# World-centered perception of 3D object motion during visually guided self-motion

**Kazumichi Matsumiya**

Research Institute of Electrical Communication,
Tohoku University, Katahira, Sendai, Japan

**Hiroshi Ando**

ATR Cognitive Information Science Laboratories,
Hikaridai, Kyoto, Japan, &
Multimodal Communication Group,
Universal Media Research Center,
National Institute of Information and Communications Technology,
Hikaridai, Kyoto, Japan

We investigated how human observers estimate an object's three-dimensional (3D) motion trajectory during visually guided self-motion. Observers performed a task in an immersive virtual reality system consisting of front, left, right, and floor screens of a room-sized cube. In one experiment, we found that the presence of an optic flow simulating forward self-motion in the background induces a world-centered frame of reference, instead of an observer-centered frame of reference, for the perceived rotation of a 3D surface from motion. In another experiment, we found that the perceived direction of 3D object motion is biased toward a world-centered frame of reference when an optic flow pattern is presented in the background. In a third experiment, we confirmed that the effect of the optic flow pattern on the perceived direction of 3D object motion was not caused only by local motion detectors responsible for the change of the retinal size of the target. These results suggest that visually guided self-motion from optic flow induces world-centered criteria for estimates of 3D object motion.

Keywords: optic flow, 3D object motion, reference frame, self-motion

## Introduction

As we are moving through the environment, our self-motion produces a complex pattern of motion in the retinal image. For example, forward self-motion creates a radial pattern of motion in the retinal image (Gibson, 1950). This complex pattern of motion is referred to as optic flow. It has been suggested that optic flow is used for the visual control of locomotion (Warren, Kay, Zosh, Duchon, & Sahuc, 2001). Human observers can make accurate heading judgments from the optic flow (Crowell & Banks, 1996; van den Berg, 1992; Warren, Morris, & Kalish, 1988), even during pursuit eye movements (Li & Warren, 2000; Royden, Banks, & Crowell, 1992; Royden, Crowell, & Banks, 1994; Warren & Hannon, 1988).

As we move in the environment, we often estimate object movements. To estimate them accurately, the visual system has to separate information about the movement of the object from the optic flow produced by self-motion. In such a situation, the visual system can use extraretinal information (proprioceptive and vestibular information or an efference copy of the motor command) to compensate for the effects of self-motion on the retinal image (Gogel, 1990; Wallach, 1987). The mechanisms of the compensation processes when extraretinal information is available have been recently studied using an immersive virtual reality system (Jaekl et al., 2005; Tcheang, Gilson, & Glennerster, 2005).

However, recent studies have suggested that by using retinal information alone, the visual system can compensate for the effects of self-motion in estimating an object's motion during self-motion (Rushton & Warren, 2005; Warren & Rushton, 2007, 2008). Rushton and Warren (2005) have proposed that optic flow processing divides retinal motion signals into components due to self-motion and those due to object movements. This process is referred to as flow-parsing. Warren and Rushton (2007, 2008) provided evidence that flow-parsing is implicated in stationary observers' estimation of the trajectory of an object moving on a 2D fronto-parallel plane during visually guided self-motion from optic flow. In addition, other recent studies have shown that the presence of optic flow in the background affects observers' ability to judge the velocity of a moving object (Brenner, 1993; Brenner & van den Berg, 1996; Gray, Macuga, & Regan, 2004), the judgment of the time to collision with an approaching or receding object (Gray & Regan, 2000), and the perceived direction of an object's motion-in-depth (Gray et al., 2004). These results indicate that judgments concerning 3D object motion are not independent of visually guided self-motion from optic flow, suggesting

that the effects of optic flow on 3D object motion may be explained by the flow-parsing account.

Flow-parsing plays an important role in interpreting how objects move in a stationary environment. According to the flow-parsing account, the visual system can divide retinal motion signals into a component of motion due to self-movement and a component of motion due to object movement. The former arises from an observer's own movement in a stationary, world-centered reference frame. This is because the observer usually moves in an environment in which most of the objects are stationary in the stationary, world-centered reference frame. On the other hand, the latter also has to be represented in the stationary, world-centered reference frame.

Consider the following case. An observer is walking toward a plane in a stationary world consisting of textured walls. The plane is rotating around the horizontal axis of the plane, whose central location is fixed in the world (Figure 2a). In this case, the observer's retinal image is a combination of a radially expanding pattern of motion created by the self-movement of the observer and a vertically contracting pattern of motion created by the object movement of the plane. If the visual system decomposes the retinal motion into self- and object-movement components, the component of retinal motion due to self-movement can be subtracted from the total retinal motion in order to compensate for self-movement. As a result, the movement of the plane relative to the world can be estimated, leading to the interpretation of object movement in the world-centered reference frame. However, if the visual system does not decompose the retinal motion, the total retinal motion is used to estimate the movement of the plane. As a result, the movement of the plane relative to the eyes, head, or body of the observer can be estimated, leading to the interpretation of object movement in the observer-centered reference frame.

We investigated whether the presence of optic flow, which simulates self-motion in the background, induces the interpretation of object movement in the world-centered reference frame for a 3D structure from motion and motion-in-depth for stationary observers.

## General methods

### Apparatus

Stereo images, generated on four synchronized workstations at 30 frames/s, were presented in an immersive virtual reality system by projectors positioned outside. The projectors were directed to four screens (three walls and one floor) of a room-sized cube, which served as the virtual reality system (Figure 1). For each screen, a pair of projectors with polarizing filters was set; one projector
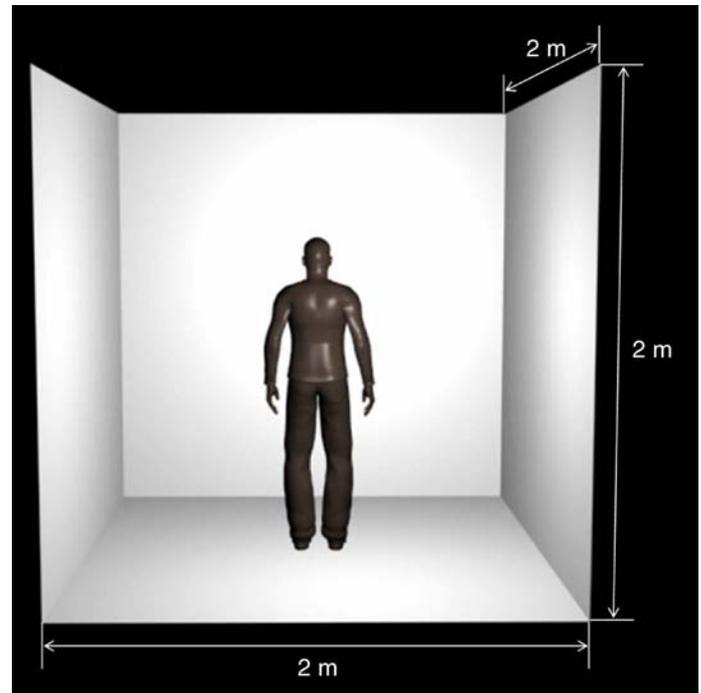


Figure 1. Four screens of a room-sized cube (three walls and one floor). Each screen is subtended 2 m in width and 2 m in height. A pair of projectors was directed at each screen. Observers stood at the center of the floor screen. In all experiments, the observers were stationary in the virtual environment presented on the four screens.

created images for the right eye and the other projector created images for the left eye. Observers wore polarizing glasses to see the stereo images while holding a joystick and standing at the center of the floor in the virtual reality system, and they remained stationary at all times. Because of differences in the stature of the observers, their head positions were measured with a position tracking system in order to correct the calculations to produce the stereo images.

