Pergamon

PII: S0042-6989(96)00060-0

*Vision Res.*, Vol. 36, No. 19, pp. 3119–3127, 1996

Printed in Great Britain
0042-6989/96 $15.00 + 0.00

# Non-visual Information in Structure-from-motion

WIM J. M. van DAMME,*† WIM A. van de GRIND*

*Received 26 June 1995; in final form 18 January 1996*

**We examined whether non-visual signals improve visual perception of three-dimensional structure-from-motion. Observers discriminated curvature in quadratic surfaces defined by random dot cinematograms with limited lifetime. They either explored visually a static surface by making head movements that were fed back to the display (HM condition) or they viewed statically the same surface which now rotated (NHM condition). Both conditions showed a clear build-up of performance as lifetime increases, but with different time constants for the HM and NHM condition. A second experiment showed that these differences could not be caused by differences in motion detection for the HM and NHM conditions. We suggest that non-visual information is combined with visual information at a high stage of visual processing, and that it does not mainly serve as input for a retinal stabilization process. Copyright © 1996 Elsevier Science Ltd.**

Structure-from-motion Self-motion Object-motion

## INTRODUCTION

The perception of depth is often regarded as a process depending only on visual input. In many psychophysical experiments concerning visual depth perception, much care is taken to restrict or even avoid movements of the subject's head, e.g. by means of a chin-rest and/or head-rest. The reason for this restriction of head movement is that a visual stimulus simulating a three-dimensional object is only correct from one specific viewing position. Any other viewing position could lead to a different interpretation of the same stimulus that is not necessarily consistent with what the experimenters intend to simulate. It is known, for example, that pilots undergoing training in a flight simulator can see easily that the simulation is not real by making small head movements. Apart from this theoretical inconsistency, it is not clear yet what the perceptual effects of head movements in a three-dimensional task are.

In the animal kingdom, many species use head movements to obtain specific visual information necessary for navigation or hunting. Flying species such as birds and insects use motion parallax to judge distance. For example, landing pigeons can judge "time to contact" from head-bobbing (Green *et al.*, 1994) and a sitting locust can judge distance to prey from lateral head movements (Sobel, 1990). The fact that they rely heavily on motion parallax and not on other ways of gathering this information (e.g. stereopsis) suggests that head movements are advantageous during three-dimensional visual tasks.

Experimental evidence for an effective use of proprioceptive information in three-dimensional vision by human observers is sparse (Ono *et al.*, 1986). Rogers and Graham (1979) showed that active observers could recognise a three-dimensional surface without difficulty by making active head movements. They found no qualitative differences between active judgements (with head movements fed back to the display) and "passive‡" judgements (static observer viewing a dynamic display), but their subjects reported that perceived depth was more pronounced in the self-produced motion parallax condition than in the externally generated parallax condition. However, their experiments were not performed in a dark room, so additional visual information other than the stimulus might have served as a reference in both conditions for the retrieval of depth.

The possibility that observers use additional visual information in the optic array instead of non-visual information to judge or disambiguate depth stimuli has been considered in more detail by Rogers and Rogers (1992). They did not report whether this extra information (both non-visual and visual) affected the depth judgements quantitatively, but it helped in disambiguating the sign of the simulated depth. This was also pointed out by Hayashibe (1991).

Stappers (1992) showed that human observers are not very good at judging the relative depth and the size of

*Department of Comparative Physiology, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

†To whom correspondence should be addressed at: Department of Physiology, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands (Tel +3110 408 7569; Fax +3110 436 7594; *Email* DAMME@FYS1.FGG.EUR.NL).

‡The term passive vision is not entirely correct here. It refers to the situation where an observer is moved, for example in a wheelchair, but does not generate the movement him/herself. In that case, no efferent copy signals are available. Proprioceptive signals still exist in this situation. For this reason, we will refer to the two conditions as the head-movement (HM) and the no head-movement (NHM) condition.

simulated three-dimensional objects if the feedback of ego-movements to the visual stimulus was distorted. In his experiments, the visual information signalled a different self-motion than the proprioceptive information. A similar task with a veridical feedback showed much better performance. This suggests that human observers are able to use proprioceptive information in a quantitative way during visual tasks.

Cornilleau-Pérès and Droulez (1994) compared performance in the detection of three-dimensional curvature in the self-motion condition with two kinds of object motion: object translation (OT) and object rotation (OR). The two kinds of object motions were used because they differ in oculomotor response: it is known that, because of the vestibulo-ocular reflex (VOR), stabilization of gaze is more accurately performed by a moving observer than by a static observer pursuing a moving object (Buizza *et al.*, 1980; Ferman *et al.*, 1987). Cornilleau-Pérès *et al.* found that OT always led to the worst performance of the three conditions, and that OR always had the best performance of the three conditions. They concluded that non-visual information about self-motion is used mainly as a retinal stabilization factor, and that it does not directly improve the processing of depth from motion, van Damme and van de Grind.

On the other hand, Oosterhoff *et al.* (1993) had found that human observers obtained significantly lower just-noticeable differences of three-dimensional curvature when they were allowed to make head movements compared to a static viewing condition with an object rotation similar to the one in Cornilleau-Pérès and Droulez (1994). According to Cornilleau-Pérès and Droulez (1994), performance in such a task should not improve by the addition of non-visual information about self-motion, but clearly it did in the Oosterhoff *et al.* (1993) experiment. Oosterhoff *et al.* (1993) explained their results by suggesting that proprioceptive information could also be used at some higher stage in the structure-from-motion process. For example, proprioceptive information on the observer's head velocity could be combined with the optic flow to reach estimates of depth and distance.

