectors and directions. These distance estimates in turn are used to refine the estimates of the ego-motion parameters. This sequence of computations is repeated until the sum of squared errors reaches an acceptable level. The criterion to stop the iteration is derived from the sum of squared errors for the true solution. (In our computations the estimate must be better than or equal to the least squares sum for the true solution; choosing a lower criterion level is not sensible because the true solution would be rejected.)

We used the same display and psychophysical procedures as used in the experiments. The input to the algorithm consisted of the direction vectors of the landmarks as seen from the simulated viewpoint and the corresponding flow-vectors. Only the flow-vectors of landmarks that were visible on the screen were used in the computation. These flow-vectors contained components due to the simulated ego-motion (the signal) and the error term (the noise) as defined above. From this input the algorithm computed the direction of heading and the ego-rotation. This computation was done for 10 frames spaced equally throughout the motion sequence.

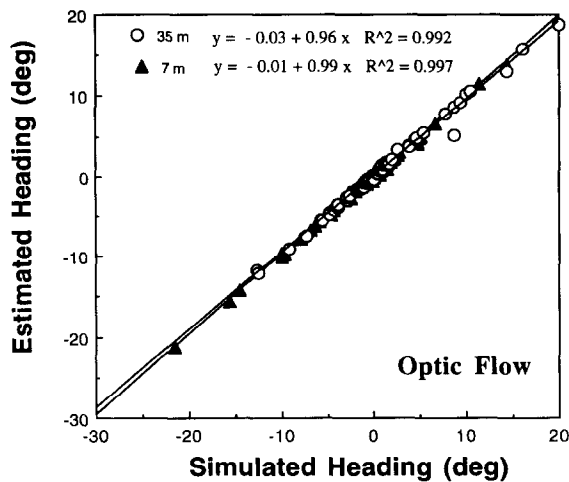


FIGURE 6. Retinal heading, as estimated by the optimal observer on the basis of the optic flow alone, did not depend on the simulated depth range. For both depth ranges the slope of the linear relation between estimated and simulated heading eccentricity was practically 1.

For each frame the horizontal heading error, defined as the signed difference between the true heading direction and the least squares estimate (projected on the ground plane) was computed. The horizontal heading error averaged across the 10 frames was added to the true heading direction in the last frame to obtain the "perceived heading" by the optimal observer in that trial. This was appropriate, because in both the experiments and the simulations, the refresh interval of the points divided each trial into 10 epochs with independent flow-fields (as far as the error terms are concerned). Thus the optimal observer and our human observers had a similar amount of information.

We also applied this analysis to the results of the first experiment (using the same procedures as were used to obtain the data of Fig. 2). In that experiment noise was not presented. For comparison with the optimal observer we therefore simulated the human measurement error by perturbation of the flow-vectors by 10% (SNR = 10). Figure 6 shows the estimated heading by the optimal observer as a function of the simulated heading in the same format as Fig. 2(a). The scatter in the estimated heading is less than in the human data. This might mean that the measurement error in humans is larger than 10% or that other sources of variation than those arising from the motion measurement affect performance. More importantly, there is no bias in the heading estimates of the optimal observer for either depth range. The difference in performance shows that the bias is not inherent to reduction of the information in the optic flow by reduction of the depth range. Rather, it reflects a property of the mechanism of human heading perception. This view was reinforced by the analysis of the last experiment. As shown in Fig. 5 heading discrimination thresholds of the optimal observer (dashed line) were often *worse* than for humans. Note, that 17 out of the 27 thresholds obtained for the three naive subjects were equal to or better than the thresholds for the optimal observer. Because the least-squares solution sets

the limit to the information that can be obtained from the optic flow by any algorithm, the difference shows that humans use *more* information than is contained in the optic flow.

We examined the effect on the characteristics of the optimal observer of two potential sources of information that humans could use in addition to the optic flow. First, as mentioned in the Introduction, the ground plane provides independent (static) depth cues that might be used to select the most distant points in the display for estimation of the ego-rotation. Thus, the optimal observer algorithm was modified in the following way. The ego-rotation was found using the initial estimate of the algorithm but now based on only those landmarks that were closer to the horizon than 3 deg. Subtracting this rotational component from the perturbed flow an estimate of the translatory flow was obtained. The direction of heading was estimated from the latter set of perturbed flow-vectors (the rotational component was set to zero throughout the optimization of the latter estimate).

