Pergamon

0042-6989(94)E0032-G

Vision Res. Vol. 34, No. 18, pp. 2331–2336, 1994

0042-6989/94 $7.00 + 0.00

The Visual Perception of Three-Dimensional Shape from Self-Motion and Object-Motion*

V. CORNILLEAU-PÉRÈS,† ‡ J. DROULEZ†

Received 25 October 1993; in revised form 7 February 1994

To evaluate the influence of egomotion on the three-dimensional visual processing of structure-from-motion (SFM), we compared the visual discrimination between planar and spherical surfaces during subject-translation, object-translation, or rotation of the object in depth. Performance was the best for object-rotation, intermediate for subject-translation, and the poorest for object-translation—and thus increased with the quality of retinal image stabilization achieved in the different conditions. This suggests that the major role of self-motion information was to stabilize retinal images. In view of previous results, we propose that the interactions between self-motion information and SFM are reduced to functional complementarity, in the sense that self-motion can lift visual ambiguities but does not improve the sensitivity of SFM processes.

Structure-from-motion Self-motion Object-motion

INTRODUCTION

The pattern of retinal motion (or optic flow) due to translations of the eye in space provides rich information about the three-dimensional (3D) layout of the environment (Gibson, 1950; Rogers & Graham, 1979). Similarly, when a rigid object does not only rotate around the eye but also translates, its 3D shape can be identified readily (e.g. Cornilleau-Pérès & Droulez, 1989). The two situations (self-motion, object-motion) are equivalent in the sense that motion parallax, defined as the variations of image velocities over the retina, is used as a depth cue by the visual system. However, in the case of self-motion extraretinal signals of different origins (vestibular, proprioceptive, efferent copies of motor commands . . .) are known to interact with visual information and contribute to the perception of head and eye movements (for reviews see Dichgans & Brandt, 1978; Howard, 1982; Droulez & Darlot, 1989). The theoretical analysis shows that processing the 3D structure of rigid objects from visual motion is a non-linear problem which has received no simple resolution up to now, but can be simplified if self-motion information is provided (e.g. Aloimonos, Weiss & Bandyopadhyay, 1988). In particular, Longuet-Higgins and Prazdny (1980) showed that if the translation direction is given, the problem becomes linear and rather easy to solve. Therefore, by interacting with visual processes, self-motion information from various sources could contribute to improve the visual perception of 3D shapes. Indeed it was reported that

(i) the amount of perceived depth is higher for subject-motion than for object-motion (Rogers & Graham, 1979; Ono & Steinbach, 1990);
(ii) self-motion information is effective in disambiguating the sign of relative depth within a simulated surface (Rogers & Rogers, 1992).

In these studies the 3D variables (amount of depth, sign of relative depth) could not be determined uniquely from the optic flow. Some additional information was needed, related to the object proximity (for the amount of depth) and motion direction (for the sign of depth) respectively. Therefore results (i) and (ii) reveal the ability of the visual system to use self-motion signals to lift ambiguities regarding 3D shape. The question arises, however, whether the processing of visual motion can cooperate with self-motion information, in the sense of an increase of sensitivity and robustness to noise. To address this question, we took advantage of the fact that the planarity of a translating surface can be uniquely specified from motion parallax (Droulez & Cornilleau-Pérès, 1990), and we used a task of discrimination between planar and curved surfaces, for which subjects showed good accuracy during object-motion (Cornilleau-Pérès & Droulez, 1989). Performance for object-motion and subject-motion were compared under identical relative translations between the observer and the object.

Beside a possible direct improvement of the visual perception of 3D shapes, self-motion is likely to have an indirect influence on visual perception through eye

*A brief report of this work appeared previously (Cornilleau-Pérès & Droulez, 1992).