The results of the above-mentioned studies indicate that it is important to analyse more specifically the effect of head movements in both a three-dimensional task (structure-from-motion) and in a two-dimensional task (motion perception). If retinal stabilization has an improving effect on the perception of three-dimensional structure-from-motion, it is likely that it also improves the perception of motion in itself. The perception of motion and the perception of structure-from-motion have common characteristics, as was shown by Treue *et al.* (1991). They found that the point lifetime threshold for perceiving structure-from-motion was similar to the threshold for estimating velocity, and concluded that velocity measurements are used in the structure-from-motion process. This again stresses the importance of examining the effects of head movements in both a two-dimensional and in a three-dimensional task.

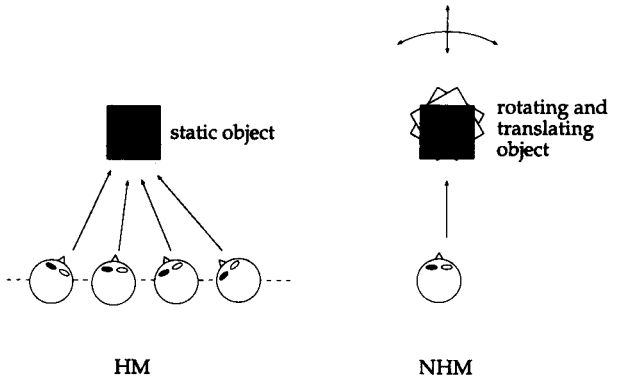


FIGURE 1. The two conditions HM (head movement) and NHM (no head movement). HM: the observer moves through the world and the objects are static or dynamic (but in this figure static). NHM: the observer is static and the object is moving in the world (in this figure the object rotates around a vertical axis).

In the present paper, the performance of human observers in a two-dimensional task (perception of motion) and in a three-dimensional task (perception of three-dimensional shape or three-dimensional curvature from motion) will be compared, in two conditions (Fig. 1): (1) the observer induces the motion parallax by moving the head; this condition will be called "with head movements" or HM; and (2) the observer is static and the parallax is determined entirely by the object-movement; this condition will be called "no head movement" or NHM.

In both conditions, the visual information (differential image velocity) is principally the same, but in the HM condition there is also non-visual information available. In both tasks, stimuli consisting of random dots with finite lifetimes are used.

## EXPERIMENT 1: THE THREE-DIMENSIONAL TASK

### *Design*

In the three-dimensional task, subjects were presented a series of pairs of random dot cinematograms, where each cinematogram simulated a curved quadratic surface. The surfaces were cylinder-like, randomly oriented around the line of sight and with a different curvedness* ($C_1$ and $C_2$, respectively). The order of display within a pair was random and, within one series, $C_1$ and $C_2$ were kept constant. The task of the subjects was to indicate after viewing of each pair, which one of the two simulated surfaces appeared the flattest.

This curvedness-discrimination paradigm was similar to that of van Damme *et al.* (1994). That experiment, performed with only static observers, showed that curvedness-discrimination performance obeys Weber's Law quite well. They found Weber fractions of about 0.35. This means, for example, that when observers have

*Curvedness is a nomenclature adapted from Koenderink (1990).

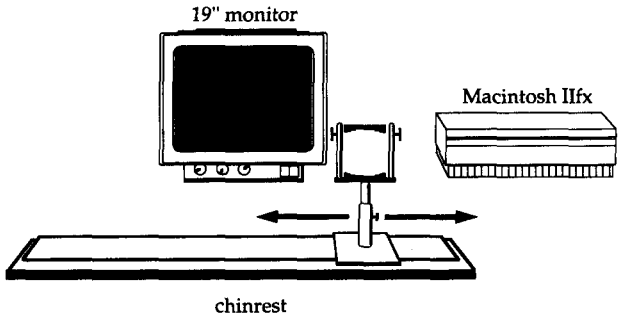


FIGURE 2. Experimental set-up. Subjects viewed the screen monocularly in a chin-rest, which could be moved along a rail, parallel to the screen. The position of the chin-rest, and thus the position of the eye, was measured continuously and fed back to the random dot display. In this way a three-dimensional surface could be simulated, in front of or behind the screen.

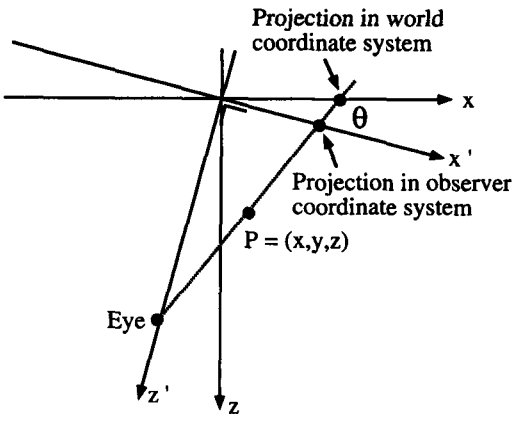


FIGURE 3. The perspective projection in the world coordinate system $(x, y, z)$ and in the egocentric coordinate system $(x', y', z')$. Both $y$ and $y'$ are pointing out of the paper. For an active observer, this plane of projection oscillates with respect to the frontoparallel plane. This is not the case for a static observer (see the Appendix).

to indicate whether or not they see the difference between pairs of simulated surfaces, one with a curvedness of 10/m, the other with a curvedness of 13.5/m, they would score 75% correct (this follows from the staircase procedure that they used).