Figure 7 shows the estimated heading direction as a function of the simulated heading direction in the, same format as in Fig. 2. Heading estimates by the algorithm were biased towards the fixation point by about 10% when the final depth range was 7 m. During the simulation the depth range decreased from 12 to 7 m. Thus, the magnitude of the bias that was to be expected from the nearness of the horizon increased continuously throughout the trial. When the estimates were based on the last 160 msec of the motion sequence the heading bias of the optimal observer increased to 20%. Practically no bias occurred for the distant horizon. Thus, the direction and the magnitude of the bias closely correspond to the human data. As shown in Fig. 5 (dotted

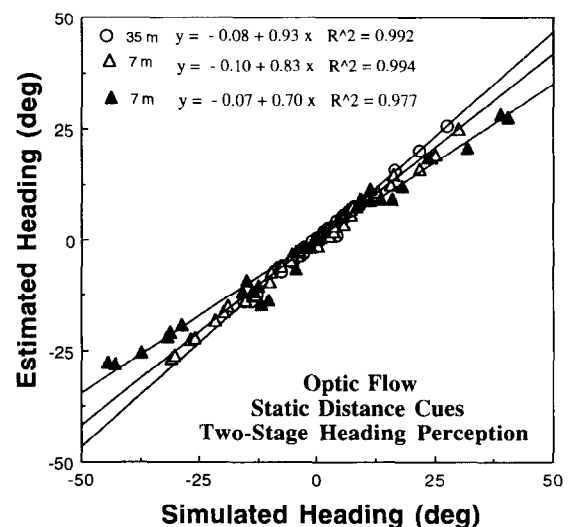


FIGURE 7. Retinal heading, as estimated by the optimal observer on the basis of the optic flow and static depth cues, did depend on the simulated depth range. For large depth ranges the slope of the linear relation between estimated and simulated retinal heading was practically 1 (circles). This slope decreased for the reduced depth range by 10% when estimates were based on the entire motion sequence (open triangles) and by 20% when the estimate was based on only the last 160 msec of the motion sequence (solid triangles).

line), heading thresholds became more precise when the static depth cues were used for estimation of the ego-rotation. However, thresholds for some subjects (AB, EB, JK) were still better at  $SNR_v = 0.5$  than for the optimal observer that used optic flow and independent depth cues, suggesting that humans might use even more information.

Knowing that the motion is across the ground plane provides strong constraints on the depth estimates of the landmarks. When this constraint was also implemented in the algorithm (Appendix B) the thresholds of the optimal observer using optic flow, static depth cues and shape knowledge were lower than the best human thresholds (Fig. 5, solid line).

## DISCUSSION

Previous experimental (van den Berg, 1992) and simulation (Perrone, 1992) studies have suggested that human heading perception in the presence of ego-rotations is based on a two-stage process involving (1) a global estimate of the ego-rotation from the compound flow-field followed by (2) a global estimate of the ego-translation from the flow-field with the component corresponding to the estimated ego-rotation subtracted. Van den Berg (1992) suggested that the estimate of the ego-rotation was based on the retinal motion of the furthest points in the display. As this selection precedes the perception of the environmental lay-out from the optic flow it should be based on independent information on depth. The present study provides two additional pieces of evidence to support this idea: (a) when the depth range is reduced, subjects underestimate the eccentricity of heading relative to the fixated point in the ground plane, and (b) human perception of heading is more robust against noise than heading estimates of an optimal observer, that uses optic flow only.

The bias of the perceived heading towards the fixation point was 10–20% of the simulated retinal heading when the depth range was reduced from 35 to 7 m. Because the number of visible points, their distribution on the screen and the vertical extent of the stimulus on the screen covaried with the depth range, either one or a combination of these factors might have caused the result. Note, however, that the optimal observer showed no such bias in the perceived heading when presented with the same stimuli. Thus, the optic flow was equally informative of heading for the two depth ranges. The bias did not disappear when the confounding differences were removed to as large an extent as possible.