## Experiment 1

### Purpose

To examine whether visually simulated self-motion induces a world-centered frame of reference, we applied a psychophysical method developed by Wexler, Lamouret, and Droulez (2001). Their method takes advantage of ambiguity in perceiving a 3D surface from motion. They considered that a set of dots moving in a frontal plane simulated a surface rotating around the horizontal axis while the observer was moving toward the surface (Figure 2a). In this case, forward self-motion of
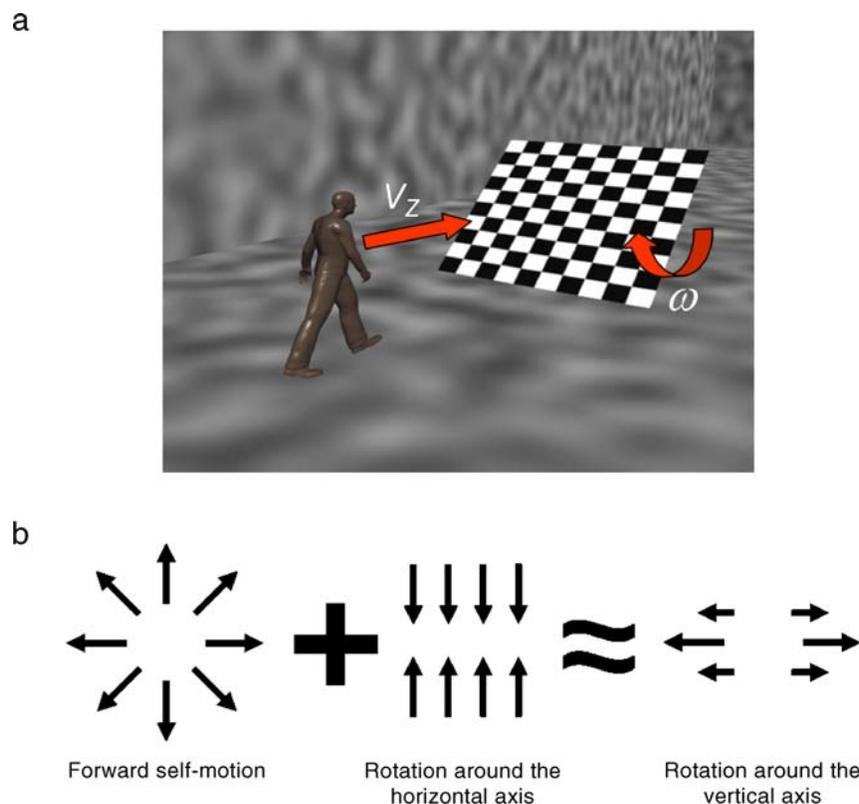
Figure 2. Illustration of the method developed by Wexler, Lamouret et al. (2001). (a) Walking observer and rotating plane. While the observer is walking toward the plane, the plane is rotating around the horizontal axis that interacts at the center of the plane and is parallel to the plane. $V_Z$ represents the walking velocity and $\omega$ represents the rotation velocity of the plane. (b) Retinal image motion created by simulating forward self-motion toward the plane combined with plane rotation around the horizontal axis. If observers perceived the 3D configuration in a world-centered frame of reference, they would judge the rotation of the plane correctly. However, if observers perceived the 3D configuration in an observer-centered frame of reference, they would misjudge the plane rotation as a simulation of the plane rotating around the vertical axis.

the observer created a radially expanding pattern of motion in the retinal image, and rotating object motion of the surface created a vertically contracting pattern of motion in the retinal image (Figure 2b). The rotation velocity of the surface and the moving velocity of the simulated self-motion were selected to cancel out because the expanding and contracting movements in the upper and lower visual fields were canceled out. As a result, the expanding movements remained almost entirely in the left and right visual fields (Figure 2b). Such an optic flow pattern of the rotation surface in the retinal image can have two different interpretations. If the visual system uses a world-centered frame of reference to perceive the 3D surface from motion, the observer should have an impression of a surface rotating around the horizontal axis. However, if the visual system uses an observer-centered frame of reference, the observer should have an impression of a surface rotating around the vertical axis. Thus, the perception of a 3D surface from motion changes depending on the frame of reference used. In the present study, a set of random dots moving in a frontal plane was presented in the central visual field in a virtual world. The movement of the dots was created by simulating a surface

rotating around the horizontal axis while an observer was moving toward the surface. In addition to the moving dots presented in the central visual field, a virtual room consisting of textured walls was presented in the peripheral visual field.

## Methods

Four virtual worlds were generated (Figure 3). First, the *moving room* consisted of frontal, left, and right walls with a ground plane and ceiling (Figures 3a and 3b). Their surfaces had a filtered noise pattern. The simulated size of the moving room subtended 1,300 cm in width, 300 cm in height, and 1,750 cm in depth (Figure 3a). Second, the *moving floor* consisted of a ground plane with a filtered noise pattern (Figure 3c). The simulated size of the moving floor subtended 1,300 cm in width and 1,750 cm in depth. The moving room and the moving floor moved at a constant speed of 200 cm/s in the virtual world, in order to simulate the forward linear translation of observers (the red arrows in Figures 3b and 3c). At the beginning of the trials, the front of the ground plane was set 1,250 cm
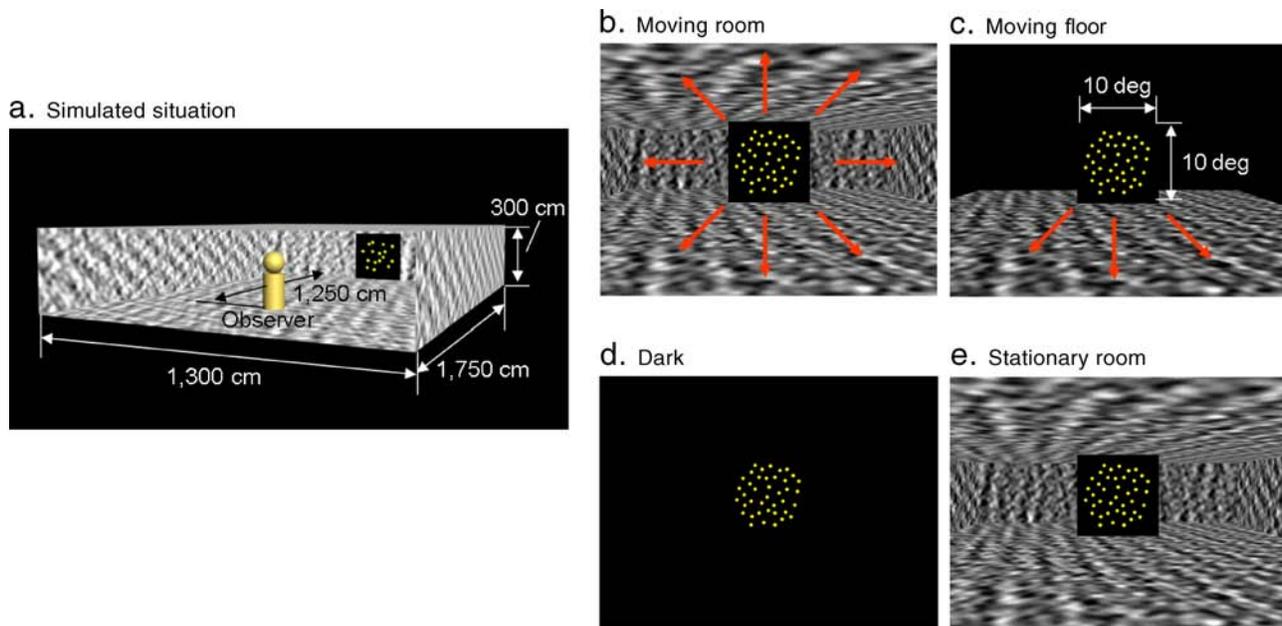
Figure 3. Virtual environments used in Experiment 1. (a) Simulated situation for the moving-room and stationary-room conditions: The simulated size of the virtual room subtended 1,300 cm in width, 300 cm in height, and 1,750 cm in depth. The surfaces of the virtual room had a filtered noise pattern. At the beginning of the trials, the frontal wall was set 1,250 cm ahead of the observer. A set of yellow dots was presented on the black frontal plane, which was set 1,240 cm ahead of the observers in the virtual room. (b) Moving room: The flow pattern simulating forward self-motion at the speed of 200 cm/s was provided by moving textured surfaces. The red arrows represent the flow of the simulated self-motion. (c) Moving floor: The frontal, left, and right walls and the ceiling were painted in black. (d) Dark: All surfaces were painted in black. (e) Stationary room: This room was the same as the moving room except that it was stationary.