In the present experiment, $C_1$ and $C_2$ were chosen to be 5/m and 8/m, respectively. The difference in curvedness is therefore 3/m which corresponds to $0.6\times$ the curvedness itself. According to van Damme *et al.* (1994), this difference in curvedness should be easily detectable and should result in more than 75% correct score. Pilot experiments confirmed that this was indeed the case.

*Apparatus*

The two conditions (HM and NHM) were compared with the same experimental set-up that was described in van Damme *et al.*, 1993, 1994).

With this set-up (Fig. 2), feedback of head-motion to the display was obtained by means of a movable chin/headrest. The chin-rest could be moved along a rail (length 0.37 m), parallel to a computer screen (Trinitron GDM 1950/1952). As the chin-rest was moved along the rail, a potentiometer below the chin-rest was turned. In this way, the voltage over the potentiometer varied as the chin-rest moved. The voltage was proportional to the position of the chin-rest and was sampled by an A/D interface card (National Instruments NB-MIO 16L) at a rate of 100 Ksamples/sec in a Macintosh IIfx computer.

The entire feedback loop of this set-up has some delay (the time between change in the chin-rest position and the update of the stimulus on the screen). It was not possible to measure this delay exactly, but instead an estimation can be given. The sampling of the chin-rest voltage and the drawing of the stimulus occurred within a single frame that was synchronized with the refresh rate of the monitor. Each A/D conversion took 0.01 msec (because of the 100 Ksamples/sec rate). Since the delay caused by the mechanical part of the chin-rest can be neglected the overall delay is well below 16 msec, i.e. one frame of the monitor.

In the HM condition, the subjects were free to move head- and chin-rest together in a way they felt comfortable. In that situation, chin-rest motion proved to be close to sinusoidal with a frequency of about 0.7 Hz. Subjects did not use the complete extent of the rail on which the chin-rest was supported but only up to a factor of 0.85 on average (this was determined in pilot experiments). In the NHM condition, the chin/head-rest was fixed in the centre of the rail, so that the line of sight was perpendicular to the computer screen on which the stimuli were shown. The movement of the stimulus was now induced by mimicking a sinusoidal movement of the chin-rest with a frequency of 0.7 Hz, and with an amplitude of $0.85\times$ the maximum amplitude of the chin-rest. In this way, the motion parameters of the visual stimuli were similar in both the HM and the NHM conditions.

In the HM condition, the plane of projection (which is the same as the CRT screen if the observer is located in the middle of the chin-rest) made a small rotation around a vertical axis whenever the observer moved. If just some chin-rest movement were mimicked in the NHM condition, then this rotation would not be accounted for. To make the visual input as similar as possible in both situations, a correction was used in the projection calculations for the NHM condition to compensate for this effect (see Fig. 3 and the Appendix).

Although this procedure leads to "similar" visual stimuli, it did not guarantee that the retinal image sequences were exactly the same in the two conditions. A small difference remained, since the head movement of active observers could never be exactly sinusoidal. The retinal image of active observers therefore would be less stable than the retinal image of static observers. If there were an effect of this small difference, then it would be that active observers have a worse performance.