The reported bias is to be expected if human heading perception follows the two stage scheme. Each flow-vector is consistent with an ego-rotation about an axis perpendicular to the plane subtended by that flow-vector ( $\mathbf{p}_i$ ) and the direction ( $\mathbf{d}_i$ ) towards the landmark to which the flow-vector belongs. The local estimate of the ego-rotation equals the vector-cross product of these quantities ( $\mathbf{d}_i \times \mathbf{p}_i$ ). Averaging these local estimates of the ego-rotation across all points along the horizon, and taking the vertical component we obtained an estimate

of the horizontal ego-rotation. The horizontal ego-rotation thus obtained is veridical for the distant (35 m) points but underestimated by about 25% when the far points are located at 7 m. Subtracting an underestimated rotatory component results in a flow-field that radiates outward from a point that is not coincident with the heading direction, but shifted towards the fixation point. Consistent with this analysis, the heading estimate of the optimal observer (using static depth information) showed a bias of 20% towards the fixation point, when only the last 10 frames were used (Fig. 7). In our displays the vantage point approached the far end of the plane. Thus, the percentage underestimation should increase throughout the trial. The somewhat lower underestimates in some of our subjects therefore indicates that they based the heading estimate on a longer interval than the last 10 frames.

One might consider an alternative explanation for the bias (M. Banks personal communication). Fixating a point to the side while moving forward generates retinal motion spiralling outward from the fixation point. In this pattern a strip of exactly aligned flow-vectors occurs that runs from the fixation point towards the heading direction. If our subjects identified the heading direction by taking the intersection of this strip with the horizon they would show a similar bias if the depth range was reduced. We believe this strategy to be unlikely at best. First, motion vectors are nearly aligned along strips with a different orientation in the fronto-parallel plane (resulting in a different intersection point with the horizon), which, combined with the limited accuracy of the local motion measurement, would make precise heading discrimination unlikely. Secondly, the perturbations in the second experiment disrupt this alignment considerably, making the task impossible. For example, the perturbed flow vectors deviate by as much as 45 deg at  $SNR_v = 1$ . Yet, heading discrimination amounted to 3–5 deg in most subjects at this same noise level.

The idea that ego-rotation is estimated using the furthest points in the display is consistent with a number of studies. Brandt, Wist and Dichgans (1975) observed that circular vection is induced when the moving surround is presented in the background and absent when it is presented in the foreground. Ohmi, Howard and Landolt (1987) reported that when a moving and a stationary surround were presented at different depths and viewed monocularly, the moving part evoked circular vection when it was perceived as the most distant. Despite the presence of unambiguous (although possibly weak) depth cues, spontaneous depth reversals and concomitant changes in vection occurred. Similarly, linear vection evoked by a looming display was inhibited by stationary points when their apparent distances were larger than those of the looming points, but not when their apparent distances were smaller (Ohmi & Howard, 1988). Thus, vection depends on the motion of the most distant surface.

Sudden changes in retinal background velocity during pursuit of a moving target induce changes in the perceived target velocity, unless the target velocity is also

changed so as to maintain the relative velocity (Brenner, 1991). The use of relative motion promotes the constancy of the perceived target motion in the presence of ego-motion. In that study, the background was a fronto-parallel plane surrounding the target. In recent experiments the target moved across a ground-plane that was bound by a distant wall. Simulation of ego-translation parallel to the wall resulted in retinal motion of the background that changed in accordance with the simulated distance. Therefore, relative motion could take many different values. For several simulated target distances no change in target motion was perceived when the changes in ego-motion and target motion maintained the target's velocity *relative to the distant wall* (Brenner & van den Berg, 1993). The latter result is again suggestive of humans using motion of the most distant parts of the display to estimate their rotation, and using that estimate to obtain the object velocity.

Robust heading perception despite degradation of the flow-field by noise, is not a new finding. This has been reported before for linear motion without ego-rotation, for motion on a curved path (Warren, Blackwell, Kurtz, Hatsopoulos & Kalish, 1991) and for differential motion components such as curl, shear and divergence (DeBruyn & Orban, 1990). Using a different type of noise, van den Berg (1992) showed robust heading perception for combinations of ego-rotation and ego-translation. In the latter study subjects could distinguish the landmarks from the noise because of the latter points' erratic motion paths. Here we report robust heading perception for similar ego-motion when no such discrimination is possible.

