User Performance of Gaze-Based Interaction with On-line Virtual Communities

Howell Istance[1]   **Aulikki Hyrskykari**[2]   **Stephen Vickers**[1]   **Nazmie Ali**[2]

[1]Human-Computer Interaction Research Group,
School of Computing
De Montfort University, The Gateway,
Leicester, LE1 9BH, UNITED KINGDOM

{hoi,svickers}@dmu.ac.uk

[2]TAUCHI,
Department of Computer Sciences
FIN-33014 University of Tampere
FINLAND

ah@cs.uta.fi, nazmieali@gmail.com

Keywords
Gaze control, games, evaluation, virtual communities

Introduction

We present the results of an investigation into gaze-based interaction techniques with on-line virtual communities. The purpose of this study was to gain a better understanding of user performance with a gaze interaction technique developed for interacting with 3D graphical on-line communities and games. The study involved 12 participants each of whom carried out 2 equivalent sets of 3 tasks in a world created in Second Life. One set was carried out using a keystroke and mouse emulator driven by gaze, and the other set was carried out with the normal keyboard and mouse.. The study demonstrates that subjects were easily able to perform a set of tasks with eye gaze with only a minimal amount of training. It has also identified the causes of user errors and the amount of performance improvement that could be expected if the causes of these errors can be designed out.

Gaze driven mouse and keyboard emulation

The idea and the implementation of "Snap Clutch", the gaze interaction technique used, is described more thoroughly elsewhere (Istance et al., 2008). In short, the principle is to use a gaze dwell to emulate a variety of input events. Four different emulator modes are available at any one time to the user, and a simple off-screen glance is used to change to one of these modes. A 'mode' in this context means a way of mapping between gaze data as inputs, and mouse or keyboard events as outputs. So, for example, a 'pie menu' mode can be constructed where the first gaze dwell generates a right button click at that position on screen. The second dwell generates a left mouse button click, then the mode is automatically de-activated, and so subsequent dwells generate no events.

The four modes can be chosen according to the actual application being controlled. In this study we used the following four modes:

- Glance up mode : Unconstrained looking around
    - no action on dwell (control off),
    - rotate left when looking inside the left hand edge of the screen
    - rotate right when looking inside the right hand edge of the screen

- Glance right mode: Right button click
    - a gaze dwell causes a right button click

- Glance left mode: Left button click
    - a gaze dwell causes a left button click

- Glance down mode: Locomotion
  - no action on dwell
  - constant streaming of 'W' keystroke events when the user looks in the main part of the screen
  - streaming of 'A' and 'D' keystroke events when the user looks into small square regions in the left and right hand sides of the screen
  - streaming of 'S' keystroke events when the user looks inside a thin strip inside the bottom edge of the screen causing the avatar to walk backwards

An approach to user performance investigations

The initial pilot study (Istance et al, 2008) showed that using Second Life with our gaze-based technique resulted in task completion times that were distinctively longer than when using conventional interaction techniques. In order to achieve parity of gaze interaction with normal keyboard and mouse, it is important to be able to identify usability issues with gaze control in terms of what influences the speed of interaction (time of task completion) and the types of errors made.

Partitioning task time into 'productive' time and 'error' time has long been a feature of usability engineering (Gilb, 1984). The time spent in a specific error condition represents the potential saving in task completion time if the cause of that specific error can be designed out so the user no longer makes that error. The relative savings in task completion times by addressing each of the types of errors identified represents a kind of cost-benefit analysis of redesigning different features of the user interface.

The Experiment

Subjects and apparatus. The study involved 12 participants. Ten of them were students and two were university lecturers who were experienced users of gaze interaction. Ages varied from 20 to 56, the average being 29. All subjects were able-bodied. The trials were carried out using a Tobii T60 screen integrated eye tracker, and the window contents during all of the trials were recorded using screen capture software.

Tasks. Two sets of three tasks were devised to be carried out within a purpose built world within Second Life. The world represented the computer science building at the university.

- The *locomotion task* required the subject to walk from the main entrance, up the main stairs (Figure 1), go into a room where there were display panels about individual modules and then report the module code from a particular panel. The difference between the two sets of tasks was the actual panel the participant was asked to report the code from.

- The *object manipulation* task required the subject to change a slide or request a web page from the main lecture theatre. In one task set, the participant changed the slide. This involved a right click on a panel button object to display a pie menu and then a left click to select the 'Touch' option. In the other set, the equivalent task was to request a web page to be displayed. This involved a left click on a panel

Figure 1. A locomotion task – searching for the target from the upstairs.

button object near the stage, and another left click to cancel to the request.