†Laboratoire de Physiologie de la Perception et de l'Action, UMR-C 9950, CNRS-Collège de France, 15 rue de l'Ecole de Médecine, 75270 Paris cedex 06, France.
‡To whom all correspondence should be addressed.

movements. Due to the interplay of different reflexes (vestibulo-ocular ...) retinal image stabilization is far better during head movements than during target movements (Buizza, Léger, Droulez, Berthoz & Schmid, 1980; Baloh, Beykirch, Honrubia & Yee, 1988; Bronstein & Gresty, 1988). Therefore, the improvement found in the condition "self-motion" by previous researchers may have been due to a difference in the retinal image stabilization. In order to separate the possible direct and indirect influences of self-motion information, we designed *two different conditions "object-motion"*. In the first the object-motion was exactly equal to the head-motion performed in the condition "subject-motion" (a pure translation). In the second we simulated a perfect retinal stabilization by combining the previous translation with a rotation around the viewer's eye, so that no eye movements were required to fixate the stimulus. In this second condition the resulting movement was close to a pure rotation of the surface around one of its own tangents. Hence we compared three conditions: subject-translations, object-translations and object-rotations. Prior to the experiment, we describe the geometry of the 3D movements involved in the three conditions, and show how the depth information available from the optic flow was kept strictly identical in the three conditions.

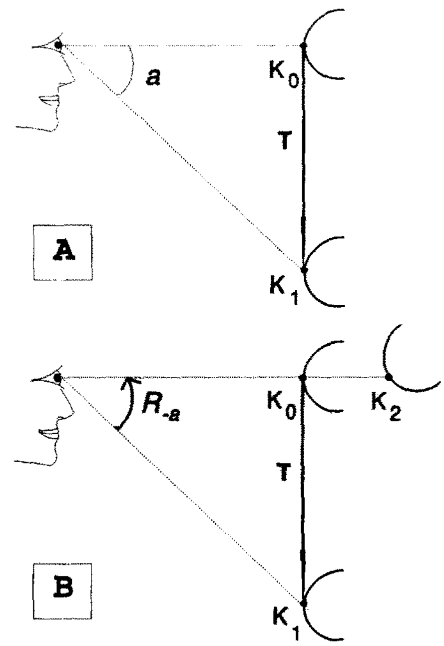

FIGURE 2. The geometry of object movements. (A) In condition OT, the subject is viewing a surface translating with a translation T. Point K moves from position K_0 to position K_1. The retinal image is perfectly stabilized if the eye rotates with an angle a around an axis perpendicular to the plane of the figure. (B) In condition OR, a rotation of angle $-a$, R_{-a}, is combined with the translation T. The resulting movement is a rotation around point K, plus a translation in depth (point K moves from K_0 to K_2). The motion parallax information is exactly the same as in condition OT, but point K remains apparently stationary in front of the subject.

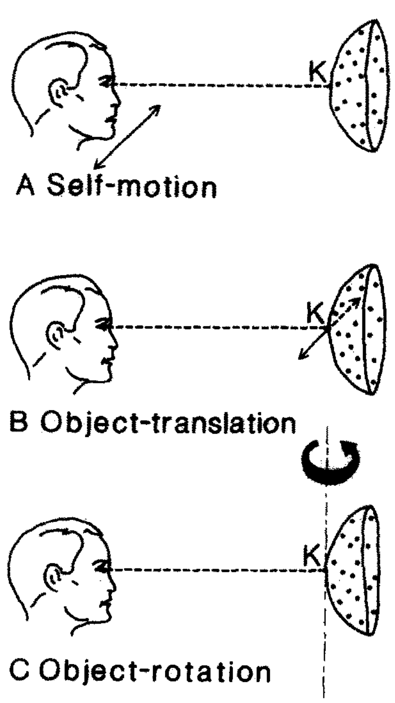

FIGURE 1. Experimental conditions. (A) Condition SM (self-motion). Head-translations are transmitted to the graphic station on which is simulated the motion parallax due to a dotted surface. (B) Condition OT (object-translation). The head-translations recorded in condition SM are used to translate the surface in front of the stationary subject. (C) Condition OR (object-rotation). The translations used in condition OT are combined with rotations around the subject's eye. The resulting surface movement is a rotation around point K.

METHODS

Description of the 3D movements

The relative movement between the eye and an object is usually decomposed into a rotation around the eye and a translation. The translation component (see, e.g. Longuet-Higgins & Prazdny, 1980) yields motion parallax information (i.e. the relative velocity between image points depends on their relative distances from the eye). By contrast, rotations of the object around the eye (or rotations of the eye around itself) only displace the image over the retina as a whole and are devoid of any depth information.