Accurate heading perception in simulations of pure ego-translation requires no more than a two frame point lifetime [45.5 msec (Warren *et al.*, 1991)]. For combinations of rotation and translation longer lifetimes were required on the order of 120 msec [eight frames (van den Berg, 1992)]. All these data point to a high degree of temporal and spatial integration in human motion processing relevant to the perception of ego-motion. However, it is difficult to assess the efficiency of human heading perception from these data.

Here, using the same psychophysical techniques to obtain thresholds for the optimal observer, we find that humans can tolerate more noise than any observer that uses only the information that is contained in the optic flow (Fig. 5). Could this result from overestimating the total noise ( $SNR_i$ ) for the subject relative to the total noise for the optimal observer? Recall that for the human observers the total noise was determined by the discrimination threshold for motion and the perturbations of the flow-vectors in the stimulus. We have adopted a rather high constant level for the human discrimination threshold (low  $SNR_i$ ) and neglected anisotropies and base speed dependency in motion discrimination (DeBruyn & Orban, 1988) that might indicate a lower discrimination threshold and consequently a higher  $SNR_i$  for our subjects. We felt confident in doing so because in pilot studies we found that the thresholds do not improve for  $SNR_v > 8$ . This indicates

that for  $SNR_v$  higher than this value the internal noise dominates the response, suggesting an overall  $SNR_i$  (ignoring the different speeds of the dots) of 5–10. Nevertheless, even if we would assume that the speed discrimination threshold is only 2% (raising  $SNR_i$  to 50), this would have negligible effects. For example, in Fig. 5 the human data points collected at  $SNR_v = 2$  would shift to the right in the graph by less than 0.04 units as a result of the change in  $SNR_i$  from 10 to 50; the effect would be even less than 0.01 unit for lower values of  $SNR_v$ . Consequently, we conclude that humans use additional information on depth (and possibly shape) that is independent of the optic flow. We suggest that depth order, as specified by perspective, texture gradients and the nearness to the horizon enables humans to select the furthest points in the display for estimation of the ego-rotation. From our data, we cannot distinguish between the separate contributions of these cues. It should be noted that we do not wish to suggest that human performance always exceeds that of an optimal observer that uses only the optic flow. To the contrary, perception of heading when moving through a cloud of dots is error-prone (van den Berg, 1992; Royden *et al.*, 1993), but not so when moving across the ground plane. Possibly a higher sensitivity to noise in the random cloud is caused by the absence of the static depth cues which enable the selection of those parts of the flow-field that provide the most reliable information for the estimation of the ego-rotation.

## REFERENCES

- van den Berg, A. V. (1992). Robustness of perception of heading from optic flow. *Vision Research*, 32, 1285–1296.
- van den Berg, A. V. (1993). Perception of heading. *Nature*, 365, 497–498.
- Brandt, T., Wist, E. R. & Dichgans, J. (1975). Foreground and background in dynamic spatial orientation. *Perception & Psychophysics*, 17, 497–503.
- Brenner, E. (1991). Judging object motion during smooth pursuit eye movements: the role of optic flow. *Vision Research*, 31, 1893–1902.
- Cutting, J. E., Springer, K., Braren, P. A. & Johnson, S. H. (1992). Wayfinding on foot from information in retinal, not optical, flow. *Journal of Experimental Psychology: General*, 121, 41–72.
- DeBruyn, B. & Orban, G. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research*, 28, 1323–1335.
- DeBruyn, B. & Orban, G. (1990). The role of direction information in the perception of geometric optic flow components. *Perception & Psychophysics*, 47, 433–438.
- van Doorn, A. J. & Koenderink, J. J. (1982). Temporal properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, 45, 179–188.
- van Doorn, A. J. & Koenderink, J. J. (1984). Spatiotemporal integration in the detection of coherent motion. *Vision Research*, 24, 47–53.
- Duffy, C. & 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, 1329–1345.
- Ferman, L., Collewijn, H. & van den Berg, A. V. (1987). A direct test of Listing's law—II. Human ocular torsion measured under dynamic conditions. *Vision Research*, 27, 939–951.
- Finney, D. J. (1966). *Probit analysis*. Cambridge: Cambridge University Press.
- Gibson, J. J. (1966). *The perception of the visual world*. Boston, Mass.: Houghton-Mifflin.

