Remote Control of Vehicles

Jacqueline Viveash, Joanna White, Jenny Boughton, Stuart King and Martin Kaye
QinetiQ ltd
Cody Technology Park
Building A50
Ively Road, Farnborough
Hampshire, GU14 0LX
United Kingdom

The remote guidance of vehicles or tools in inaccessible or hazardous environments is of considerable military interest. The control performance achieved with these systems is dependent on the information received by a human operator. Sensors such as Night Vision Goggles (NVG) mounted on the head, or cameras mounted on the vehicle or tool can be used to provide such information, and how many sensors and where the sensors are positioned will depend on the task to be performed. Two experiments have been conducted to assess the effects on performance of presenting monocular, stereoscopic, or enhanced stereoscopic (hyperstereoscopic) information to the operator. One experiment used a head-mounted system to investigate the effects of these viewing systems on depth perception under static and dynamic conditions. In the other experiment cameras were mounted on a remotely controlled vehicle to investigate the effects the different viewing systems on vehicle control. Two tasks were used in this second investigation; one was a driving task, the other a manipulation task. The first experiment showed that there was an effect of motion on the results; depth was estimated more accurately under static than under dynamic conditions. Neither stereopsis nor hyperstereopsis had a measurable effect on depth perception. It was concluded that further experimentation should take place using a more appropriate task. The second experiment indicated that remote control performance was task dependent. For the driving task there was no significant difference between the performance measured under monocular and stereoscopic conditions. In the manipulation task the best performance was achieved using stereoscopic presentation techniques, the hyperstereoscopic presentation of information producing a 38% reduction in task completion time over the monocular time. For both tasks, regardless of whether the differences were significant or not, the better performance was always achieved using stereo rather than monocular presentation techniques. Thus it was concluded that there are advantages in using stereo rather than monocular presentation techniques for remote control tasks. In manipulative tasks the performance gain can be as high as 38%. Such an improvement in performance is clearly of importance for tasks such as bomb disposal and in-flight refuelling. In both experiments some subjects complained of eyestrain when using the hyperstereoscopic systems. Further work should be conducted to determine the optimum convergence setting for different tasks, and the amount of disparity easily tolerated by the majority of the population. On this basis of the above results a new apparatus has been designed to evaluate performance when head mounted displays such as night vision goggles (NVG) are used. The head-mounted apparatus consists of pairs of mirrors to reflect the visual scene into each eye of the individual. The outer mirrors will be positioned to produce effective interpupillary distances (IPD) of 2x, 3x or 4x the individuals IPD. The monocular and 1xIPD configurations will also be investigated. Further experiments are planned which will investigate the relationship between eyestrain and hypersteropsis.

Introduction

There is considerable military and civilian interest in the remote guidance of robotic tools and vehicles. This interest arises because remote control systems are required for use in hazardous and inaccessible environments. The efficiency of control and the performance achieved with these vehicles is highly dependent on the information received by the human operator, and performance advantages can be gained by the correct presentation of this information. Sensors mounted on the head, vehicle or tool can be used to provide such information, and how many sensors and where the sensors are positioned will depend on the task to be performed.

Paper presented at the RTO HFM Symposium on "Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures", held in La Coruña, Spain, 15-17 April 2002, and published in RTO-MP-086.

One aspect that has to be considered when determining the best camera layout is whether or not there will be any improvement in control performance if stereo rather than monocular information is provided to the operator. This question will be of particular importance when the operator's task requires both distance and depth perception (Viveash, 2000).

Initially let us consider binocular and monocular vision. Although two-eyed vision does provide the primary cues to depth perception there are many people with one-eyed vision who have very good depth perception. This is because there are at least seven monocular cues (overlapping contours, or obscuration: motion and linear perspective, texture, light and dark shading, accommodation of the eye and aerial perspective) which are also used in distance and depth perception (Boff and Lincoln, 1988).

Stereopsis is the function of the binocular vision system, which amounts to a detailed comparison of the two retinal images on the basis of parallax geometry, the two retinal images are fused by the brain and yield a vivid and highly detailed perception of three-dimensional space. Typically the stereoscopic threshold varies from 1.6 to 24 seconds of arc. Targets of larger disparities may be seen as double images with no accompanying sensation of depth, or only one image may be seen and the other suppressed (Boff and Lincoln, 1988).