ahead of observers (Figure 3a). Third, the *dark room* consisted of walls, a ground plane, and a ceiling covered with black uniform surfaces (Figure 3d). Fourth, the *stationary room* was similar to the moving room except for the fact that it was stationary (Figure 3e). These virtual worlds were presented on the four screens of the virtual reality system (Figure 1). A sensation of motion-in-depth was created by changing the size of the simulated object as well as its binocular disparity. The simulated objects were rendered using anti-aliasing and geometric perspective projection from the observers' eyes with binocular disparity.

In the central visual field of 10 deg in diameter, a set of moving yellow dots was presented on a black frontal plane (10 deg in width and 10 deg in height; Figure 3c). The simulated location of the black frontal plane was set 1,240 cm ahead of observers in the virtual world using binocular disparity. Fifty yellow dots moved within the black frontal plane. The movement of the dots was created by simulating a surface rotating around the horizontal axis while an observer was moving toward the surface. In other words, the simulated forward self-motion provided a radially expanding pattern of motion in the display, and the surface rotation around the horizontal axis provided a vertically contracting pattern of motion in the display. Therefore, the dot movements were a combination of the radially expanding motion pattern and the vertically contracting motion pattern. In order to cancel out the

expanding and contracting movements of the dots in the upper and lower visual fields, the rotation velocity of the simulated surface ($\omega$) was determined by

$$\omega = \frac{V_Z}{E_Z \tan\sigma}, \tag{1}$$

where $V_Z$ is the moving velocity of the simulated forward self-motion, $E_Z$ is the distance between the eye and the surface, and $\sigma$ is the initial inclination of the surface (Wexler, Lamouret et al., 2001). In this experiment, $V_Z$ was 200 cm/s, $E_Z$ was 1,300 cm, and $\sigma$ was 45 deg. The expanding movements of the dots remained almost entirely in the left and right visual fields. In other words, the flow pattern of the dots consisted of an approximately 1D horizontal expansion, as is illustrated on the right side of Figure 2b. In addition to the moving dots presented on the frontal screen of the virtual reality system, one of the four virtual worlds was presented on the four screens of the virtual reality system.

The virtual room (except the stationary room) started to move when observers pressed a button on the joystick. After one second, the yellow dots in the central visual field moved for one second while the virtual room was presented. Then, the yellow dots disappeared from the display. The observers' task was to indicate whether the surface's perceived axis of rotation was horizontal or

vertical. The type of virtual rooms was randomly varied from trial to trial.

The experimental run comprised 40 trials (4 virtual rooms × 10 times). Each observer performed two experimental runs.

Two female and two male observers with corrected-to-normal vision participated in this experiment. The two female observers were experienced in other psychophysical experiments but did not know the purpose of this study. The two male observers were authors of this paper.

## Results and discussion

Figure 4 shows the mean percentages of perception of a surface rotating around the horizontal axis for the four virtual worlds. The data are the means of the four observers. In Figure 4, when the moving room simulating self-motion was presented in the virtual world, the observers mostly perceived a surface rotating around the horizontal axis. When the moving floor was presented, the rate of perception of a surface rotating around the horizontal axis was reduced compared to the rate for the moving room. When the dark room or stationary room was presented, the observers mostly perceived a surface rotating around the vertical axis. These results indicate that the rate of perception of a

surface rotating around the horizontal axis increases when the large moving textured background stimulus is presented in the peripheral visual field ($F(3, 12) = 78.82$, $p < 0.0001$). This suggests that the presence of optic flow simulating self-motion leads to a bias toward a world-centered interpretation in perceiving the rotating surface for stationary observers.

Several studies have shown that the magnitude of the perception of self-motion increases with the size of the display (Anderson & Braunstein, 1985; Brandt, Dichgans, & Koenig, 1973; Howard & Heckmann, 1989). This suggests that the large moving textured background stimulus used in the present study facilitates the perception of self-motion. Therefore, it is possible that the perception of self-motion might be required to induce a world-centered frame of reference in perceiving a 3D structure from motion. However, recent studies have shown that even though a relatively small field of view was used, the presence of optic flow, which simulated self-motion, changed the perceived trajectory of a probe (Rushton & Warren, 2005; Warren & Rushton, 2007, 2008). This suggests that the perception of self-motion is not necessary to compensate for its effects in estimating object movement during self-motion. Thus, the findings of Rushton and Warren imply that the world-centered interpretation of a 3D structure from motion may not depend on the magnitude of self-motion perception.
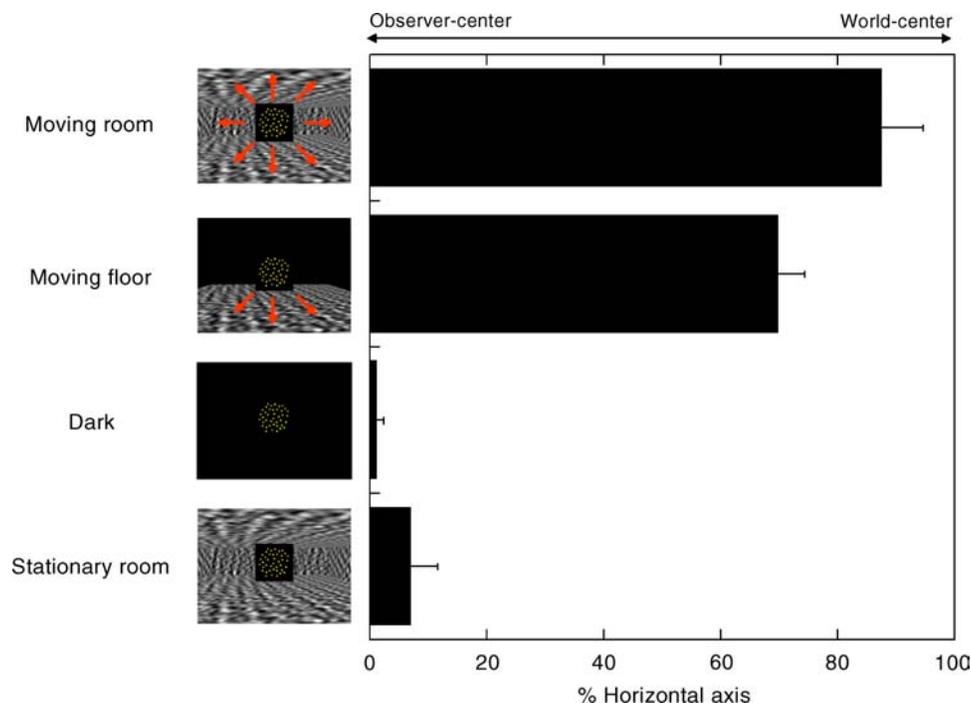


Figure 4. Results of Experiment 1. The graph shows the percentage of observers reporting a horizontal axis created from the pattern of yellow central dot motion across all visual background conditions. Four observers participated in this experiment. Data were averaged for the four observers. In the moving-room and moving-floor conditions, the flow pattern simulating forward self-motion at the speed of 200 cm/s was provided by moving textured surfaces in the peripheral visual field. The perception of the horizontal and vertical axes corresponds to biases toward world- and observer-centered frames of reference, respectively. Error bars represent the *SEM*.