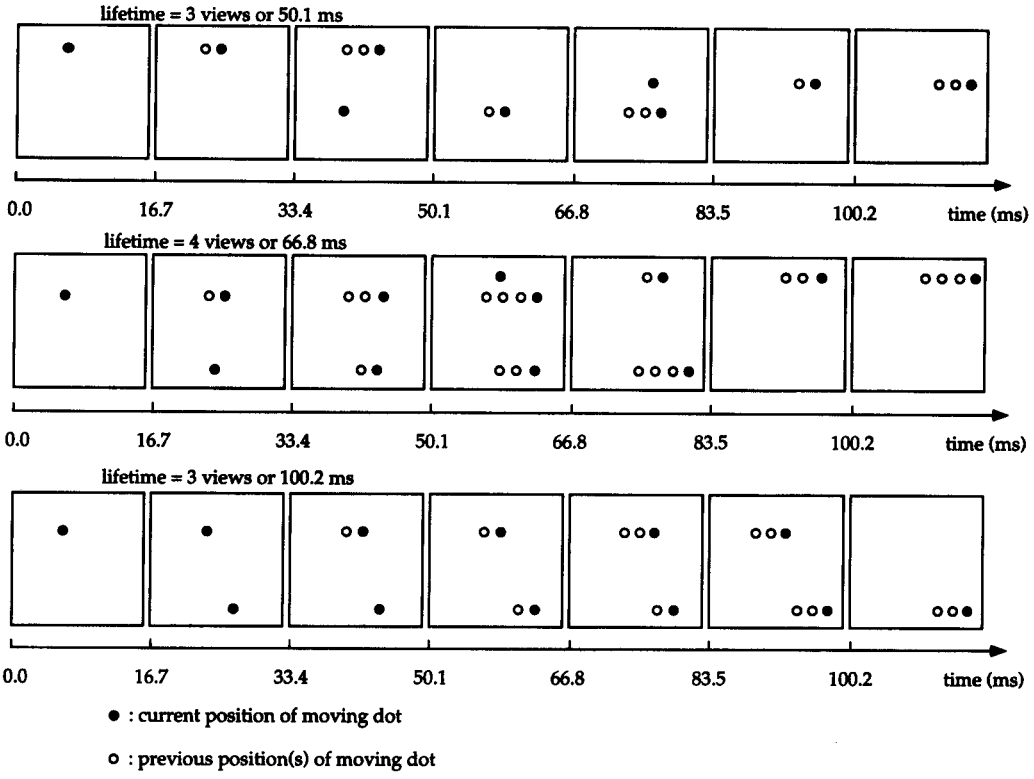


FIGURE 4. Three examples of the use of limited lifetimes and different frame durations in a cinematogram. Each panel shows eight individual frames from the, say, infinitely long sequence that the monitor displays. The position of a dot on the screen is indicated by a filled circle. Previous positions are drawn as open circles, and serve as a clarification of the path that the dot travelled (and of the age of this dot). Top: view duration = 16.7 msec and lifetime = three views or 50.1 msec; middle: view duration = 16.7 msec and lifetime = four views or 66.8 msec; bottom: view duration = 33.4 msec and lifetime = three views or 100.2 msec.

*Stimuli*

The stimuli were random dot cinematograms of quadratic surface patches with dots of limited lifetime. When a random dot disappeared, a new one was created at a random position and given an age of one. At the start of each presentation, all the dots were assigned a random age, so that not all dots reached the end of their lifetime simultaneously. This prevented a sudden refresh of all dots simultaneously during the presentation. Lifetime of the dots could be expressed in number of frames or in milliseconds. If one view was drawn within one video frame, then one view lasted 16.7 msec (the monitor that was used to display the stimuli operated on a 60.0 Hz vertical retrace rate). If one view was drawn within two video frames, then one view lasted $2 \times 16.7$ msec = 33.4 msec, etc. The view duration was taken as an additional parameter in this experiment.

There were 10 different lifetimes possible in the range of 2–11 views, but it depended on the view duration what values were actually presented to the subjects (and because of technical problems, some lifetime settings were skipped). Because the delay between stimulus movement and head movement in the HM condition increases when the duration of the stimulus views increases, only three values of view duration were used: one, two or three video frames, equivalent to 16.7, 33.4 and 50.1 msec. Any higher view duration would result in too slow a feedback in the HM condition (the movement of the dots would be far from continuous). Figure 4 illustrates the use of different lifetimes and view durations.

For both the HM and NHM condition, there were three view durations per subject (except for subject MV who did not measure the 50.1 msec condition). For each view duration, a separate session was conducted, and each session was repeated five times. Each session provided a percentage correct answers for each lifetime, and the percentages of the five repeated sessions were averaged.

The images were viewed from 1.00 m. They were presented through a square mask made of black card that was attached to the display. The size of the stimulus (mask) was $15 \times 15$ cm ($8.5 \times 8.5$ deg). Pixel size was 0.351 mm (0.02 deg). Each image contained 360 dots and was visible for 3 sec. All images were viewed monocularly in an otherwise completely dark room.

*Subjects*

Three male subjects participated in this experiment (WD, MV and FM). They all had normal or corrected to normal vision.

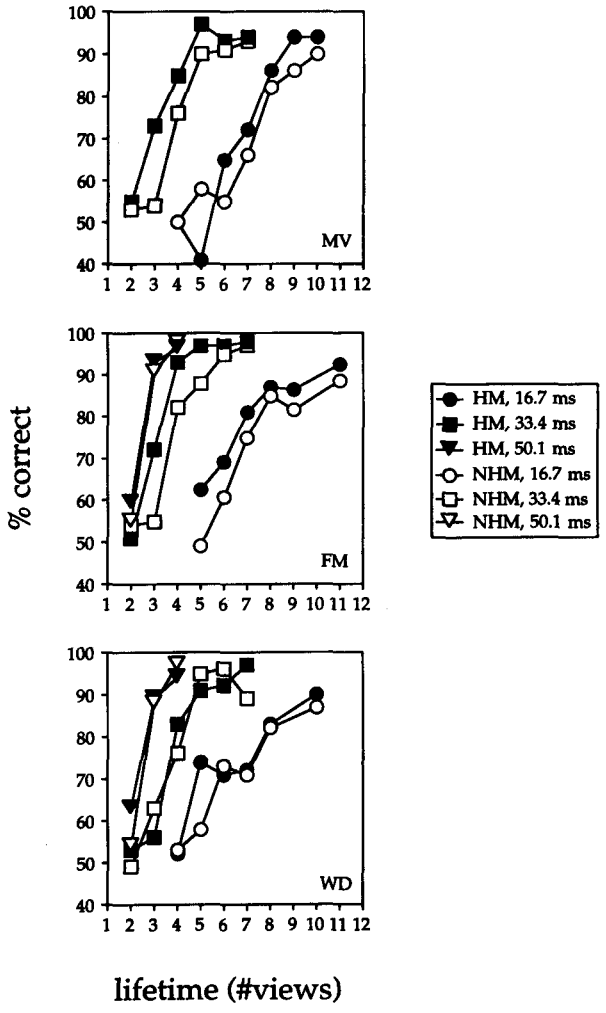


FIGURE 5. Percentage correct answers for all subjects (FM, MV and WD) for the view durations 16.7, 33.4 and 50.1 msec and for the two conditions HM and NHM. For subject MV, only the view durations 16.7 and 33.4 msec were measured.

