Comparison of head gaze and head and eye gaze within an immersive environment

Norman Murray and Dave Roberts
Centre for Virtual Environments,
Business House
University of Salford,
Salford, M5 4WT, U.K.
n.murray@salford.ac.uk d.j.roberts@salford.ac.uk

Abstract

For efficient collaboration between participants, eye gaze is seen as being critical for interaction. Tele-conferencing systems such as the AcessGrid allow users to meet across geographically disparate rooms but as of now there seems no substitute for face to face meetings. This paper gives an overview of some preliminary work that looks towards integrating eye gaze into an immersive Collaborative Virtual Environment and assessing the impact that this would have on interaction between the users of such a system. An experiment was conducted to assess the difference between users abilities to judge what objects an avatar is looking at with only head gaze being viewed and also with eye and head gaze data being displayed. The results from the experiment show that eye gaze is of vital importance to the subjects correctly identifying what a person is looking at in an immersive virtual environment. This is followed by a description of how the eye tracking system has been integrated into an immersive collaborative virtual environment and some preliminary results from the use of such a system.

1   Introduction

This paper presents an experiment designed to investigate the impact of combining eye and head gaze within a immersive Collaborative Virtual Environment (CVE). Currently most Immersive Projection Technology (IPT) displays only track the user's head and hand(s). This limits the amount of information that can be conveyed from one person to another when they are collaborating between IPTs. One aspect of interaction that could be used to enhance the communication across IPTs would be the integration of eye gaze. Eye gaze is an important interaction resource in collaboration but it is typically not supported in today's communication technology. Some of the uses we make of gaze include:

- gaze being used to direct the visual attention of others [3];

- gaze being used to determine your actions according to the gaze direction of those listening to you [6];

- gaze can be used to determine whether others are paying attention [18];

- gaze can be used alongside speech to address and prompt another speaker [10];

- gaze is used when handling objects [16];

- gaze may be used in proposing courses of action, for example, answering the phone or making a phone call [13].

Research needs to be done to as assess the impact that the inclusion of eye gaze into IPTs has and how it can support the above tasks. The experiment detailed here will examine the use of gaze for the first item in the above list, namely how gaze can be used to direct the visual attention of others.

Eye gaze can be maintained in some limited way with video based systems such as im.point [17]. This allows three users at different terminals to meet at a round virtual/real table, see Figure 1. The geometric positioning of the users around the table helps maintain the perception of eye contact between them. By using video based technology the gestures made at the table can be transmitted to the other viewers, although the users have limited movement and gaze direction due to the nature of the interface. With IPTs, video based technology cannot be currently used as the user is typically wearing shutter glasses that make it hard for the eyes to be seen and the user is usually very

mobile within the IPT making it hard to be tracked by video as they conduct there task. Systems such as the US's Tele-Immersion Challenge [14] and Blue-C immersive system [11] exist, but they cannot capture the level of detail required for eye tracking.

Figure 1. Immersive Meeting Point (im.point)

Research has been conducted into the use of eye gaze models to replicate eye gaze as a communicational resource in two and three way consversations. These use various techniques to replicate eye gaze such as basing the eye gaze movement according to when the person is speaking and when the person is listening based on research on face to face dyadic interaction [1, 2, 8]. Studies by Garau [4] and Lee [9] using these models showed that inferred gaze significantly outperforms random gaze. These studies have been conducted using a simple head and shoulders view of the avatar. A further study by Garau [5] examined the use of a shared immersive environment, where subjects used a HMD and a 4 sided IPT. All of the studies have shown that using accurate models of eye gaze improves the level of realism and can have a significant positive effect on the users responses to interaction. These experiments have used eye gaze models to simulate eye movement using conversation analysis and have typically been limited to only two or three users. To represent the user's actions accurately within an immersive system we need to be able to capture their eye gaze and use this to test how it affects collaboration and interaction within an immersive collaborative environment.

The experiment we conducted presented subjects with an avatar looking at objects within the environment. The subjects will see the avatar with head and eye tracking or with just head tracking, and try to ascertain which objects the avatar is looking at. We will then attempt to gauge whether eye tracking aids the subject within the experiment and whether the lack of eye tracking hinders the subjects during the experiment.