Using two cameras to provide a stereo view on a monitor may result in double images and distortion. These problems arise because the two cameras point to an object at a fixed point (i.e. the cameras are converged to a set distance), and usually unlike the eyes, the convergence of the cameras does not alter when other objects at different distances are observed. As a result whilst the objects in the plane of convergence will appear as a single image, images of objects out of the plane of convergence can appear as a double image, and the scene appears distorted. Furthermore, unlike the visual system, there is no integral mechanism, which automatically suppresses one of the images. Another aspect that has to be considered is that varying the distance between the cameras will also affect disparity and stereoscopic thresholds. Placing the cameras further apart will result in an enhanced disparity. Thus, another aspect that needs to be considered is whether enhanced disparity produces a better performance in all tasks.

Experiment 1

The aim of this experiment was to evaluate what effect the monocular, stereoscopic or enhanced stereoscopic presentation of information had on a subject's depth perception under both dynamic and static conditions. To this end measures of the subject ability to estimate the distance from a Stop sign were made under the two conditions; (a) when the subject was moving and (b) the subject was static and an object (a walking experimenter) was moving.

Apparatus and Method

Twelve subjects took part in the experiment. All subjects were tested for normal vision. That is they were tested for 6/6 acuity without spectacles, or corrected to 6/6 with contact lenses (but not with spectacles). In addition they were tested for normal stereoscopic vision using the Randot and Titmus fly stereoscopic tests (for a description of the tests see Boff and Lincoln, 1988).

The apparatus used to produce the three viewing presentations and the Stop sign target were common to both conditions. Enhanced stereopsis (hyperstereopsis), stereopsis or monocular viewing was produced using a head mounted Variable Latency Asynchronous Display (VLAD). Essentially, for the purposes of this experiment, the VLAD apparatus consisted of two cameras whose individual images were sent to the subject's eyes via reflecting optics. In order to produce hyperstereopsis the distance between the two cameras (i.e. the effective interpupillary distance or IPD) was variable. The Stop sign was a standard road sign with white letters on a red background The word STOP was in capital letters and 560mm wide, each letter was 150.5 mm high, and had a line thickness of 33 mm.

Under both conditions the cameras were positioned at one of four interpupillary distances (IPD). The positions were IPD x 0, IPD x 1, IPD x 2, and IPD x 4. In the IPD x 0 condition the same image was sent to both eyes. When the same image is presented to each eye, i.e. with no disparity, the configuration is known as a biocular presentation. However, in order to draw parallels between the camera configurations and one and two-eyed performance, it will be referred to as the 'monocular' condition throughout this paper. The different experimental procedures used under dynamic and static conditions were as follows.

Dynamic condition: This part of the experiment took place on a 50 metre linear track. The subject was seated on the track trolley that was moved in a straight path using the trolley winch. The trolley travelled at a velocity of 0.22 ±. 0.03 m/s. The vehicle was stopped, by the subject operating a push button at the specified distance (an estimated distance of 3 m) from the Stop sign.

Static condition: In the static condition the subject viewed two markers, one fixed (the Stop sign) and one moveable (an experimenter). The subject's task was to verbally direct the walking experimenter to halt at an estimated distance of 3 m from the Stop sign.

In each trial all the subjects made eight runs at each IPD. Under both conditions the dependent measure was distance between the Stop sign and the trolley or marker. These distances were measured using a Laser Digital Distance Meter (Bosch DLE 30) with an accuracy better than 1%.

Results

A three-way ANOVA was performed on the factors Motion, IPD and Run. Differences between Motion, IPD and Run were analysed using Tukey HSD tests.

IPD	Static	Dynamic
0	5.84	7.76
1	5.69	7.77
2	5.42	8.28
4	5.10	8.00

Table 1: Mean Distance Estimation Scores (in metres)

Significant main effects of Motion ($F(1, 10) = 13.84$, $p<0.01$) and Run ($F(7, 70) = 16.43$, $p<0.01$) were found. All motion judgements were greatly over-estimated with the static condition providing greater accuracy. There were no significant differences between the different IPD's. Mean distance estimation scores are shown in Table 5-1.