Consider an object translating in front of a subject [Fig. 2(A)] with a translation T. In order to stabilize the image on the retina, the eye should rotate in the plane of the figure, by an angle a. As stated above, this rotation does not modify the depth information available in the optic flow. Similarly, if the object now rotates by an angle $-a$ in the figure plane, the depth information remains unchanged. In this case [Fig. 2(B)], the situation is exactly the same as in Fig. 2(A), except that a perfect ocular pursuit of the object is simulated. The resulting movement is a rotation of the object in depth combined with a small translation in depth. We call this condition "object-rotation". It yields exactly the same depth information as the pure object-translation T, but no eye movements are required to stabilize the image on the retina.

The three conditions that we compared were therefore:

—subject-translations
—object-translations
—object-rotations.

We recorded the subject-translations in the first condition, and used this recording to move the object in the two last conditions. This insured that the relative translation between the eye and the object (a spherical or a planar surface) was kept strictly identical in the three conditions. These translations were combined with rotations around the subject's eye in the last condition, as explained in Fig. 2.

Apparatus

During self-motion, a long thin bar had one tip fixed on a light helmet worn by the subject, and was mobile in a system of pulleys. Three potentiometers adjusted to these pulleys transformed into voltages the three coordinates of head position. These signals were transmitted to a graphic station which generated the visual stimuli. The absolute error on head position averaged 1.7 mm over a parallelepiped of 30 cm (horizontal) $\times$ 20 cm (vertical) $\times$ 6 cm (depth) centred on the median subject's head position, and never exceeded 5 mm. The images were updated at a rate of 30 Hz (the refreshment rate of the monitor is 60 Hz), and head position was therefore sampled at this rate. The delay in the feedback-loop was 55 msec. The stimuli were displayed on a 1024×1280 monitor, where the image area is 27.5×34.4 cm.

Stimuli

The visual stimuli were spherical (with convexity facing the subject) or planar patches covered with 400 dots (pixel size 0.02 deg). Their images were calculated under polar projection, with the centre of projection equating the position of the observer's eye.

In its initial position the surface was tangent to a frontoparallel plane in a point K positioned in the screen centre (Fig. 1). The dots were randomly spread around K within a radius of 4 deg, with uniform density. The distance between the observer and the display screen was 72 cm. In order to weaken the correlation that exists between dot speed relative to the screen and curvature, the simulated viewing distance (between observer's eye and point K) was randomly chosen between 75 and 85 cm.

Subjects

Four subjects with normal uncorrected vision, aged 25–31 yr, participated. One was the first author, and three were naive paid volunteers.

Procedure

The subject was sitting and had to indicate whether a surface appeared as planar or curved under three conditions (Fig. 1).

—*SM* (*self-motion*): the subject translated actively his head horizontally or vertically, at the pace of metronome clicks (0.5 Hz) and viewed the simulated stationary surface presented on a display screen.
—*OT* (*object-translation*): the subject was stationary (head fixed in a chinrest). The simulated surface moved with the translations recorded in condition SM.
—*OR* (*object-rotation*): similar to OT, except that the surface translations were systematically composed with rotations around the subject's eye so that the central image point K remained stationary. The resulting movement was a rotation around a frontoparallel axis tangent to the surface, and mimicked the visual stimulation in condition OT *if the ocular pursuit of point K was perfect*.

While viewing the stimuli, the subject was asked to fixate as accurately as possible a target of diameter 0.05 deg located in the centre of the surface patch, in point K.

Each trial consisted of three blocks (one for each condition) of 48 image sequences (24 spheres of constant curvature and 24 planes in random order) lasting 6 sec each. Since condition SM was always performed before OT and OR, three trials were repeated to determine the possible role of fatigue or learning. Conditions OT and OR were performed in random order. Viewing was monocular (each eye was alternatively covered during successive trials), and the experimental room was dark. Three naive subjects and the first author participated.