- van de Grind, W. A., van Doorn, J. A. & Koenderink, J. J. (1986). The distribution of human motion detector properties in the monocular visual field. *Vision Research*, 26, 797-810.
- Hildreth, E. C. (1992). Recovering heading for visually-guided navigation. *Vision Research*, 32, 1177-1192.
- Howard, I. P. (1981). *Human visual orientation*. Chichester: Wiley.
- Koenderink, J. J. & van Doorn, J. A. (1987). Facts on optic flow. *Biological Cybernetics*, 56, 247-254.
- Longuet-Higgins, H. C. & Prazdny, K. (1981). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, 208, 385-397.
- McKee, S. P. (1981). A local mechanism for differential velocity detection. *Vision Research*, 21, 491-500.
- McKee, S. P. & Welch, L. (1985). Sequential recruitment in the discrimination of velocity. *Journal of the Optical Society of America A*, 2, 243-251.
- Mikami, A., Newsome, W. T. & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interaction in MT and V1. *Journal of Neurophysiology*, 55, 1328-1351.
- Ohmi, M. & Howard, I. P. (1988). Effect of stationary objects on illusory forward self-motion induced by a looming display. *Perception*, 17, 5-12.
- Ohmi, M., Howard, I. P. & Landolt, J. P. (1987). Circular vection as a function of foreground-background relationships. *Perception*, 16, 17-22.
- Perrone, J. A. (1992). Model for the computation of self-motion in biological systems. *Journal of the Optical Society of America A*, 9, 177-194.
- Regan, D. & Beverley, K. I. (1982). How do we avoid confounding the direction we are looking and the direction we are moving? *Science*, 215, 194-196.
- Rieger, J. H. & Lawton, D. T. (1985). Processing differential image motion. *Journal of the Optical Society of America A*, 2, 354-359.
- Rieger, J. H. & Toet, L. (1985). Human visual navigation in the presence of 3-D rotations. *Biological Cybernetics*, 56, 377-381.
- Royden, C. S., Banks, M. S. & Crowell, J. A. (1992). The perception of heading during eye movements. *Nature*, 360, 583-585.
- Tanaka, K., Hikosaka, K., Saito, H., Yujie, M., Fukuda, Y. & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *Journal of Neuroscience*, 6, 134-144.
- Tweed, D., Fetter, M., Andreaki, S., Koenig, E. & Dichgans, J. (1992). Three-dimensional properties of human smooth pursuit eye movements. *Vision Research*, 32, 1225-1238.
- Warren, W. H. & Hannon, D. J. (1988). Direction of selfmotion is perceived from optical flow. *Nature*, 336, 162-163.
- Warren, W. H. & Hannon, D. J. (1990). Eye movements and optical flow. *Journal of the Optical Society of America A*, 7, 160-169.
- Warren, W. H., Morris, M. W. & Kalish, M. (1988). Perception of translational heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 646-660.
- Warren, W. H., Blackwell, A. W., Kurtz, K. J., Hatsopoulos, N. G. & Kalish, M. L. (1991). On the sufficiency of the velocity field for perception of heading. *Biological Cybernetics*, 65, 311-320.
- Williams, D. W., Phillips, G. & Sekuler, R. (1986). Hysteresis in the perception of motion direction as evidence for neural cooperativity. *Nature*, 324, 253-255.

**Acknowledgements**—We thank Ans van Doorn and Jan Koenderink for their algorithm implementing the least-squares solution from the optic flow. The patience of our subjects René van Rijn, Hans van der Steen, Jan van der Kuil, Rien Grund and Aldo Ferraresi is greatly appreciated.

## APPENDIX A

In order to obtain a similar distribution of points on the screen for the large and the small depth range we adopted the following procedure when the large depth range (35 m) was presented.

(1) 80 points were randomly distributed in the plane (depth 12 m, width 24 m). Thus, the furthest points in this rectangle would be projected about 4 deg lower on the screen than points at 35 m distance.