*Results*

The results of Experiment 1 are summarized in Fig. 5. It is clear that for two subjects (FM and MV), the HM condition provides a higher percentage of correct responses than the NHM condition, for nearly all lifetimes. For subject WD, there is no difference in performance between the HM and the NHM condition. The figure also shows that the duration of a single view strongly affects the performance. For view durations of 16.7 msec, the 75% level (which can be taken as threshold level) is reached at a lifetime of about seven views. For view durations of 33.4 msec, however, this level is reached at a lifetime of about four views, and when view duration is set to 50.1 msec, the 75% level is reached at a lifetime of three views. These results clearly indicate that a build-up in time takes place in the structure-from-motion process. At low view durations, the individual views are not visible long enough to be of sufficient use for the SFM system. For view durations of 50.1 msec, the SFM process is more rapid: the slope of the psychometric curve is steepest, and all subjects can do the task with a score higher than 90% within four views. An analysis of variance (ANOVA) performed on the data averaged over subjects, showed a significant effect of HM/NHM ($P = 0.03$) in addition to the obvious significant effects of lifetime (in number of views) and view duration (for both effects $P < 0.001$).

Figure 6 shows the same data as in Fig. 5, but now the lifetime is expressed in milliseconds rather than in number of distinct views. In Fig. 6 the data of all the different view durations were collected in a single plot for each subject. An ANOVA, performed on the data averaged over subjects, again showed significant effects of lifetime in msec and HM/NHM (for both effects $P < 0.001$).

*Discussion*

When the results of Experiment 1 are compared to those of Treue *et al.* (1991), the same build-up in time of the SFM process is found. They found typical thresholds of 69–81 msec for detection of SFM, whereas in the present experiment, thresholds of 110–120 msec in a SFM discrimination task were found. The difference is possibly due to the difference in task (detection vs discrimination) and procedure between the present experiment and theirs. Treue *et al.* used a reaction time

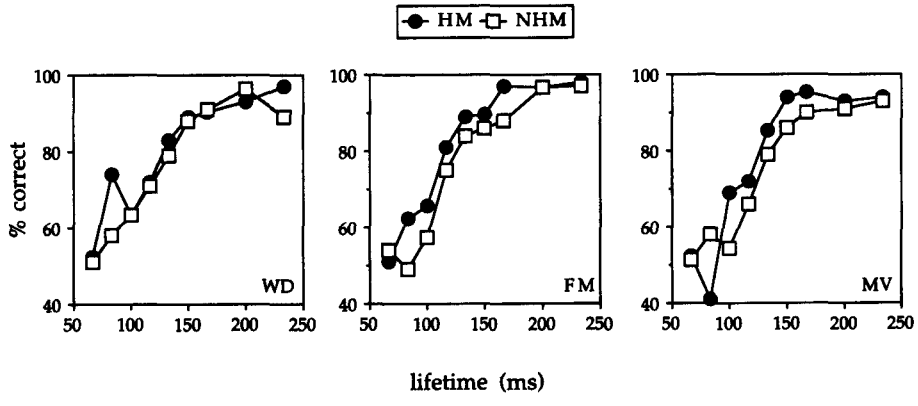


FIGURE 6. Percentage correct answers for the three subjects (FM, MV and WD) as a function of lifetime, now expressed in msec, and for the two conditions HM (●) and NHM (□).

paradigm for detecting structure in a limited lifetime display whereas, in the present experiment, the discrimination performance was measured.

Head movements decrease integration time with up to about 20 msec, which is considerable given the fact that the SFM process reaches the 100% level at 180–200 msec in the present experiment and at about 130 msec in the experiment of Treue *et al.* In almost none of the cases was the performance of moving observers worse than that of static observers. This is in agreement with what was found by Oosterhoff *et al.* (1993), but not with the conclusion of Cornilleau-Pérès *et al.* (1994) that best performance can be expected from a static observer viewing a rotating object. It is difficult, however, to compare two studies with a different experimental design. As far as can be told from the reports, the main differences between the two studies are the task, the stimulus size, the stimulus presentation time and the delay time in the feedback loop of active observers. Cornilleau-Pérès *et al.* (1994) used a detection task with a stimulus of 20 deg, a presentation time of 6 sec and a delay of 55 msec. These different conditions could provide the key to the different results but exactly how they are responsible for the differences is a subject for further research (see for example Dijkstra *et al.*, 1995).

The difference in performance between moving and static observers vanishes for large lifetimes (>180 msec). The results indicate that the visual system can benefit from proprioceptive information quite early in the SFM process (possibly when the visual information is not very effective yet), and the results of Oosterhoff *et al.* (1993) show that this advantage remains if the same task is performed with a display containing dots with infinite lifetime.

It is still possible, however, that subjects obtained some advantage of a better retinal stabilization in the HM condition that could explain the higher scores. A more stable retinal image could allow a more accurate measurement of velocities. Indirectly, this could lead to a qualitatively higher performance in processes that require velocity as input. Treue *et al.* (1991) already suggested that velocity measurements were used for the SFM process and their experiment provided evidence for common characteristics of velocity measurements and the SFM process. To check whether velocity measurements or more generally the perception of motion are influenced by head movements, a second experiment was designed, in which performances of both active and static observers in a two-dimensional motion detection task were measured.