- The *application control* task required the user to change the appearance of their avatar. In one set, the task was to remove the moustache and in the other it was to raise the height of the eyebrows. The participant right-clicked the avatar to display a pie menu with an option to edit 'appearance'. This caused a dialogue box to appear. The user had to select a group of features to edit from the vertical panel of buttons, then scroll down a list of features to make the required feature visible. A horizontal slider was used to change the selected feature.

Procedure. The subjects were split into two groups. One half did one task set with the keyboard and mouse, followed by the other task set using gaze control. The other 6 subjects started with gaze control followed by keyboard and mouse. None of the 10 student participants had used Second Life or gaze control previously.

Each subject was given a 15 minute introduction to Second Life in the form of simple training exercises using a keyboard and mouse. This was followed by a 15 minute introduction to using gaze control. This also contained a series of simple training exercises.

Each task set began with the avatar standing by the main entrance to the building and the tasks were completed in the same order for all subjects, locomotion, object manipulation and application control. The task was first explained and then the subjects were asked to complete it. They were reminded of the tasks as needed during completion. After both task sets, subjects were given a brief questionnaire to complete. They were advised that they could withdraw from the trial at any time and that there were no penalty for doing so. No reward for participation in the trial was offered and the whole session took between 45-55 minutes to complete.

Analysis and Results

We identified four different categories of errors the participants made when performing the tasks. Videos of the trials were annotated using an open source video annotation application called Elan (ELAN, 2008). At the outset the data from one subject was marked up by two people and the consistency of the outcomes was checked. A number of minor adjustments to the original definitions of the error types were made as a result, otherwise there was a high degree of agreement between the analyses. The four main error catergories were the following:

1. Locomotion error – being of one of following

   - unintentional motion backwards (the gaze first moves through the 'move backwards' zone of the screen after glancing down to change into locomotion mode)

   - unintentional rotation - left or right (the person means to glance off screen to change modes when in 'no control' mode, but rotates instead)

   - turn overshoot (person deliberately turns while in 'locomotion' mode but turns too far and has to correct this.

   - walk overshoot (person tries to stop, but the change to 'no control' mode takes too long the person walks too far, and has to reverse)

2. Mode change error – an unintentional change of the mode; a subject tries to rotate while in locomotion mode – left or right – and changes mode by mistake by looking too far off screen

3. Accuracy error – a subject tries to click on a target but misses because of inaccurate pointing. If this resulted in the wrong selection being made this error included recovery from this.

4. Misunderstanding error - a subject misunderstands / mishears / forgets what to do – e.g. a subject goes in the wrong direction and later corrects the direction
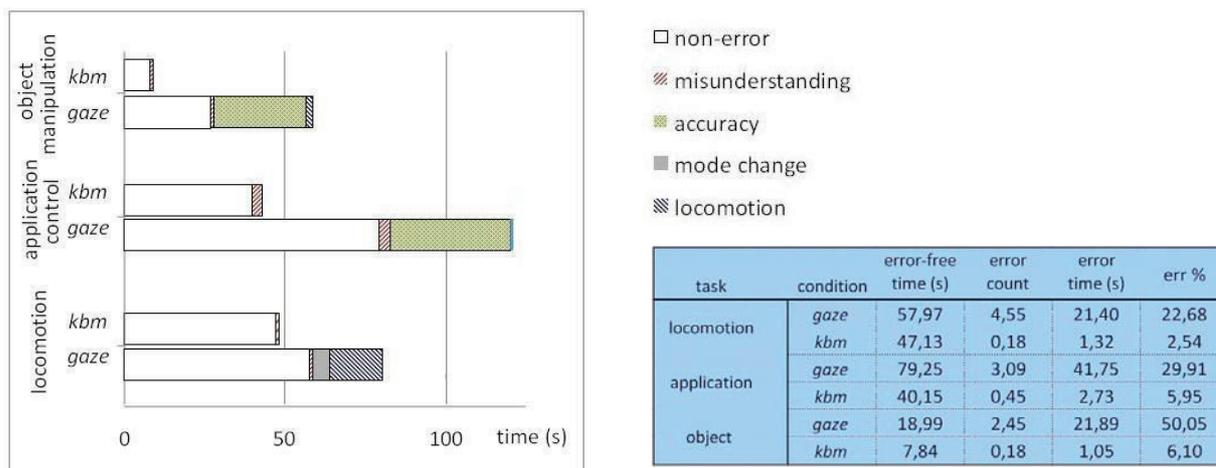
task	condition	error-free time (s)	error count	error time (s)	err %
locomotion	gaze	57,97	4,55	21,40	22,68
	kbm	47,13	0,18	1,32	2,54
application	gaze	79,25	3,09	41,75	29,91
	kbm	40,15	0,45	2,73	5,95
object	gaze	18,99	2,45	21,89	50,05
	kbm	7,84	0,18	1,05	6,10

Figure 2. Average task completion times partitioned into error times (in four types of errors) and non-error times

For each subject and for each task, the time spent in each error condition was summed and this was deducted from the total task time, leaving the non-error time for each trial.