Discussion

An observer's ability to judge distance varies with the conditions under which the judgement is made, and one of the most important factor's in this judgement is the amount of information about the conditions available to the observer. The more information available, the more likely the subject is to perceive the distance accurately. Further when the information is inadequate the less accurate the perception and the greater variability in the distance estimate (Sedgwick, 1986). This would, to a certain extent, explain why the distance estimations were both inaccurate and highly variable in both trials. However, although all distance judgements were over-estimated the results show that static judgements produced significantly more accurate estimations than dynamic judgements (those made whilst moving). It is not surprising that a significant effect of motion was reported. In the static condition subjects had a greater number of perspective cues available in their field of view than with the dynamic condition, when they were physically positioned closer to the Stop sign. It is thought that the environment in which the trial was run was rich in monocular cues. This is likely to have largely contributed to the lack of effect between the different IPD's. The use of stereo cues would have been limited at best or not utilised at all.

Conclusion

This experiment demonstrated that there was a significant effect of motion on the perception of depth; depth was estimated more accurately under static than under dynamic conditions. As a result it was concluded that any further experimentation should take place using more appropriate tasks.

Experiment 2

The aim of this experiment was to investigate the effects of the different viewing systems on vehicle control. Two tasks were used in this experiment. One was a driving task, the other a manipulative task.

Apparatus and Methods

Twelve subjects took part in the experiment. The subjects' vision was tested as described in the previous experiment.

A robotic vehicle and stereoscopic viewing systems were common to both tasks. The robotic vehicle used in the experiment was controlled in both velocity and direction using a single joystick. The viewing system was provided by mounting either one or two cameras on a wood and aluminium plate, which was fixed to the chassis of the vehicle.

The inter-camera distance (ICD) was set by a series of parallel fixing holes in the platform and three ICD's were used: 0, 22 and 66 mm. The 0 mm condition was the 'monocular' condition in which a single camera was centrally mounted on the platform. In the 'stereopsis' condition, the 22 mm ICD was the minimum possible separation obtained with the two cameras mounted side by side, and it gave an apparently natural stereoscopic view within this model environment. The 66 mm ICD gave an impression of an 'enhanced stereopsis' condition. A detailed description of the apparatus is provided elsewhere (Viveash et al 2002).

Driving task. The driving task consisted of guiding the vehicle around a circuit through seven pairs of upright wooden markers. These were placed 480 mm apart, a distance 100 mm wider than the car chassis. The car needed to be square on to the markers to go cleanly through the gap. To reduce monocular cues to depth, the markers were varied in size (diameter and height) and their bases, where they stood on the floor, were obscured from view. In the driving task, performance assessment was the total time taken for the run. Each subject made eight runs at each ICD.

Manipulation task. The manipulation task required the subjects to capture five metal rings on a probe attached to the front of the vehicle. The rings were of five different diameters and were hung in a line from a supporting bar. The vehicle returned to a starting position each time a ring was captured. The probe was asymmetrically mounted on the vehicle and rose at an oblique angle, so that it was impossible simply to establish visual alignment between the probe and the ring, and then advance the car. Instead, subjects found it necessary to exert continuous control, rather than simply executing a pre-programmed movement. In the manipulation task, assessment was based on the total time taken to complete the task.

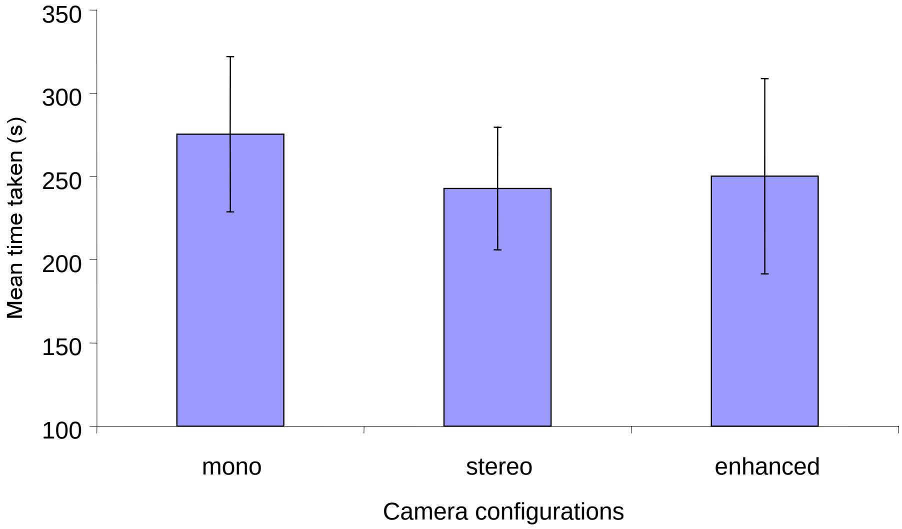
After completing both trials the subjects were presented with a questionnaire that asked them to rank order their preferences for the three ICD's used in the experiment, rank order 1 being the preferred choice.