*Actually, they created noise by modulating the luminance of each pixel between frames by luminance addition of a spatially and temporally uncorrelated random noise pattern. The difference between this method and the more conventional method is not relevant for the purposes of our experiment, so we refer to the original article for more details.

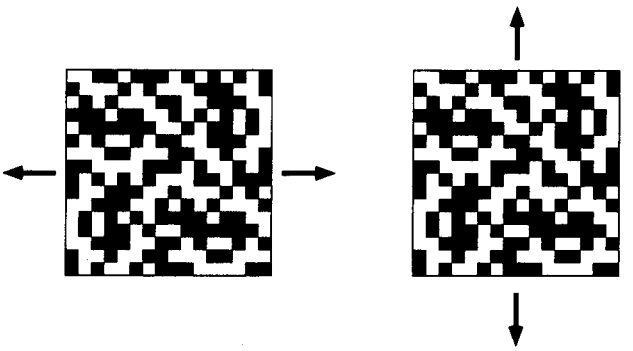


FIGURE 7. Basic illustration of the stimulus used in Experiment 2. Subjects indicated the perceived direction of the moving random pixel arrays. In the condition "horizontal motion", this could be left or right; in the condition "vertical motion", this could be up or down.

## EXPERIMENT 2: DETECTION OF MOTION

### *Design*

The purpose of the second experiment was to examine whether observers that make head movements show a different performance in a motion detection task than observers who do not move their head. If the performance of moving observers were to differ from that of static observers in a three-dimensional task because of the difference in retinal stabilization, then their performance will also differ from that of static observers in a two-dimensional task.

Therefore, the sensitivity for detecting motion was measured both while observers move their head and when the head was fixed. The design of this experiment was based on the work of Fredericksen *et al.* (1993, 1994). They modelled human motion perception by an array of bi-local detectors, each sensitive to a specific combination of spatial displacement (span) and time interval (delay) and thus "tuned" to a specific velocity (where the tuned velocity is the ratio of span and delay). They defined motion sensitivity as the reciprocal threshold for detecting motion. In their experiment, static observers were asked to indicate the direction of perceived motion in a noisy random pixel array (see Fig. 7).

In a forced-choice paradigm, the amount of noise (or more precisely, the signal-to-noise ratio or SNR*) was manipulated according to the responses of the observers, and in this way a threshold SNR was determined that represented the threshold for detecting motion. In their experiment, observers were always static, but in the present experiment, both static observers and observers moving the head were tested. For reasons of conformity, the two conditions will be called HM and NHM, although there was no feedback of head movement to stimulus movement in the HM condition of the second experiment.

A directional discrimination task was used to obtain motion detection thresholds as a function of retinal

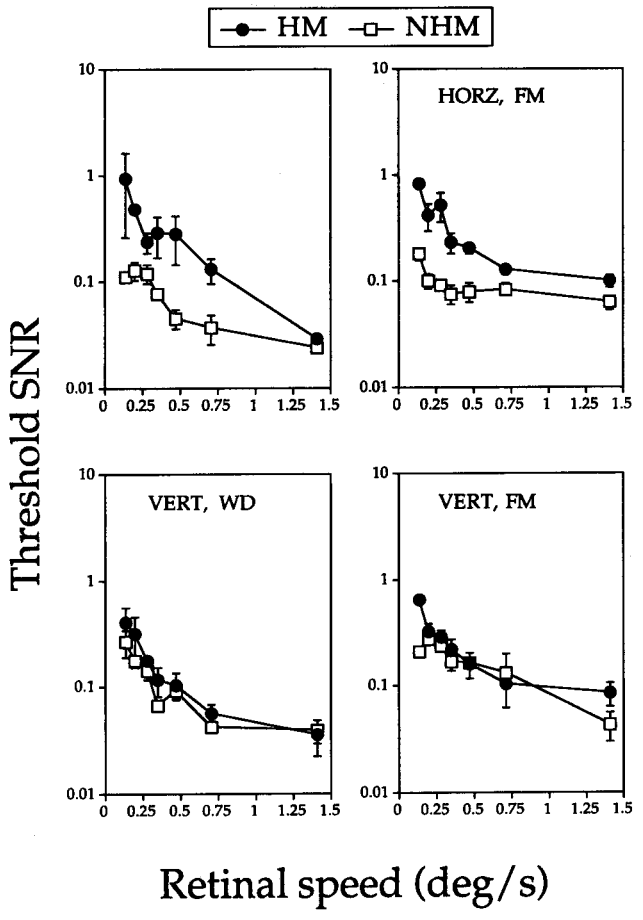


FIGURE 8. Threshold luminance signal-to-noise ratios for detecting horizontal (upper two panels) and vertical (lower two panels) motion for two subjects in the HM (●) and NHM (□) conditions as a function of retinal speed.

velocity of the moving random pixel arrays. There were two directional conditions: horizontal and vertical. In the horizontal motion case, there was randomly a rightward or a leftward moving pixel array. In the vertical motion case, there was randomly an upward or a downward moving pixel array. A staircase procedure was used in which subjects had to indicate the perceived direction of the moving pixel array. The staircase tracked the 79% correct level and ended when the tenth turning point was reached, after which the last six turning points in the staircase were averaged and stored as the threshold.