The outcomes of the trials for each of the three tasks are shown in Figure 2. Each task represented one example from the three main different categories of tasks performed in virtual environment (Hand, 1997; Bowman, 1999). Data from the locomotion task is at the bottom of the graph, the application control task data is in the middle and the object manipulation task data is shown at the top. The total lengths of the bars show the average total task completion times in seconds including errors.

There were significant problems with calibrating the eyetracker for one subject. She was able to complete all 3 tasks in the gaze condition but there were far greater accuracy errors than for any other subject (the error time was more than 3 standard deviations from the mean of all other subjects' error times for the application control and object manipulation tasks). Consequently all data from this subject was removed from the analysis.

The results show that all subjects were able to complete the three tasks using eye gaze. The non-error part at the bottom of the bars enables comparison of task times if the cause of the errors can be removed by design changes. The gaze:keyboard-mouse ratio of non-error times for the locomotion task is 1:1.2. The corresponding ratio for the application control task is approximately 1:2, and for the object manipulation task is approximately 1:2.5.

The error-free times in the gaze condition are encouraging, particularly for the locomotion task. With only a short training session, subjects would be able to complete the locomotion task using gaze nearly as quickly as with key commands if the cause of the locomotion errors could be removed. The reasons behind the locomotion errors are in part due to the speed of movement of the avatar in response to key commands generated by the emulator. This causes overshoot or undershoot of movement which then have to be corrected. This is largely due to the processing pipeline on a single computer (eyetracker – emulator – Second Life browser, and additionally, in the experiment condition, the video capture software). There may also be optimisations to the emulator software that could improve performance here. Another source of locomotion errors is the location of the backwards motion zone at the bottom of the screen. This meant that the gaze position first had to travel through this zone after changing into locomotion mode and the latency in the system caused an unwanted backwards movement as a result. These can be addressed by modifications to the behaviour of the locomotion mode and examining in detail the causes for response latency.

The biggest cause of errors in the application control and the object manipulation tasks was the difficulty of hitting the small control objects in the dialog boxes to change appearance. This was exacerbated by

some latency in generating click events probably due to the processing pipeline. The best solution here to reduce these errors is probably to include some kind of zoom facility as is common with 2D gaze driven interfaces.

Subjects were asked in the post trial questionnaire to identify the most difficult aspects of the gaze control conditions. The majority of subjects said the application control task was the most difficult due to the lack of accuracy of gaze pointing. A smaller number said the slowness in generating click events, particularly in the object manipulation task, caused problems sustaining cursor position.

Conclusion

In summary, the study has been successful in revealing the extent and causes of performance differences between the gaze and keyboard-mouse conditions. It has enabled specific design changes to be identified to address these differences and has given an indication of the performance improvements that are likely to result from these changes. Importantly however, it has demonstrated the feasibility and potential for gaze based interaction with 3D virtual communities.

References

Istance, H., Bates, R., Hyrskykari, A., & Vickers, S. (2008). Snap Clutch, a moded approach to solving the Midas Touch problem. In *Proc. of the ACM Symposium on Eye Tracking Research & Applications (ETRA 2008),* ACM, New York, NY, 221-227,
DOI= http://doi.acm.org/10.1145/1344471.1344523

Bowman, D., Johnson, D., and Hodges, L. (1999) Testbed evaluation of virtual environment interaction techniques. In *Proc. of the ACM Symposium on Virtual Reality Software and Technology (VRST '99).* ACM, New York, NY, 26-33.   DOI= http://doi.acm.org/10.1145/323663.323667

Gilb, T. (1984) The "impact analysis table" applied to human factors design. In *Proc. Interact '84, First IFIP Conference on Human-Computer Interaction*, London, September 4-7, Vol. 2, pp. 97-101.

Hand, C. 1997. A Survey of 3D Interaction Techniques. *Computer Graphics Forum 16*, 5, 269-281.

ELAN (2008) A tool for the creation of annotations on video and audio resources. Freely available at http://www.lat-mpi.eu/tools/elan/