Results

The main aim of the experiment was to assess the effects of three different camera configurations on vehicle control performance. The mean data for the driving task are shown in Figure 1 and those for the manipulation task in Figure 2.

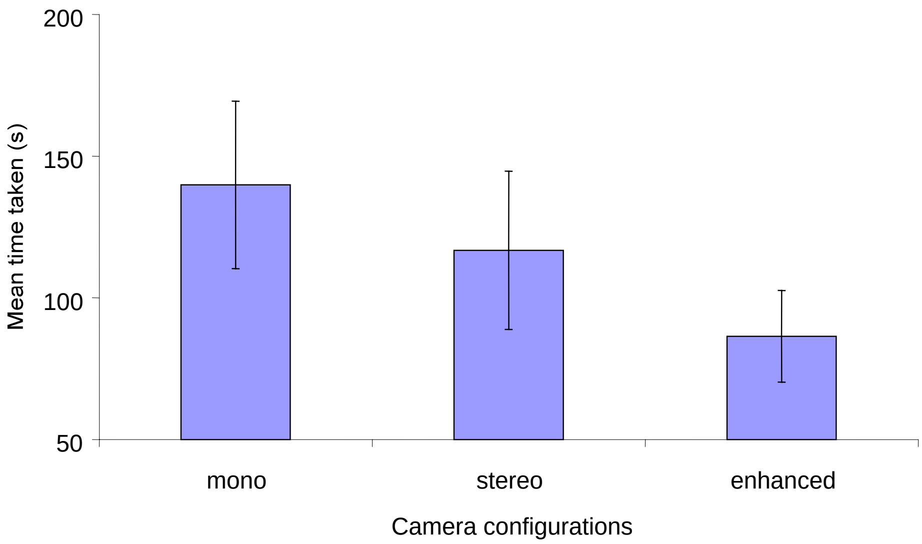
Driving Task. There was no significant main effect due to the different visual conditions. Mean times for the driving task at the three ICD separations are shown in Figure 1. The number of markers knocked down under the three display conditions varied very little (19±2).

Manipulation task. There was a significant main effect of camera configuration on the time taken ($F(2, 14) = 10.06$, $p<0.01$), and post-hoc comparison showed that monocular and stereo presentations resulted in significantly longer task duration's than the enhanced disparity presentation. Mean times for the three ICD configurations on the manipulation task are shown in Figure 2. It was also noted that the subjects had the greatest number of unsuccessful attempts to remove the rings under monocular presentation conditions and the least number under the enhanced disparity condition.



Mean time to complete in the driving task experiment. Figure 1

± standard error bars shown

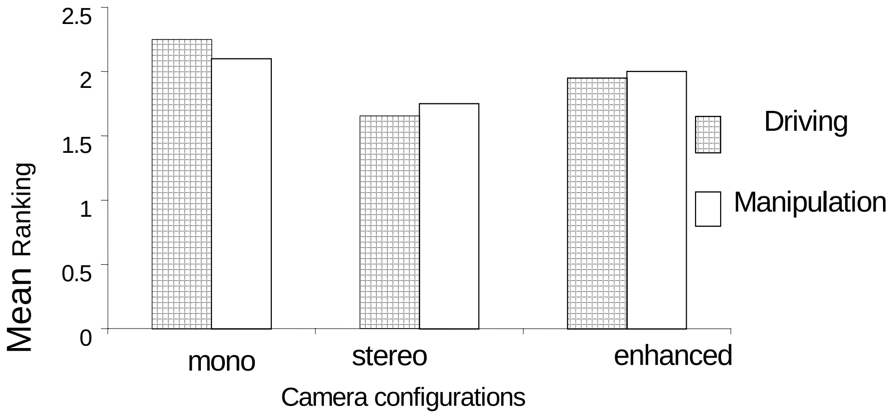


Mean time to complete in the manipulation task experiment. Figure 2

± standard error bars shown

Subjective

Rankings of the preferences for the methods of visual presentation for the two tasks are shown in Figure 3, in which lower scores indicate stronger preference. Although better performance was obtained in the enhanced condition than in the stereo condition, several subjects commented that it gave them feelings of eyestrain.