## Experiment 2

### Purpose

We devised an experiment to distinguish between a bias toward a world-centered interpretation and a bias toward an observer-centered interpretation in the perception of the 3D object-motion direction. Observers looked at a sphere moving along a 3D trajectory from their left or right side toward the sagittal plane, as illustrated in Figure 5a. In Figure 5a, the red circle represents the moving sphere, and the black arrow represents the motion direction of the sphere. Forward self-motion was simulated by moving a room with textured surfaces in the 3D virtual world. We defined the trajectory angle ($\theta$) as an angle to the line between observers and the target (Figure 5a). The trajectory angle varied, as illustrated by the black arrows in Figure 5b. The endpoints of target motions ($x$, $z$) were the same in depth but were different in horizontal locations (Figure 5b). When forward self-motion was simulated by moving the virtual room, the $z$ component of the target motion was consistent with the motion of the virtual room for all motion trajectories. In this situation, the use of an observer-centered frame of reference leads to the interpretation that observers are stationary and the world (the textured room) is moving (Figure 5b). On the other hand, the use of a world-centered frame of reference leads to the interpretation that observers are moving and the world (the textured room) is stationary (Figure 5c). If observers use an observer-centered frame of reference to estimate the motion direction of the target, the total retinal motion represents

object movements alone. As a result, different motion directions of the target should be perceived according to the presented trajectory angles regardless of the optic flow in the background (the black arrows in Figure 5b). However, if observers use a world-centered frame of reference, the total retinal motion includes self- and object-movement components, and the component of retinal motion due to self-movement is subtracted from the total retinal motion. As a result, the perceived direction of the target should be constant across the presented trajectory angles, because the $z$ component of the target motion is not regarded as a motion component in the world-centered frame for this experiment (the black arrows in Figure 5c).

### Methods

In Experiment 2, three virtual worlds were generated: first, the moving room; second, the stationary room; third, the dark room. These worlds were the same as those in Experiment 1 except that the moving yellow dots and the central frontal black plane (see Figure 3) were not presented; instead, a red virtual ball was presented at a height of 150 cm above the ground plane. For each trial, an observer tracked the ball trajectory with pursuit eye movements when the ball moved in the virtual environment, even though it has been demonstrated that the perceived trajectory of the target does not depend on whether observers track the target by eye movements or by continuously looking at a stationary marker (Welchman, Tuck, & Harris, 2004). The initial position of the ball was selected to be on the left ($-18$ deg) or right ($+18$ deg)
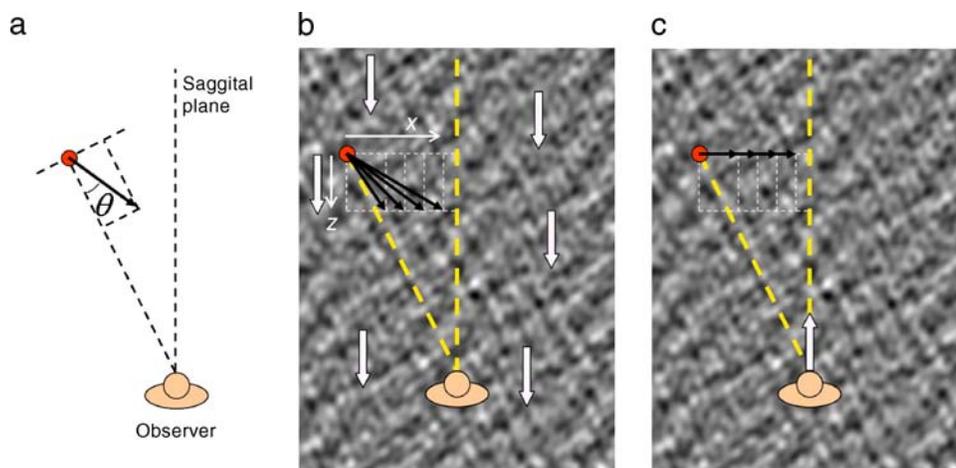


Figure 5. Top view of simulation in Experiment 2. (a) Observer and typical target trajectory at angle $\theta$ to the line between the observer and the target. The target moves toward the sagittal plane. (b) Simulated situations in a virtual environment. The endpoints of target motions are the same in depth but are different in horizontal locations. In the moving-room condition, a virtual room consisting of textured surfaces moves toward the observer. This motion is created by simulating linear forward self-motion at the speed of 200 cm/s. (c) Observers generally interpret the pattern in (b) as simulating self-motion, leading to perceived self-motion. If the simulated self-motion triggered the use of a world-centered frame of reference to estimate the target trajectory, the perceived direction of the target motion would be constant, as illustrated in the figure.

| Trajectory angle, $\theta$ (deg) | $x$ at trajectory endpoint (cm) | $z$ at trajectory endpoint (cm) | Speed (cm/s) |
|---|---|---|---|
| 5.8 | 90 | 200 | 219 |
| 12.5 | 120 | 200 | 233 |
| 18.5 | 150 | 200 | 250 |
| 23.5 | 180 | 200 | 269 |

Table 1. Parameters for Experiment 2.

randomly for each trial. The ball moved toward the sagittal plane of the observer along a 3D trajectory. The values of the trajectory angle, the magnitude of the $x$ and $z$ distances, and the speeds of the target are given in Table 1.

Before each trial began, the observers stood at the center of the floor in the virtual reality system, and one of three virtual worlds was presented with a marker. The observers were asked to fixate on the marker on the left or right side of the display. A virtual room moved toward the observers for 3 sec. One second after the room moved with the marker, the marker was replaced with the virtual ball. The ball moved for 1 sec. After that, the ball disappeared from the display before reaching the sagittal plane. One second after the ball disappeared, a virtual vertical pole and the ground plane with the noise pattern were presented in the display. The observers' task was to indicate, by moving the virtual vertical pole in depth with the joystick, the position where the ball would arrive in the sagittal plane of the observers in the future.

One female and three male observers with corrected-to-normal vision participated in this experiment. The one female and the two male observers were experienced in other psychophysical experiments but did not know the purpose of this study. One of the three male observers was one of the authors of this paper.

## Results and discussion

Figure 6 shows the apparent depth position from the observers as a function of the trajectory angles ($\theta$ in Figure 5a). In Figure 6, open squares represent the mean data of all four observers, and each solid symbol type represents a different observer. The blue dashed line represents the prediction line calculated from the target trajectories when the observers used an observer-centered frame of reference (see the black arrows in Figure 5b). Note that, in the stationary-room condition, the observer-centered frame of reference overlaps the world-centered frame of reference. The red dashed line represents the prediction line when observers use the world-centered frame of reference in the moving-room condition (see the black arrows in Figure 5c). In the dark-room condition (Figure 6a), the apparent depth position increased with the trajectory angle of the target. The apparent depth positions were set to be much smaller than expected by the blue dashed line in Figure 6a. In the stationary-room condition (Figure 6b), the apparent depth positions increased with the trajectory angle of the target in a way similar to those in the dark-room condition, and the magnitude of the apparent depth positions slightly increased as compared with that in the dark-room condition. However, in the moving-room condition (Figure 6c), the apparent depth positions were set to be close to values represented by the red dashed line.