(2) For each point the (Fick) direction angles as seen from the vantage point were computed from their cartesian coordinates (horizontal  $x$ ; vertical  $y$ ; depth  $z$ ) relative to the vantage point,

$$\begin{aligned}\text{horizontal angle } \phi &= \text{atan}(x/z) \\ \text{vertical angle } \theta &= \text{atan}(y/(x^2+z^2)^{1/2})\end{aligned}$$

note that  $y$  equalled the eye height (1.3 m) for all points, because the points were located in the ground plane.

(3) For each point, a new point was found by reduction of the vertical angle by 4 deg ( $\theta'$ )

$$\begin{aligned}\phi' &= \phi \\ \theta' &= \theta - 4\end{aligned}$$

and the new cartesian coordinates were computed with the inverse expressions of step (2)

$$\begin{aligned}y' &= y \\ x' &= y' \sin(\phi') / \tan(\theta') \\ z' &= y' \cos(\phi') / \tan(\theta')\end{aligned}$$

Because each new point was rotated upwards by 4 deg relative to the original point the pattern of the points on the screen remained the same. However the simulated distances of the new points were larger, such that the furthest points were now found at about 35 m from the vantage point.

## APPENDIX B

We present the computational details of the responses of the optimal observer when using three sources of information: the optic flow, static distance cues and knowledge of the orientation and shape of the ground plane.

In our simulations the optic flow is generated through motion of the simulated vantage point "V" relative to the stationary simulated world. The vantage point can be conceived of as a camera. Its position and orientation are defined with two vectors and a "torsion angle". One vector defines the position relative to the world ( $\mathbf{vp}$ ), the other the direction in which the camera is oriented ( $\mathbf{dv}$ ). The torsion angle ( $\psi$ ) is defined as the angular rotation about the viewing direction, and corresponds to the Fick angle of torsion (Howard, 1981). In our simulations of eye rotations  $\psi$  was always set to zero. Because eye movements closely follow Listing's law (Ferman, Collewijn & van der Berg, 1987; Tweed, Fetter, Andreaki, Koenig & Dichgans, 1992) this was not entirely realistic. However, in a pilot study it was found that simulated eye torsion in keeping with Listing's law did not affect heading thresholds. Thus, eye torsion was neglected.

The "environment" consisted of  $N$  points in the plane. They were fully characterized by their three-dimensional position vectors. Changing  $\mathbf{vp}$  in each frame created the unperturbed flow-field on the screen. The camera motion was a simple translatory motion with constant speed  $T$  in the direction  $\mathbf{t}$ . To simulate eye rotation,  $\mathbf{dv}$  was also changed so as to keep the "camera" directed at the same point in the plane throughout the presentation. In the perturbed flow-fields each point in the plane was displaced by a vector  $\delta_i$  in step with the change of the vantage point.

The error term ( $\epsilon_i$ ) was computed for each flow vector  $\mathbf{p}_i$  that belonged to the landmark "i" seen in the direction  $\mathbf{d}_i$ . Notice that in these computations  $\mathbf{t}$ ,  $\mathbf{d}_i$  and  $\mathbf{dv}$  were direction vectors, i.e. with unit length. The following steps can be distinguished.

The ego-rotation was given by:

$$\mathbf{R} = T_j D_v (\mathbf{t} \times \mathbf{dv}) \quad (\text{B1})$$

the unperturbed flow-vector for each landmark was given by:

$$\mathbf{p}_i = -T_j D_i (\mathbf{t} - (\mathbf{t} \cdot \mathbf{d}_i) \mathbf{d}_i) - \mathbf{R} \times \mathbf{d}_i \quad (\text{B2})$$

where  $D_i$  denotes the distance between the landmark and the vantage point and  $D_v$  the distance between the fixation point and V. The vector

$\mathbf{p}_i$  is located in the normal plane of  $\mathbf{d}_i$ . The magnitude of the error term was proportional to the unperturbed flow of the landmark:

$$\|\epsilon_i\| = \|\mathbf{p}_i\|/\text{SNR}_v \tag{B3}$$

The direction of  $\epsilon_i$  was chosen in the plane perpendicular to the direction vector  $\mathbf{d}_i$ . To this end a local reference frame was defined ( $\mathbf{e}_1, \mathbf{e}_2$  unit vectors):

$$\begin{aligned} \mathbf{e}_1 &= \mathbf{t} \times \mathbf{d}_i / (\|\mathbf{t} \times \mathbf{d}_i\|) \\ \mathbf{e}_2 &= \mathbf{e}_1 \times \mathbf{d}_i \end{aligned}$$