2   Experiment Goal

The aim of the experiment was to assess the impact that the inclusion of eye gaze could have on an avatar. We aim to contribute to the understanding of eye gaze as a communicational resource and demonstrate that it can be supported in some limited way in a telecommunication system that does not constrain the movement of the users. Currently within IPTs, the users are typically tracked via a tracker on their head and having either one or both of their hands tracked. The relation between the head and the body, i.e. what should happen to the avatar when the person looks from left to right is usually inferred from the position of the hands, i.e, is the user rotating their head about the neck or are they actually rotating their whole body. Current systems typically only track this limited form of movement by the user, but studies have shown that immersive CVEs provide a powerful means for communication [15]. There are systems that are becoming more common place that could track the users bodies in greater detail, for example, see [12] that utilises an optical tracking system to track hand movements for gesture recognition within a system. Such systems could be extended to track the user's arm and leg movements within an immersive CVE.

For the experiment we were to test the differences observed when a subject tried to distinguish what objects in a scene the avatar in the environment was looking, comparing head gaze to eye and head gaze combined. As there was to be no other form of communication between the subject and the avatar they were to view, it was sufficient that the subject should view prerecorded eye gaze and head information. This allowed greater repeatability with multiple test subjects giving their impression of the same scenario. Initially a user was set in front of the Barco Trace with the ASL eye tracker fitted and a log file was generated that recorded their head movements and eye gaze as they were told to look at the objects within the scene. During the experiment the subjects are replayed the movements of the user that had been previously recorded. There were several reasons for doing this. It would mean that each subject was shown the same gaze data each time for the experiment, so we can know for sure that each subject within the experiment was receiving the same data. We could also check the prerecorded data to make sure that the eye calibration was stable throughout and by adding gaze lines (see Figure 5) make sure that the correct data has been logged. The subject has no other forms of interaction with the avatar that they are viewing other than to see their head, eye and body movements. This would also mean that we would only need one experimenter and the subject to run the experiment, as having another user at the eye tracker would also require someone to monitor the eye trackers calibration state throughout the use of the eye tracker.

The initial test does not consider convergence. Although we accept this may be important it would require binocular eye tracking or simulation of the second eye movement based on the object of gaze and the distance to the object. This papers hopes to demonstrate the basic principles of using eye gaze to communicate interest in an object before we study improvements.

2.1 Apparatus

The hardware used for eye tracking is an ASL Model 501 head mounted eye tracker, see Figure 2. The head mounted optics allow the user to have more freedom of movement although this is still limited by the length of the cabling. The hardware consists of an eye camera and a near infra red emitter that illuminates the eye. The eye tracker measures the subject's line of gaze with respect to the head, that are output as x and y coordinates, and it also measures the pupil diameter. When the pupil diameter is zero this usually means that the user is blinking, although it could also be due to pupil recognition being lost for some other reason. The eye tracker operates at a sampling rate of 60Hz with a system accuracy of 0.5 degree across the visual angle, allowing unlimited head movement, a visual range of 50 degrees horizontally and 40 degrees vertically, and weighs 250 grams.

Figure 2. ASL H5 Head Mounted Eye Tracker

Figure 3 shows the IPT used, a Barco Trace system and its integrated VICON optical tracker (three of the five cameras mounted on the Trace can be seen at the top of the figure). A standard PC is used to drive the display with dual processor XENON 3.2 GHz with a Nvidia Geforce FX3000 graphics card and 2GB of RAM. The VICON system uses reflective markers for tracking, usually with five

markers placed on each item to be tracked. The current system tracks both handheld devices and the head. The users head is tracked by the markers on the StereoGraphics CrystalEyes active shutter glasses. The optical tracking system can track objects to approximately 1mm positional accuracy and rotational accuracy of less than 1 degree. The display is approximately 1.4m in width by 1.1m in height with a resolution of 1280 by 1024. The subjects only wore the shutter glasses during the experiment, there was no need to use a device to navigate the environment, although they could physically move in front of the display as their head was tracked.
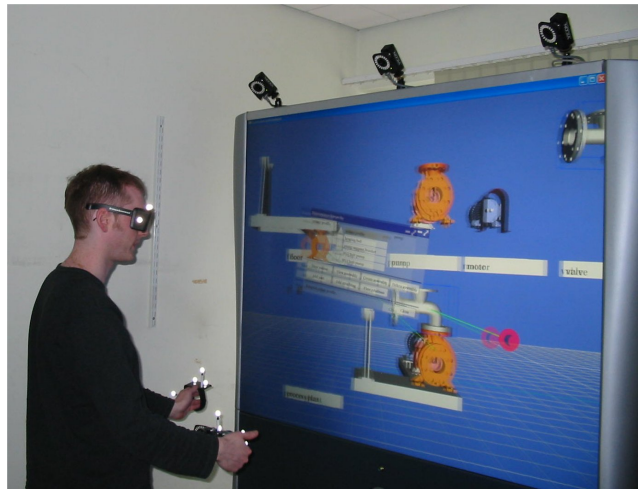
Figure 3. Barco Trace integrated with Vicon optical tracking system.