For each block of 48 image sequences we calculated the percentage of correct responses. Prior to the experiment, training sessions were performed in order to accustom the subjects with the task, and to determine the curvature of the simulated sphere yielding about 75% correct responses for condition SM.

All the spherical surfaces had their convexity facing the subject. During training sessions, the subject sometimes perceived them as concave, and deforming. However, with the indication that the spherical surfaces were convex, this perception usually disappeared with practice.

RESULTS

Figure 3 reports the percentages of correct responses (PCR) measured in the three conditions. The differences between the three conditions are similar across subjects, and across trials, indicating little or no influence of fatigue and learning. The fact that the distributions of performance did not differ across trials was verified with a non-parametric ANOVA based on Friedman's test (Krauth, 1988, p. 286).

First the visual sensitivity to surface curvature was far higher in condition SM than in condition OT, although the relative movement between the subject's head and the object were strictly identical. Second, for *object-motion*, the performance were highly dependent on the level of stabilization of the image on the screen.

Curvature sensitivity was optimal when the image was perfectly stabilized (condition OR), whereas responses were close to chance level for pure object translations (condition OT). Overall, the performance was close to perfection in condition OR (average PCR: 91% in horizontal; and 93% in vertical), close to threshold in condition SM (76% and 73%), and close to chance level in condition OT (52% for both directions of movement). The difference between conditions was assessed by using a non-parametric test, the Wilcoxon matched-pairs signed-ranks test (Krauth, 1988), in two cases:

- for each direction of movement we gathered the responses from all subjects and found differences that were significant to the level $P < 0.001$ between conditions;
- for each subject, we gathered the results obtained for the two directions of movement and found a significant difference (to the level $P < 0.05$) between conditions.

Image velocity as a possible bias in our paradigm

Concerning our psychophysical task, the question arises whether subjects could base their responses on artifacts such as the average dot speed rather than on perceived curvature. Dot positions and velocities relative to point K, as seen from the subject's eye, are identical in the three conditions. Therefore, artifacts linked to the 2D repartition and movement of dots were of same potential strength in the three conditions, but should be the easiest to use in condition OR yielding the best retinal image stabilization. So we restrict the discussion to this condition. A previous paper (Cornilleau-Pérès & Droulez, 1993) gave a series of arguments (including control experiments) showing that, in condition OR, subjects report on perceived 3D structure, and do not use other cues incidently related to surface curvature. Since the motion and curvature parameters differ in the present experiment, we verified experimentally that subjects did not base their responses on the 2D dot speed. Dot speed averages higher for spheres than for planes