With respect to this reference frame each perturbation term  $\epsilon_i$  was given a random direction  $\alpha_i$ :

$$\epsilon_i = (\sin(\alpha_i)\mathbf{e}_1 + \cos(\alpha_i)\mathbf{e}_2)\|\mathbf{p}_i\|/\text{SNR}_v \tag{B4}$$

the displacement  $\delta_i$  was simply  $D_i \epsilon_i$ .

The optimal observer was given the directions  $\mathbf{d}_i$  to the landmarks and the perturbed flow vectors  $\rho_i (= \mathbf{p}_i + \epsilon_i)$ . The optimal observer now picked those  $\mathbf{t}, \mathbf{R}$  and  $T/D_i$  that minimized  $E$ , with

$$E = 1/N \sum_i \|\rho_i - \mathbf{p}_i\|^2$$

Notice that  $N$  indicates the number of visible landmarks on the screen. Following Koenderink and van Doorn (1987) the variational problem was solved iteratively using Lagrange multipliers; the minimum of  $E$  is found for locations where the gradient of  $E$  is a linear combination of the gradients of the constraint functions. In case of the optimal observer using only the optic flow there is one constraint function:

$$\|\mathbf{t}\| - 1 = 0$$

resulting in,

$$\text{grad}(E) + \lambda \text{grad}(\|\mathbf{t}\| - 1) = 0. \tag{B5}$$

This equation was iteratively solved by Koenderink and van Doorn (1987). When the optimal observer uses "knowledge" that points were

located in the ground plane, a second constraint function applies to minimize  $E$ :

$$T/H(\mathbf{d}_i \cdot \mathbf{n}) - T/D_i = 0$$

where  $H$  denotes the height of the vantage point above the ground plane and  $\mathbf{n}$  is the unit normal vector to the ground plane. This function states that the landmarks should be in the ground plane, resulting in

$$\text{grad}(E) + \lambda_1 \text{grad}(\|\mathbf{t}\| - 1) + \lambda_2 \text{grad}((\mathbf{d}_i \cdot \mathbf{n})T/H - T/D_i) = 0. \tag{B6}$$

Rearranging terms one obtains expressions that are suitable for iterative solution. The expressions for the ego-motion parameters obtained from equations (B5) and (B6) are identical and can be found in Koenderink and van Doorn (1987). The expression for the "distances" ( $T/D_i$ ) are different. The expression derived from equation (B5) can be found in Koenderink and van Doorn (1987) for equation (B6) one finds:

$$T/D_i = \frac{(-\lambda_2 N - 2((\rho_i + \mathbf{R} \times \mathbf{d}_i) \cdot (\mathbf{t} - (\mathbf{t} \cdot \mathbf{d}_i)\mathbf{d}_i)))}{2\|(\mathbf{t} - (\mathbf{t} \cdot \mathbf{d}_i)\mathbf{d}_i)\|^2}$$

where  $\lambda_2$  is chosen such that it minimizes the variance of the height of the landmarks with respect to the ground plane,

$$\lambda_2 = \frac{-\sum a_i b_i + 1/N \sum a_i \sum b_i}{\sum a_i^2 - 1/N \sum a_i \sum a_i}$$

with

$$a_i = -N/(2(\mathbf{d}_i \cdot \mathbf{n})\|(\mathbf{t} - (\mathbf{t} \cdot \mathbf{d}_i)\mathbf{d}_i)\|^2)$$

and

$$b_i = ((\rho_i + \mathbf{R} \times \mathbf{d}_i) \cdot (\mathbf{t} - (\mathbf{t} \cdot \mathbf{d}_i)\mathbf{d}_i))/((\mathbf{d}_i \cdot \mathbf{n})\|(\mathbf{t} - (\mathbf{t} \cdot \mathbf{d}_i)\mathbf{d}_i)\|^2).$$

The "distance", of the vantage point above the ground plane estimated from the optic flow is given by:

$$T/H = \lambda_2 1/N \sum a_i + 1/N \sum b_i.$$