In addition, we used the slopes of the fitting lines to the data for each observer to calculate a difference index; this index quantifies the difference in slopes. We defined the slope difference index as $\text{SDI} = 100*(1 - S/S_{\text{dark}})$, where SDI is the slope difference index, $S_{\text{dark}}$ is the slope of the fitting line in the dark-room condition, and $S$ is the slope in the moving-room or the stationary-room condition. The
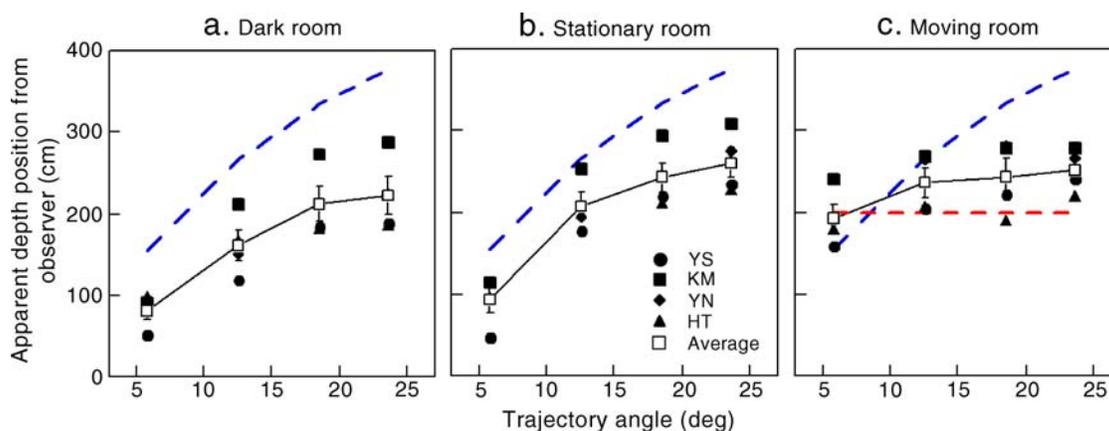


Figure 6. Results of Experiment 2. The graphs show the judged depth position as a function of trajectory angles ($\theta$ in Figure 4a) with respect to the line connecting the observer and the start point of the target. (a) Dark room. (b) Stationary room. (c) Moving room. Error bars represent the *SEM*. The blue dotted line represents the prediction line calculated from the target trajectories when observers use an observer-centered frame of reference. Note that, in the stationary-room condition, the observer-centered frame of reference is the same as the world-centered frame of reference. The red dotted line represents the prediction line when observers use a world-centered frame of reference in the moving-room condition.

index is 0% when the slope $S$ is the same as $S_{dark}$. The index is a positive value when slope $S$ is smaller than $S_{dark}$, and it is 100% when slope $S$ is zero. The index is a negative value when slope $S$ is larger than $S_{dark}$. Figure 7 shows the slope difference index for the moving-room and the stationary-room conditions. The gray bars represent the mean slope difference index of all four observers. The symbols represent different observers. As shown in Figure 7, in the moving-room condition, the mean slope difference index is 59%. This index is significantly greater than 0% ($t = 6.69$, $p < 0.01$). In the stationary-room condition, the mean slope difference index is −20%, and there is no significant difference from 0% ($t = 1.89$, $p = 0.16$ ns). These results confirm that the slope in the moving-room condition is significantly different from that in the dark-room condition, whereas the slope in the stationary-room condition is not.

These findings indicate the following characteristics:

1. the observers perceive the different motion directions of the target according to the motion trajectory angles in the conditions of the dark room and stationary room,
2. the observers tend to perceive the constant motion direction of the target across the motion trajectory
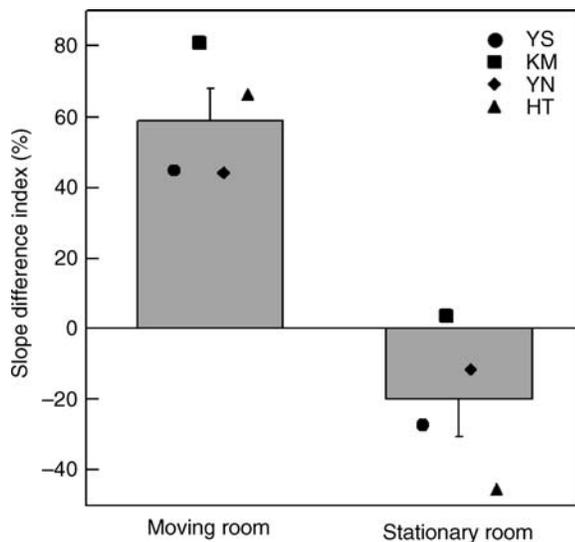


Figure 7. Comparison of the slopes in the moving-room and stationary-room conditions on the basis of the dark-room condition in Experiment 2. For each observer in each condition, we calculated the slope difference index that was defined to be SDI = 100*(1 − $S/S_{dark}$), where $S_{dark}$ is the slope of the fitting line to an observer's data in the dark-room condition and $S$ is the slope to the observer's data in the moving-room or stationary-room condition. A value of 0% indicates that the slope is the same as that in the dark-room condition. A positive value indicates that the slope is smaller than that in the dark-room condition. A negative value indicates that the slope is larger than that in the dark-room condition.

angles in the moving-room condition as compared with the other conditions.

Thus, these findings suggest that observers use the world-centered frame of reference to perceive the direction of 3D object motion during visually guided self-motion from the optic flow.

In addition, the results of Experiment 2 showed that the magnitude of the apparent depth positions tended to increase slightly as compared with that in the dark-room condition, although the slope of the fitting line to the apparent depth positions in the stationary-room condition was not significantly different from that in the dark-room condition (Figure 6b). This suggests that the observers might be able to use the world-centered frame of reference to perceive the direction of the 3D object motion even without visually guided self-motion from the optic flow. However, the effect of presenting the stationary room seems to be much weaker than that of presenting the moving room. In fact, the results of Experiment 1 indicate that the presence of the stationary room strongly biases toward the use of an observer-centered frame of reference, although the performance in the stationary-room condition slightly shifts toward a world-centered interpretation as compared with that in the dark-room condition (Figure 4). Alternatively, it is possible that the difference in apparent depth positions between the moving-room and the stationary-room conditions can be caused by local motion detectors that assess the change of retinal size of the target. In Experiment 3, we tested this possibility.