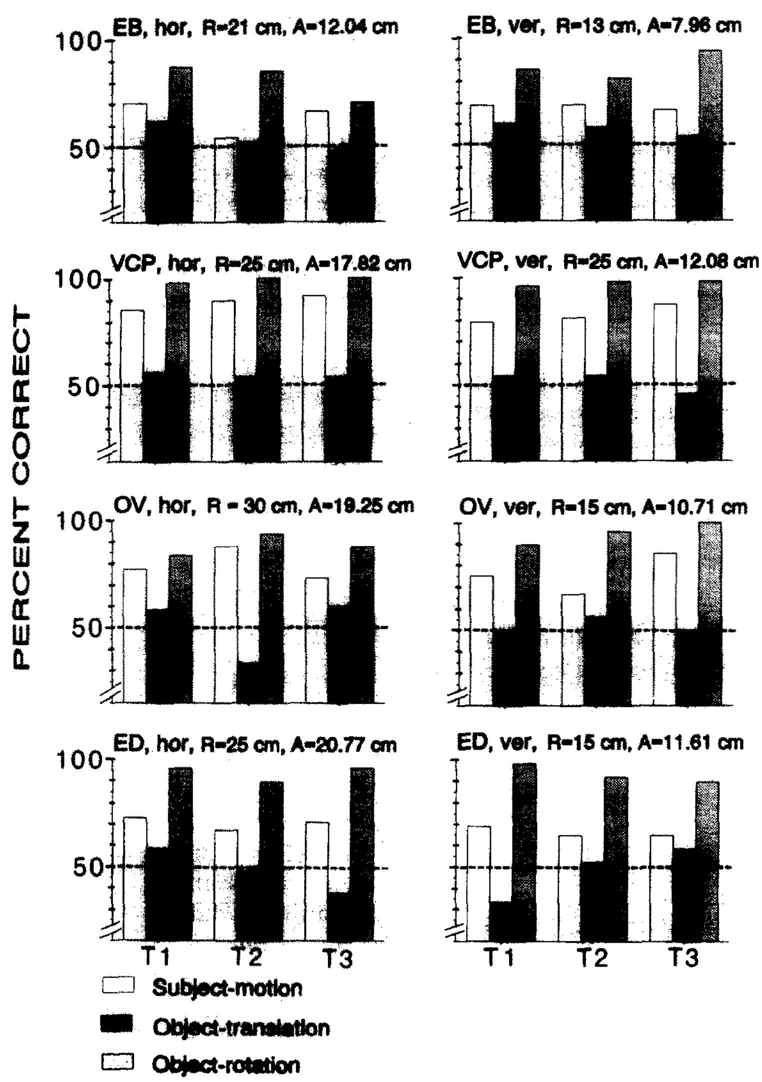

FIGURE 3. Results. The percentage of correct responses in sphere-plane discrimination is plotted for three successive trials (T1, T2, T3) for four subjects and two motion-directions (left, horizontal; right, vertical). R is the radius of the spherical surface, A is the total average extent of head-translations.

TABLE 1. Control experiment: average dot-speed as a possible bias for curvature detection

	Average dot speed (deg/sec)			PCR (%)	
Subject	Plane cond OR	Sphere cond OR	Plane or sphere cond ORc	Cond OR	Cond ORc
VCP	0.080	0.100	0.092	99.3	98.6
OV	0.088	0.110	0.101	91.7	92.4

Columns 1 and 2: dot speed relative to the screen as averaged over three trials (condition OR, horizontal motion) for two subjects (VCP and OV), for spheres and planes. Column 3: average dot speed for spheres and planes in condition ORc. Columns 4 and 5: percentage of correct responses in conditions OR and ORc (see text) for the two subjects.

(Table 1). Hence we designed a condition ORc similar to OR except that dot speed was not correlated to surface curvature. For each image-sequence motion amplitude was multiplied by a factor g randomly chosen in an interval $[1-a, 1+a]$, a being such that the average dot speed on the screen was equal for planes when the movement is amplified (g chosen in $[1, 1+a]$) and for spheres when the movement is reduced (g chosen in $[1-a, 1]$). For both subjects a was taken as 0.2. Table 1 (column 3) gives the corresponding dot speed. Only the stimuli falling in those two categories finally served in determining the subject's performance. Columns 4 and 5 of Table 1 show no noticeable decrease of performance from condition OR to condition ORc, thus confirming that dot speed is not an artifact of our paradigm.

DISCUSSION

The decrease of performance observed systematically from condition OR to condition SM contradicts the assertion that self-motion information contributes to improve the processing of 3D structure-from-motion. However, this conclusion should be taken cautiously because the visual stimulation was not strictly equivalent in the two conditions. The spatial variations in image velocity were identical, but not the global retinal motion or retinal slip, which we define here as the retinal image velocity of point K. In condition OT, the retinal slip depends on the accuracy of ocular pursuit. Point K follows a quasi-unpredictive trajectory (only the overall direction and timing of the motion remained constant) with a speed reaching 15–40 deg/sec. For periodic and non-periodic tracking Lisberger, Evinger, Johanson and Fuchs (1981) find that the residual retinal slip depends essentially on maximum target acceleration. For accelerations of 100 deg/sec^2, close to what occurs in our experiments, the retinal slip ranged between 5 deg/sec (periodic movements), and 15 deg/sec (aperiodic movements). This range is also close to what is predicted with a gain of 0.6 for smooth pursuit at 0.5 Hz, as found in the vertical or horizontal directions for pseudo-random target motion by Collewijn and Tamminga (1984).