To compare the results of Experiments 1 and 2, it is important to have similar stimulus parameters. Therefore, the retinal velocities of the stimuli in the second experiment were chosen in such a way that they more or less covered the range of retinal velocities of the stimuli in the first experiment. The same holds for the head movements: a metronome was used in the second experiment for indicating the rhythm of the head movements so that it would be equal to the frequency of head movements in the first experiment.

*Subjects*

Two of the subjects that participated in Experiment 1 (WD and FM) also participated in this experiment.

*Stimuli*

Stimuli were moving random pixel arrays, generated by custom image generation hardware that was driven by a Macintosh computer. Stimuli were displayed on a monitor at a frequency of 90 Hz in an otherwise dark room. The stimulus size was $14 \times 14$ cm ($256 \times 256$ pixels, pixel size 0.55 mm). Viewing distance was 2.0 m. Each stimulus was visible for 1 sec. All stimuli were viewed with one eye, the same that was used in Experiment 1. A fixation dot was attached to the centre of the screen, and subjects were instructed to maintain fixation on this dot as accurately as possible.

The retinal velocities could be 0.070, 0.100, 0.141, 0.176, 0.235, 0.352 or 0.705 deg/sec. Retinal velocity can also be expressed as the ratio of span and delay. In this second experiment, the span was fixed at 1 pixel = 0.015 deg, so that the following values of delay were needed to obtain the above-mentioned retinal velocities: 222.0, 155.4, 111.0, 88.8, 66.6, 44.4 and 22.2 msec, respectively. These values correspond to a multiple of the duration of one frame of the monitor, which is 11.1 msec. The delay time value in this second experiment is comparable with the view duration of the first experiment.

*Results and discussion*

Figure 8 shows SNR thresholds for the two subjects in the NHM and HM case for both directional conditions. The detection thresholds are comparable with those found by Fredericksen *et al.* (1993). The range of retinal speeds in the present experiment extends further towards lower speeds, however. It is clear that the thresholds for detecting horizontal motion are higher for moving observers than for static observers. The sensitivity for detecting vertical motion is not affected by horizontal head movements. There is certainly no improvement in motion sensitivity for moving observers. Apparently the head movements only have an effect when the direction of head movement is the same as the direction of the visual motion (either horizontal or vertical).

The fact that active observers are better at maintaining fixation (minimizing the retinal slip) than static observers led Cornilleau-Pérès *et al.* (1994) to the conclusion that active observers are better in SFM tasks because of this advantage only and not because information about ego-motion was used in cooperation with visual motion. Should this be the case, then it is likely that sensitivity for detecting motion is higher for active observers than for static observers: active observers simply stabilize the retinal image better than static observers. The results of Experiment 2, however, show just the opposite: active observers have a lower motion-detection sensitivity than static observers.

## GENERAL DISCUSSION

The results of Experiment 1, together with the results of Oosterhoff *et al.* (1993), show that active observers are more accurate in processing structure-from-motion than static observers. Apparently, one gains some advantage when moving the head. However, this advantage is not present at the level where motion is detected. In fact there is a disadvantage for moving observers in detecting motion. This does not fit with the suggestions that, in a three-dimensional task, moving observers take advantage of a better retinal stabilization (e.g. by using the vestibulo-ocular reflex) and, as a consequence, improve the processing of structure-from-motion.

Experiment 2 shows that moving observers have a clear lower motion detection sensitivity. It may seem remarkable that the processing of structure-from-motion is improved by ego movements whereas motion detection is hampered by ego movements. A possible explanation might be that the use of a fixation mark in Experiment 2 provided subjects the opportunity for stabilizing the image in both conditions better than they would when no fixation mark was present. In that case, the movements of the observer in the HM condition would cause small instabilities in the retinal image with a decrease in performance as a consequence. However, this cannot explain the results of Experiment 1, in which no fixation mark was used.

Since it is likely that motion is processed prior to structure-from-motion, a degraded perception of motion for a moving observer should lead to a degraded perception of SFM unless qualitative proprioceptive information is available at a higher level of SFM processing. The results of this experiment show that such a cooperation of visual and non-visual information in the processing of SFM is a serious possibility, and that non-visual signals are not mainly used for retinal stabilization.

## REFERENCES


Buizza, A., Léger, A., Droulez, J., Bertoz, A. & Schmid, R. (1980). Influence of otolithic stimulation by horizontal linear head acceleration in optokinetic nystagmus and visual motion perception. *Experimental Brain Research, 71*, 406–410.

Cornilleau-Pérès, V. & Droulez, J. (1994). The visual perception of three-dimensional shape from self-motion and object motion. *Vision Research, 34*, 2331–2336.

van Damme, W. J. M. & van de Grind, W. A. (1993). Active vision and the identification of 3D shape. *Vision Research, 11*, 1581–1587.

van Damme, W. J. M., Oosterhoff, F. H. & van de Grind, W. A. (1994). Discrimination of 3-D shape and 3-D curvature from motion in active vision. *Perception and Psychophysics, 55*, 340–349.

Dijkstra, T. M. H., Cornilleau-Pérès, V., Gielen, C. C. A. M. & Droulez, J. (1995). Perception of 3D shape from ego- and object motion: Comparison between small and large field stimuli. *Vision Research, 35*, 453–462.

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

Fredericksen, R. E., Verstraten, F. A. J. & van de Grind, W. A. (1993). Spatio-temporal characteristics of human motion perception. *Vision Research, 33*, 1193–1205.