# Experiment 3

## Purpose

To determine whether the effect of the simulated self-motion on the perceived direction of 3D object motion was caused by local motion detectors that assess the change of retinal size of the target, we varied the gap between the outer edge of the target and the inner edge of the textured background (Figure 8a). In reports of psychophysical evidence, a changing-size detector consisting of local motion detectors has a small receptive field (1.5–2.0 deg; Beverley & Regan, 1982; Gray & Regan, 2000; Regan & Beverley, 1979). In Experiment 2, when the target moved along a 3D trajectory in the virtual world, the size of the target changed within the range of sizes of the receptive field. When the textured virtual room moved in order to simulate forward self-motion, both the target expansion and the flow of the textured room stimulated the changing-size detector, as illustrated in the left panel of Figure 8b. However, when the textured virtual room was stationary in the virtual world, the target expansion only stimulated the changing-size detector, as illustrated in the right panel of Figure 8b. Therefore, it is
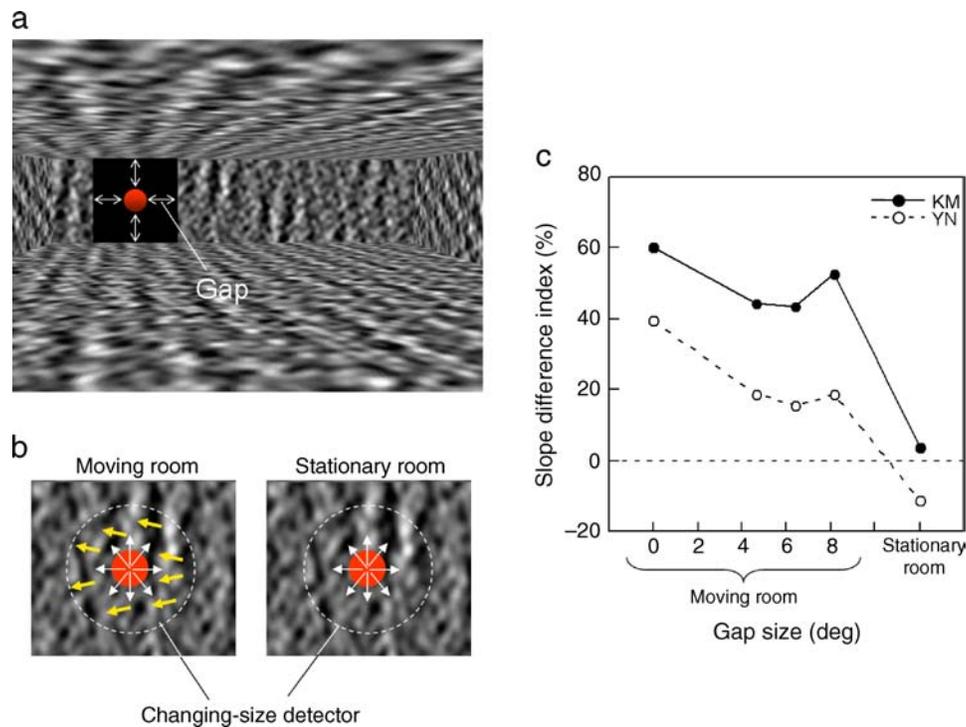
Figure 8. Stimuli and results of Experiment 3. (a) We presented a black frontal plane behind the target (red ball). We varied the gap between the edges of the black plane and the target. The center of the plane kept the same location as the center of the target, even when the target moved along a 3D trajectory. However, the size of the plane remained constant during the target motion. In this experiment, we used the same target motion trajectories and the same motion of the virtual moving room as those in Experiment 2. (b) Effect of the flow of the textured room on the changing-size detector. The left and right panels represent the moving-room and stationary-room conditions, respectively. The white dotted circle represents the receptive field of a changing-size detector. The red sphere represents the target. The white arrows indicate the target expansion. The orange arrows indicate the flow of the textured room when the virtual room moves. In the moving-room condition, both the target expansion and the flow of the textured room stimulate the changing-size detector. (c) The graph shows the slope difference index as a function of gap sizes. Symbols represent different observers. The data on the right indicate the slope difference index in the stationary-room condition from Experiment 2.

possible that the difference in performance between the moving-room and the stationary-room conditions is caused by the local changing-size detector (see Figures 6b and 6c). If this is true, performance in the moving-room condition would be equal to that in the stationary-room condition when introducing a small gap of more than 2 deg between the outer edge of the target and the inner edge of the flow pattern. This is because introducing the gap removes the flow of the moving room from the local changing-size detector, resulting in the stimulation of the target expansion alone to the local changing-size detector even for the moving-room condition (Gray & Regan, 2000; Warren & Rushton, 2007, 2008).

## Methods

In Experiment 3, we used the same apparatus and general stimulus form as in Experiment 2, but we added a frontal black plane in front of a frontal textured wall for the moving-room condition (see Figure 8a). We varied the gap between the edges of the black plane and target, as

illustrated in Figure 8a. The center of the black plane stayed at the same location as the center of the target, even when the target moved along a 3D trajectory. However, the size of the plane remained constant during the target motion. In the moving-room condition, we used four gap sizes: 0 deg, 4.6 deg, 6.4 deg, and 8.2 deg. Observers performed a task only for the moving-room condition with changes in the gap size. The observers' task was to indicate the depth position where the target would arrive by using the vertical pole. We calculated the slope difference index of the apparent depth positions for each gap size. For data analysis, we used the slope difference index of the results of the stationary-room and the dark-room conditions obtained in Experiment 2 for each observer. Two observers participated in this experiment. They also participated in Experiment 2.

## Results and discussion

Figure 8c shows the slope difference index as a function of gap size with the index of the stationary-room

condition for two observers. The symbols represent the two different observers. As shown in Figure 8c, the indexes in the moving-room condition for all gap sizes were larger than the index in the stationary-room condition. Thus, the effect of the optic flow pattern on the perceived direction of 3D object motion does not seem to be caused only by the local changing-size detector. This is consistent with the previous findings of Gray and Regan (2000), Gray et al. (2004), and Warren and Rushton (2007, 2008).

Alternatively, the presence of a gap isolating the target expansion from the background might reduce local expansion cues for the target, resulting in a difference in the performance between the moving-room and the stationary-room conditions. A textured relative motion boundary around the target was included in the case of the stationary room but not in the case of the moving room with the gap. However, if the presence of a gap reduces local expansion cues for the target, the reduction in local expansion cues should appear even in the case of the dark room in Experiment 2. In this situation, one would predict that the observers' responses in the moving-room condition with the gap are the same as those in the dark-room condition in Experiment 2. As a consequence, the slope difference index should be 0% in the case of the moving room with the gap (see the definition of the slope difference index in Results and discussion section). However, as shown in Figure 8c, the slope difference indexes were larger than 0% in the case of the moving room with the gap. This suggests that the reduction in local expansion cues due to the presence of a gap could not explain the difference in performance between the moving-room and the stationary-room conditions.

# General discussion

The present study reveals that, when optic flow is provided by using an adequately large display, stationary observers use a world-centered frame of reference, instead of an observer-centered frame of reference, to perceive 3D structure from motion and to judge the direction of 3D object motion. The effect of optic flow on the judgment of 3D object motion could not be explained by only the local motion detectors that assess the change of retinal size of the target, which is consistent with the findings of Gray and Regan (2000), Gray et al. (2004), and Warren and Rushton (2007, 2008). Thus, these findings suggest that visually guided self-motion from optic flow induces the world-centered perception of 3D object motion.

Rushton and Warren (2005) have proposed that optic flow processing divides retinal motion signals into components due to self-motion and those due to object movements. Warren and Rushton (2007, 2008) provided evidence that flow-parsing is implicated in stationary observers' estimation of the trajectory of 2D object motion during visually guided self-motion from optic flow. Our results are consistent with the flow-parsing account in Rushton and Warren (2005). In Experiment 1, the flow pattern of the target consisted of a combination of forward self-motion and the surface rotating around the horizontal axis. This combination distorted the flow pattern of the rotating surface in the retinal image (see Figure 2b). However, observers perceived the surface rotating around the horizontal axis when optic flow simulating forward self-motion was presented in the peripheral visual field. This suggests that the total retinal motion is decomposed into self- and object-motion components in perceiving 3D structure from motion, which compensates for retinal motion due to self-motion. In Experiment 2, the target moved along a 3D trajectory, and the trajectory angle varied relative to the body of observers. However, when presenting optic flow simulating forward self-motion in the background, the perceived direction of the target was constant across the presented trajectory angles. In this case, the depth component of the target motion was the same as the moving velocity of the simulated forward self-motion (see Figure 5b). This result indicates that the component of retinal motion due to self-motion is subtracted from the total retinal motion in perceiving 3D motion trajectory, which compensates for retinal motion due to self-motion. Thus, the present study extends the flow-parsing account of Rushton and Warren from 2D object motion to 3D object motion.