Several studies (Buizza *et al.*, 1980; Kasai, 1987; Baloh *et al.*, 1988; Bronstein & Gresty, 1988; Paige, 1989) documented the improvement of retinal image stabilization during head-translations, as compared to target-translations, both for phasic motion components (decrease of latencies and time to match target velocities) and for the gain of smooth tracking. In particular Buizza *et al.* showed that linear accelerations above 0.1 m/sec^2 yield an increase in the gain of the optokinetic nystagmus, which suggests that in our condition SM where horizontal head accelerations reached 0.74–1.23 m/sec^2, ocular tracking was helped by reflexes such as the otolith-ocular reflex. There exists no measurement of retinal slip during, active head-translations. However, a value of about 1 deg/sec can be proposed on the basis of measurements performed during active head-rotations which involve otolithic stimulations because their axis does not pass through the otolith organs. In particular this is found by Collewijn, Martins and Steinman (1981), and Ferman, Collewijn, Jansen, and van den Berg (1987), for motion parameters which are similar to ours (head-rotations of 10 deg amplitude, frequencies ranging between 0.33 and 1.33 Hz).

Therefore the detection of surface curvature improves with retinal image stabilization across the different conditions. Condition OR clearly yields both the optimal retinal stabilization and optimal curvature detection. Can the decrease of performance in condition SM be due to a retinal slip as small as 1 deg/sec? Acuity and contrast sensitivity are not affected by retinal image-motion up to 2 deg/sec (Westheimer & McKee, 1975; Murphy, 1975). Alternatively, Nakayama (1981) showed that global image displacements as small as 2 min arc decrease the sensitivity to velocity gradients. This result cannot be directly related to our study because viewing durations were lower than 200 msec. However, in similar viewing conditions, Nakayama found a range of common image-motion where differential motion sensitivity was severely degraded, whereas vernier acuity was not. He concluded that the goal of retinal image stabilization through eye movements might be "to ensure the optimal pick up of motion-parallax information". This conclusion supports our present interpretation, namely that retinal slip is the major factor influencing performance in curvature detection from object-motion or subject-motion.

Our experiment clearly shows that self-motion information *per se* does not improve the 3D processing of depth from visual motion. Rather, the optimal performance are obtained during the object-motion condition that yields the best retinal image stabilization, namely "object-rotation". Together with the results obtained by Rogers and Graham (1979), Ono and Steinbach (1990) and Rogers and Rogers (1992), our results lead to distinguish between 3D variables that are ambiguously defined by motion parallax (amount of depth, relative depth order), and those uniquely determined from optic flow, such as the planarity of translating surfaces.*

*Note that monocular vs binocular viewing conditions were already shown to affect the perceived depth but not the detection of surface curvature. On one hand the amount of perceived depth of a moving surface is higher in monocular than in binocular viewing in the absence of binocular disparity (Rogers & Collett, 1989). On the other hand, we found (Cornilleau-Pérès & Droulez, 1989) no such difference in the detection of surface curvature.

Self-motion information is able to lift ambiguities on the former, but seems uneffective in improving the processing of the latter. The cooperation between self-motion and the processing of 3D shape-from-motion might therefore be restricted to a functional complementarity, in the sense that self-motion information can be used to disambiguate structure from motion [this type of interaction is termed "disambiguation" by Bülthoff and Mallot (1988)]. Concerning the physiology of cortical interactions between visual motion processing and signals related to self-motion, it suggests that the first steps of the processing of 3D shape-from-motion might be independent from signals related to self-motion.

REFERENCES

Aloimonos, J., Weiss, I. & Bandhopadhyay, A. (1988). Active vision. *International Journal of Computer Vision, 2,* 333–356.

Baloh, R. W., Beykirch, K., Honrubia, V. & Yee, R. D. (1988). Eye movements induced by linear acceleration on a parallel swing. *Journal of Neurophysiology, 60,* 2000–2013.

Bronstein, A. M. & Gresty, M. A. (1988). Short latency compensatory eye movements responses to transient linear head acceleration: A specific function of the otolith-ocular reflex. *Experimental Brain Research, 71,* 406–410.

Buizza, A., Léger, A., Droulez, J., Berthoz, A. & Schmid, R. (1980). Influence of otolithic stimulation by horizontal linear acceleration on optokinetic nystagmus and visual motion perception. *Experimental Brain Research, 39,* 165–176.