Fredericksen, R. E., Verstraten, F. A. J. & van de Grind, W. A. (1994). Temporal integration of random dot apparent motion information in human central vision. *Vision Research, 34*, 461–476.

Green, P. R., Davies, N. O. & Thorpe, P. H. (1994). Head-bobbing and head orientation during landing flights of pigeons. *Journal of Comparative Physiology A, 174*, 249–256.

Hayashibe, K. (1991). Reversals of visual depth caused by motion parallax. *Perception, 20*, 17–28.

Koenderink, J. J. (1990). *Solid shape.* Cambridge, MA: MIT Press.

Ono, M. E., Rivest, J. & Ono, H. (1986). Depth perception as a function of motion parallax and absolute distance information. *Journal of Experimental Psychology, 3*, 331–337.

Oosterhoff, F. H., van Damme, W. J. M. & van de Grind, W. A. (1993). Active exploration of three-dimensional objects is more reliable than passive observation. *Perception, 22*, 99.

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 and Psychophysics, 52*, 446–452.

Sobel, E. C. (1990). The locust's use of motion parallax to measure distance. *Journal of Comparative Physiology A, 167*, 579–588.

Stappers, P. J. (1992). Scaling the visual consequences of active head movements: A study of active perceivers and spatial technology. Ph.D. Thesis. The Netherlands: University Delft.

Treue, S., Husain, M. & Andersen, R. A. (1991). Human perception of structure from motion. *Vision Research, 1*, 59–75.


*Acknowledgements*—This research is sponsored by the SPIN project "3D-Computer Vision" of the Dutch Ministry of Economic Affairs. We are very grateful to Frank Mulder and Frans Verstraten for their help during the experiments.

## APPENDIX

The method of comparing active and non-active vision that is described in this paper assumes that both conditions provide equal retinal input. This can be achieved by "recording" the movement of an active observer and use of this record to generate a stimulus for a non-active observer. However, it is not entirely correct to feed this recorded movement into the same projection algorithm that is used for an active observer: for an active observer, the plane of projection is oscillating, whereas for a static observer the plane is not (it is static). This Appendix illustrates that considerable errors in projected position can be made if this difference is not taken into account. These errors can be found by calculating the projection of a point in space in terms of egocentric coordinates instead of world coordinates.

Call $\mathbf{r} = (x, y, z)$ the world coordinates that may be visualized as being related to the CRT screen, and call $\mathbf{r}' = (x', y', z')$ the egocentric coordinates, with $z'$ pointing towards the origin of the world (the centre of the screen). Point $P$ is a generic point in space that we want to project on the plane $z = 0$. For a static observer, the plane of projection is always fronto-parallel. To obtain the projection in egocentric coordinates, we calculate the transformation from world coordinates to egocentric coordinates, which is straightforward mathematics: $\mathbf{r}' = T\mathbf{r}$, with $T$ the transformation matrix of world coordinates to egocentric coordinates.

In the chin-rest set-up, the vertical head position is always zero, and then $T$ looks like:

$$T = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$$

where $\theta$ is the angle of rotation around the $y$-axis (vertical axis).

The projection in egocentric coordinates depends on the viewing distance, the simulated depth and the position of the head. As an example, we calculated the error in projected position that is made if

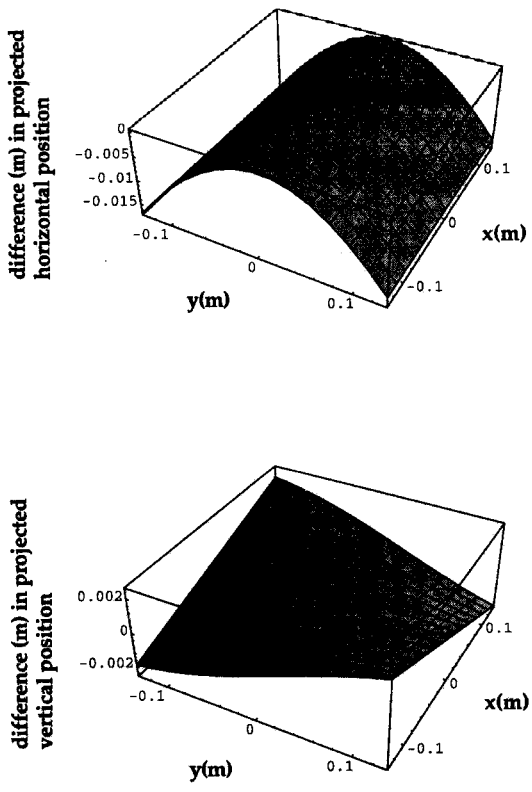


FIGURE A1. Errors in projected position that occur if the correction as described in the Appendix would not be made. For this numerical example, typical values of the experimental parameters were used. (A) Errors in horizontal projected position; (B) errors in vertical projected position.

the correction is not made. For this calculation, we took values that are typical for our experiments: a normal head movement with an amplitude of 0.157 m, viewing distance is 1.00 m and a quadratic surface with curvedness C = 8/m. The error in projected (horizontal and vertical) position is illustrated in Fig. 9.

There is no error in projected position in the centre of the image, which corresponds to the assumed point of fixation. Anywhere else, the errors in projected position are non-zero and can be as large as 1.5 cm horizontally and 0.2 cm vertically, which cannot be neglected. Hence, the correction was used for these differences in the generation of the stimuli in Experiment 1.