Gray et al. (2004) found that, for stationary observers, the perceived direction of object motion in depth shifted toward the focus of expansion of the optic flow pattern presented in the peripheral visual field. The findings of Gray et al. (2004) can be explained by the idea that the visual system uses the world-centered frame of reference for judgment of object motion during visually simulated self-motion. Figure 9a shows an example of the stimulus used by Gray et al. In Figure 9a, the purple square provides the expanding motion that simulates an object approaching the left side of the observer's body, and the black squares provide the optic flow pattern that simulates forward self-motion. Figure 9b shows the top view of the situation simulated in Figure 9a. In Figure 9b, the red arrow represents the simulated direction of the target motion. As illustrated in Figure 9b, the $x$ and $z$ components of motion are considered for the target motion. If the observer uses a world-centered frame of reference to estimate the 3D motion direction of the target, the motion of the black squares for the simulated forward self-motion is deducted from the $z$ component of the target motion. As a result, a new $z'$ component of the target motion is produced, and the new direction of the target motion is produced by the $x$ and $z'$ components of the target motion (Figure 9c). As illustrated in Figure 9c, therefore, the perceived direction of the target motion is shifted toward the focus of expansion of the optic flow pattern.
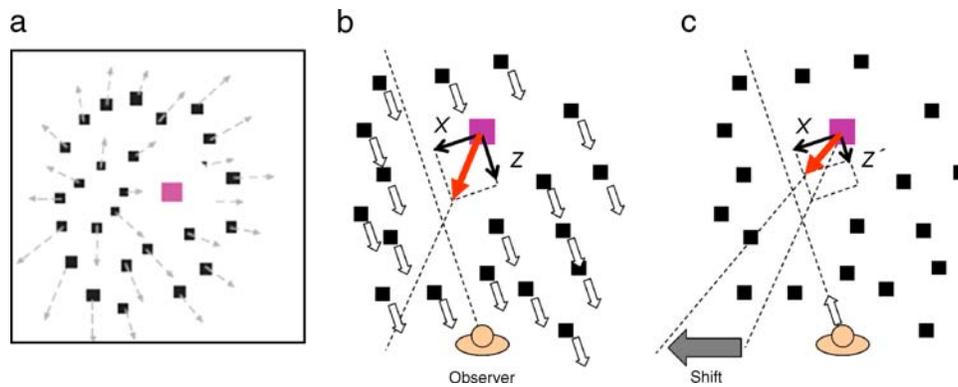
Figure 9. Explanation for the findings of Gray et al. (2004) by using a world-centered frame of reference during simulated self-motion. (a) An example of the stimulus used by Gray et al. (2004). The purple square was the target. The motion of the black squares provided the optic flow pattern of forward self-motion. (b) Top view of the situation simulated by Gray et al. The red arrow represents the simulated direction of the target motion in depth. The black arrows represent the *x* and *z* components of motion for the target motion. (c) If the simulated self-motion triggered the use of a world-centered frame of reference to estimate the target trajectory, the perceived direction of the target motion would be shifted toward the focus of expansion of the optic flow pattern, as illustrated in the figure (see the text for details).

It has been shown that voluntary self-motion induces the world-centered perception of 3D object motion (Wexler, 2003; Wexler, Lamouret et al., 2001; Wexler, Panerai, Lamouret, & Droulez, 2001). This suggests that head movements given by voluntary self-motion are needed to induce the world-centered perception of 3D object motion. However, the present study reveals that visually simulated self-motion without head movements can also induce the world-centered perception of 3D object motion. The present study used a considerably large visual field in order to present an optic flow pattern. Therefore, it is possible that the large visual field might be required to induce the world-centered perception of 3D object motion during visually simulated self-motion without head movements. However, recent studies have shown that even though a relatively small field of view was used, the presence of optic flow, which simulated self-motion without head movements, changed the perceived trajectory of a probe (Rushton & Warren, 2005; Warren & Rushton, 2007, 2008). This suggests that the size of the optic flow pattern may not be necessarily important to induce the world-centered perception of 3D object motion during visually simulated self-motion.

It has recently been reported that observers use visual direction to judge the direction of 3D object motion (Harris & Drga, 2005). Harris and Drga (2005) found that the visual direction strategy causes large systematic errors for estimates of the direction of 3D object motion. This strategy is not very useful for collision achievement or avoidance in our everyday life. Using such a strategy raises the question of how we accurately interact with 3D moving objects in the real environment. The present study seems to provide one answer for such a question. That is, using a world-centered frame of reference might not cause large systematic errors, whereas using an observer-centered frame of reference, as in visual direction, does.

In fact, we found that the perceived direction of 3D object motion tended to be more accurate when self-motion was simulated than when no self-motion was simulated (see Figure 6). Thus, the world-centered perception of 3D object motion may escape the critical problem of the visual direction strategy for collision achievement or avoidance.

Rushton and Duke (2007) also found systematic errors in judgments concerning the trajectory of 3D object motion and found that the error pattern is not explained by Harris and Drga's model. On the basis of these results, Rushton and Duke suggested that observers did not use visual direction to judge the trajectory of 3D object motion in their experiment. This implies that even though observers do not use the visual direction strategy for estimating the trajectory, systematic errors occur in judgments of trajectory. However, both Rushton and Duke's and Harris and Drga's experiments presented the target against a dark background. Therefore, in both the experiments, the observers had to use an observer-centered frame of reference to judge the trajectory. Note that the situations used in both experiments are similar to the dark-room condition in Experiment 2 of our study. Thus, using an observer-centered frame of reference may cause systematic errors in trajectory perception.

The manner in which observers responded in this study was different from that in Harris and Drga's experiment. In Harris and Drga's experiment, observers were asked to reproduce the object's trajectory by adjusting an arrow in front of them. In our study, the observers judged the position at which the target would arrive in their sagittal plane. Although the two experiments employed different methods to obtain observers' responses, both these studies found errors in trajectory estimation. Thus, it appears that the errors in trajectory estimation do not depend on the manner in which observers respond. Poljac, Neggers, and

van den Berg ([2006](#)) asked observers to point to the locations of the intersection with the plane of regard by the extrapolation of the perceived trajectory of an approaching object. They also found biased judgments in the pointing task. The method used in Experiment 2 of our study was similar to that used in Poljac et al.

In the present study, the virtual moving room was used to produce an optic flow pattern (Figure 3a). One possibility is that the room stimulus might invoke a sort of high-level/top-down mechanism in perceiving 3D object motion. According to this account, the observers know that rooms do not tend to move. As a result, such a top-down bias might contribute to the world-centered interpretation of 3D object motion. In future investigations, it would be informative to examine whether the top-down bias affects the world-centered interpretation of 3D object motion.