Bülthoff, H. H. & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America, 5,* 1749–1758.

Collewijn, H. & Tamminga, E. P. (1984). Human smooth and saccadic eye movements during voluntary pursuit of different target motions on different backgrounds. *Journal of Physiology, 351,* 217–250.

Collewijn, H., Martins, A. J. & Steinman, R. M. (1981). Natural retinal image motion: Origin and changes. *Annals of the New York Academy of Science, 374,* 312–329.

Cornilleau-Pérès, V. & Droulez, J. (1989). Visual perception of surface curvature: Psychophysics of curvature detection induced by motion parallax. *Perception & Psychophysics, 46,* 351–364.

Cornilleau-Pérès, V. & Droulez, J. (1992). Perception of three-dimensional shapes from motion during active and passive vision. *Perception (Suppl. 2), 21,* 55.

Cornilleau-Pérès, V. & Droulez, J. (1993). Stereo-motion cooperation and the use of motion parallax in the visual perception of 3D structure. *Perception & Psychophysics, 54,* 223–239.

Dichgans, J. & Brandt, T. (1978). Visual-vestibular interactions and motion perception. In Held, R., Leibowitz, H. W. & Teuber, H. L. (Eds), *Handbook of sensory pysiology, Vol. VIII perception* (Chap. 8, pp. 755–804). New York: Springer.

Droulez, J. & Cornilleau-Pérès, V. (1990). Visual perception of surface curvature. The spin variation and its physiological implications. *Biological Cybernetics, 62,* 211–224.

Droulez, J. & Darlot, C. (1989). In *Attention and performance* (Vol. XII, Chap. 17, pp. 495–525). Hillsdale, N.J.: Erlbaum.

Ferman, L., Collewijn, H., Jansen, T. C. & van den Berg, A. V. (1987). Human gaze stability in the horizontal vertical and torsional direction during voluntary head movements, evaluated with a three-dimensional scleral induction coil. *Vision Research, 27,* 811–828.

Gibson, J. J. (1950). *The perception of the visual world.* Boston, Mass.: Houghton-Mifflin.

Howard, I. P. (1982). *Human visual orientation.* New York: Wiley.

Kasai, T. (1987). Compensatory eye movements for horizontal linear head accelerations. *Neuroscience (Suppl.), 22,* S731.

Krauth J. (1988). Distribution-free statistics. *Collection: Techniques in the behavioral and neural sciences* (Vol. 2). Amsterdam: Elsevier.

Lisberger, S. G., Evinger, C., Johanson, G. W. & Fuchs, A. F. (1981). Relationship between eye acceleration and retinal image velocity during foveal smooth pursuit in man and monkey. *Journal of Neurophysiology, 46,* 229–249.

Longuet-Higgins, H. C. & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B, 208,* 385–397.

Murphy, B. J. (1978). Pattern thresholds for moving and stationary gratings during smooth eye movements. *Vision Research, 18,* 521–530.

Nakayama, K. (1981). Differential motion hyperacuity under conditions of common image motion. *Vision Research, 21,* 1475–1482.

Ono, H. & Steinbach, M. J. (1990). Monocular stereopsis with and without head movement. *Perception & Psychophysics, 48,* 179–187.

Paige, G. D. (1989). The influence of target distance on eye movement responses during vertical linear motion. *Experimental Brain Research, 77,* 585–593.

Rogers, B. J. & Collett, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *Quarterly Journal of Experimental Psychology, 41A,* 697–717.

Rogers, B. J. & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8,* 125–134.

Rogers, S. & Rogers, B. J. (1992). Visual and non-visual information disambiguate surfaces specified by motion parallax. *Perception & Psychophysics, 52,* 446–452.

Westheimer, G. & McKee, S. P. (1975). Visual acuity in the presence of retinal image motion. *Journal of the Optical Society of America, 65,* 847–850.

Acknowledgements—This work was supported by Espirit Project No. 3149 MUCOM. We thank M. Ehrette for technical assistance and Pr A. Berthoz and Dr J. R. Duhamel for useful comments on the manuscript.