Subjective rankings Figure 3.

Discussion

The results confirm that the best method of task presentation (monocular or stereo) for a given task is task-dependent. In the driving task experiment there was no significant difference between the time taken to complete a circuit for any form of presentation. In the manipulation task experiment, however, stereoscopic presentation gave significantly better performance than monocular.

The negative outcome for the driving task experiment was not unexpected, because others have had a similar result (Gold et al,1968). Such a result is thought to arise when tasks are performed in environments that are so rich in monocular cues to depth that stereopsis provides little additional information. This was disappointing in this particular trial because particular care had been taken minimise the monocular cues.

The results from the manipulation task clearly demonstrated an improvement in control performance arising from stereoscopic presentation, which reduced mean task time by approximately 16%. Moreover, there was a further improvement in performance from enhanced stereopsis presentation, which decreased task time by 38%, in comparison with the monocular condition. Other experimenters have also shown improved performance with stereopsis in manipulation tasks (Smith et al, 1979).

In the subjective assessments the majority of subjects preferred stereoscopic over monocular presentation. The enhanced stereoscopic presentation, although more popular than the monocular presentation, was found to cause feelings of eyestrain, which would make it difficult to use for long periods of time. There were no comments about the distortion of the scene, even though such comments have previously been reported for cameras converged to a point near to a vehicle (Nagata, 1996) and not to infinity as in this trial. Nevertheless, as expected, double images were seen in the foreground when viewing the enhanced presentation.

The fact that some subjects complained of eyestrain when using the enhanced stereo presentation indicates that a further investigation of enhanced stereo techniques and eyestrain is required. This investigation should determine the optimum camera convergence angles for a variety of tasks and the degree of disparity easily tolerated by the majority of the population.

Another point to note is that performance was never worse for stereo rather than monocular presentation and there is, therefore, no indication that there is ever a disadvantage to using stereo presentation techniques for remote control tasks. For manipulative tasks performance gain may be as high as 38%, and such an advantage is clearly of importance for tasks such as bomb disposal and in-flight refuelling.

Conclusions

This second experiment has demonstrated that: (1) The relative advantage of stereoscopic over monocular presentation is task-dependent; (2) There was no significant difference in control performance between stereo and monocular display presentations in the driving task; (3) There was a significant difference in control performance between stereo and monocular display presentations for the manipulation task; (4) In both trials, a stereoscopic display always resulted in performance that was at least as good as with a monocular display.

References

Viveash., J.P. (2000). "*Hyperstereopsis*" DERA (Internal publication)

Boff, K.R. and Lincoln, J.E. 1988a, Depth Perception. In *Engineering Data Compendium. Volume II: Human Perception and Performance.* Harry G Armstrong Aerospace Medical Research Laboratory, Wright Patterson Air Force Base, Ohio. pp 1053-1127

Viveash, J.P, Kaye, M.G. White, J., Boughton, J., King, S.K. (2002) The remote control of vehicles and tools. *Contempory Ergonomics* (In Press)

Sedgwich, H.A. (1986). The perception of distance. In "*Handbook of Perception and Human performance Volume I* Ed Boff , Kr. Kaufman, L., Thomas, J.P. 21, 21.9

Gold, T. 1972, The limits of stereopsis for depth perception in dynamic visual situations. In *International Symposium & Seminar (72E06), Volume II* (Society for Information Display, New York)

Nagata, S. 1996, The binocular fusion of human vision on stereoscopic displays. Field of view and environmental effects. *Ergonomics*, **39** (11), 1273-1284

Smith, D.C., Cole, R.E., Merrit. J.O. and Pepper, R.L. (1979), Remote operator performance comparing mono and stereo TV displays. The effect of visibility, learning and task factors. *Naval Ocean Systems Center, San Diego, California, TR380*

Part of this work was carried out on behalf of Eurofighter IPT

Part of this work was carried out as part of the Technology Group 5 (Human Sciences and Synthetic Environments) of the MOD Corporate Research Programme

This page has been deliberately left blank

Page intentionnellement blanche