Neurons in area MST are sensitive to the flow pattern of motion for monkeys (Bradley, Maxwell, Andersen, Banks, & Shenoy, [1996](#); Duffy & Wurtz, [1991a](#), [1991b](#), [1995](#); Saito et al., [1986](#); Tanaka, Fukuda, & Saito, [1989](#); Tanaka & Saito, [1989](#)) as well as humans (Morrone et al., [2000](#)), and are also sensitive to motion-in-depth of a moving object (Sakata, Kusunoki, & Tanaka, [1993](#)). Furthermore, it has recently been reported that visual tracking neurons in MST represent object motion in world-centered coordinates (Ilg, Schumann, & Thier, [2004](#)). These neural mechanisms may contribute to the creation of the world-centered representation of 3D object motion during visually guided self-motion. This representation may provide accurate perception of 3D object motion during self-motion.

## Acknowledgments

## References

Anderson, G. J., & Braunstein, M. L. (1985). Induced self-motion in central vision. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 122–132. [PubMed]

Beverley, K. I., & Regan, D. (1982). Adaptation to incomplete flow patterns: No evidence for 'filling-in' the perception of flow patterns. *Perception, 11,* 275–278. [PubMed]

Bradley, D. C., Maxwell, M., Andersen, R. A., Banks, M. S., & Shenoy, K. V. (1996). Mechanisms of heading perception in primate visual cortex. *Science, 273,* 1544–1547. [PubMed]

Brandt, T., Dichgans, J., & Koenig, E. (1973). Differential effects of central verses peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research, 16,* 476–491. [PubMed]

Brenner, E. (1993). Judging an object's velocity when its distance changes due to ego-motion. *Vision Research, 33,* 487–504. [PubMed]

Brenner, E., & van den Berg, A. V. (1996). The special role of distant structures in perceived object velocity. *Vision Research, 36,* 3805–3814. [PubMed]

Crowell, J. A., & Banks, M. S. (1996). Ideal observer for heading judgments. *Vision Research, 36,* 471–490. [PubMed]

Duffy, C. J., & Wurtz, R. H. (1991a). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology, 65,* 1329–1345. [PubMed]

Duffy, C. J., & Wurtz, R. H. (1991b). Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *Journal of Neurophysiology, 65,* 1346–1359. [PubMed]

Duffy, C. J., & Wurtz, R. H. (1995). Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *Journal of Neuroscience, 15,* 5192–5208. [PubMed] [Article]

Gibson, J. (1950). *Perception of the visual world.* Boston: Houghton Mifflin.

Gogel, W. C. (1990). A theory of phenomenal geometry and its applications. *Perception & Psychophysics, 48,* 105–123. [PubMed]

Gray, R., Macuga, K., & Regan, D. (2004). Long range interactions between object-motion and self-motion in the perception of movement in depth. *Vision Research, 44,* 179–195. [PubMed]

Gray, R., & Regan, D. (2000). Simulated self-motion alters perceived time to collision. *Current Biology, 10,* 587–590. [PubMed] [Article]

Harris, J. M., & Drga, V. F. (2005). Using visual direction in three-dimensional motion perception. *Nature Neuroscience, 8,* 229–233. [PubMed]

Howard, I. P., & Heckmann, T. (1989). Circular vection as a function of the relative sizes, distances, and positions of two competing visual displays. *Perception, 18,* 657–665. [PubMed]

Ilg, U. J., Schumann, S., & Thier, P. (2004). Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron, 43,* 145–151. [PubMed] [Article]

Jaekl, P., Zikovitz, D. C., Jenkin, M. R., Jenkin, H. L., Zacher, J. E., & Harris, L. R. (2005). Gravity and perceptual stability during translational head movement on earth and in microgravity. *Acta Astronautica, 56,* 1033–1040. [PubMed]

Li, L., & Warren, W. H., Jr. (2000). Perception of heading during rotation: Sufficiency of dense motion parallax and reference objects. *Vision Research, 40,* 3873–3894. [PubMed]

Morrone, M. C., Tosetti, M., Montanaro, D., Fiorentini, A., Cioni, G., & Burr, D. C. (2000). A cortical area that responds specifically to optic flow, revealed by fMRI. *Nature Neuroscience, 3,* 1322–1328. [PubMed]

Poljac, E., Neggers, B., & van den Berg, A. V. (2006). Collision judgment of objects approaching the head. *Experimental Brain Research, 171,* 35–46. [PubMed] [Article]

Regan, D., & Beverley, K. I. (1979). Visually guided locomotion: Psychophysical evidence for a neural mechanism sensitive to flow patterns. *Science, 205,* 311–313. [PubMed]

Royden, C. S., Banks, M. S., & Crowell, J. A. (1992). The perception of heading during eye movements. *Nature, 360,* 583–585. [PubMed]

Royden, C. S., Crowell, J. A., & Banks, M. S. (1994). Estimating heading during eye movements. *Vision Research, 34,* 3197–3214. [PubMed]

Rushton, S. K., & Duke, P. A. (2007). The use of direction and distance information in the perception of approach trajectory. *Vision Research, 47,* 899–912. [PubMed]

Rushton, S. K., & Warren, P. A. (2005). Moving observers, relative retinal motion and the detection of object movement. *Current Biology, 15,* R542–R543. [PubMed] [Article]

Saito, H., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., & Iwai, E. (1986). Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience, 6,* 145–157. [PubMed] [Article]

Sakata, H., Kusunoki, M., & Tanaka, Y. (1993). *Neural mechanisms of perception of linear and rotary movement in depth in the parietal association cortex of the monkey.* Oxford, UK: Oxford University Press.

Tanaka, K., Fukada, Y., & Saito, H. A. (1989). Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology, 62,* 642–656. [PubMed]

Tanaka, K., & Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology, 62,* 626–641. [PubMed]

Tcheang, L., Gilson, S. J., & Glennerster, A. (2005). Systematic distortions of perceptual stability investigated using immersive virtual reality. *Vision Research, 45,* 2177–2189. [PubMed]

van den Berg, A. V. (1992). Robustness of perception of heading from optic flow. *Vision Research, 32,* 1285–1296. [PubMed]

Wallach, H. (1987). Perceiving a stable environment when one moves. *Annual Review of Psychology, 38,* 1–27. [PubMed]

Warren, P. A., & Rushton, S. K. (2007). Perception of object trajectory: Parsing retinal motion into self and object movement components. *Journal of Vision, 7*(11):2, 1–11, http://journalofvision.org/7/11/2/, doi:10.1167/7.11.2. [PubMed] [Article]

Warren, P. A., & Rushton, S. K. (2008). Evidence for flow-parsing in radial flow displays. *Vision Research, 48,* 655–663. [PubMed]

Warren, W., & Hannon, D. (1988). Direction of self-motion is perceived from optical flow. *Nature, 336,* 162–163.

Warren, W. H., Jr., Kay, B. A., Zosh, W. D., Duchon, A. P., & Sahuc, S. (2001). Optic flow is used to control human walking. *Nature Neuroscience, 4,* 213–216. [PubMed]

Warren, W. H., Jr., Morris, M. W., & Kalish, M. (1988). Perception of translational heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 646–660. [PubMed]

Welchman, A. E., Tuck, V. L., & Harris, J. M. (2004). Human observers are biased in judging the angular approach of a projectile. *Vision Research, 44,* 2027–2042. [PubMed]

Wexler, M. (2003). Voluntary head movement and allocentric perception of space. *Psychological Science, 14,* 340–346. [PubMed]

Wexler, M., Lamouret, I., & Droulez, J. (2001). The stationarity hypothesis: An allocentric criterion in visual perception. *Vision Research, 41,* 3023–3037. [PubMed]

Wexler, M., Panerai, F., Lamouret, I., & Droulez, J. (2001). Self-motion and the perception of stationary objects. *Nature, 409,* 85–88. [PubMed]