2.2 Software

The environment was coded in C++ and made use of the Performer scene graph from Silicon Graphics and the CAVELib library from VRCO to manage the tracking and display set up and ran on the SUSE Linux operating system.

The ASL's eye tracking and calibration software run on a standard PC running windows. So that the data from the eye tracker could be easily read into the virtual environment, it was decided to integrate the eye tracker into VRPN [7]. VRPN or the virtual reality peripheral network allows you to access data from a peripheral over the network. The ASL eye tracker was integrated in to VRPN. This allows the virtual environment to then access the data of the eye tracker by linking to the VRPN server. This was useful as the eye tracker is only provided with a Windows SDK and the experiment was conducted uuder Linux.

2.3 Environment

For the experiment, the subject stood in front of the Barco Trace (see Figure 3) and wore a pair of the CrystalEyes Stereoglasses so that they could view the scene in stereo. The glasses have optical markers on them so the scene is updated according to their head movements. The experiment environment is composed of a room with an avatar and a collection of objects in the room, see Figure 4. The subject is located in the room facing the avatar, with the objects located between them. The objects are located approximately 1.5m from the user with the avatar a further 1.5m away. The objects are located in a 3 by 3 layout spaced 0.5m apart.

Figure 4. The experiment environment with objects and avatar

Figure 5. Eye tracking shown with gaze lines

The avatar model was modified with the additon of two textured spheres for eyes. To create an avatar that blinked

the rear of the eyes were coloured the same as the skin. When subject blinked the eyes were simple rotated about 180 degrees so that the rear of the eyes was in view and as these were coloured the same as the avatar skin this provided a simple method of representing the avatar as blinking.

2.4 Procedure

The scenario for the experiment was explained to the participants. The subjects then viewed the avatar and marked on a sheet the object they thought the avatar was looking at. Subjects also give a value on a Likert 1...7 scale to specify how accurate they thought their response was. After a set of these, the subject filled in a questionnaire. The questionnaire presented to the subject attempted to elicit information such as how natural the behaviour of the avatar was, whether they could identify where it was looking, how expressive the avatar was and how well they think they completed there task. The subjects then repeated the experiment with another set of objects and a further questionnaire.

The subjects performed the experiment twice, seeing the head tracked avatar and the head and eye tracked avatar. Half the group would see the eye and head tracked avatar first and the other half of the group would see the head tracked avatar first.

3 Results

The experiment was conducted using 10 subjects. Using t-tests, it was found that the subjects managed to identify a significantly greater proportion of objects using eye and head gaze, as opposed to head gaze. Some results to note from this are:

- The combination of eye and head gaze results in a greater success rate in the correct identification of the object that the user is looking at. Typically subjects managed to identify either 8 or 9 out of the 9 components.

- Using just head gaze with eyes set to look forwards only subjects had very limited success in identifying the objects that the avatar would be looking at if eye gaze information was present. Subjects managed to identify between 1 and 3 of the 9 components.

- Typically with head gaze only tracking they are selecting the *correct* object that the avatar is looking according to the information that they are receiving, although they are less sure of these results as the avatar is not directly looking at the object. They are managing to correctly identify what objects the avatar is looking at according to the head gaze information.

Regarding the questionnaire that the subjects completed after the trials with eye gaze, using t-tests, the following were found to produce significant results:

- The subjects felt that they could readily tell where the avatar was looking when viewing the avatar.

- The subjects felt that the avatars actions reflected what the user was doing.

- The subjects felt that they had completed the task more fully.

These results will be discussed in the following section.

4 Discussion

From the results it has been shown that that combination of head and eye gaze contributes to the subjects correctly identifying the objects that the other avatar within the environment is looking at. Subjects had limited success when the correct eye gaze information was missing. These initial results show that it will be possible to use eye tracker information within IPTs successfully at least within the 3m range of the system at the current definition of the display.

When only head gaze information is being provided to the subjects, they cannot correctly identify the objects that the avatar is looking at. According to the information that is being provided to them, they are correctly identifying where the avatar's gaze is being directed, but as they are not being provided with eye tracking information, the avatar's eyes are always looking directly forwards. Within an IPT this could lead to confusion between the participants as the information that they are receiving is not matching what they are interpreting. For example, in the IPT, an avatar could be pointing to an object, talking about the object and looking at the object. If the avatar appears to be looking at a different object to the one they are pointing and/or talking about, this could cause confusion. Previous studies have shown how simulated eye gaze is preferable to stationary and randomly generated eye gaze [9]. The study compared stationary eye gaze with simulated eye gaze at a desktop interface and also found that the avatar with eye gaze appeared more natural and realistic. Further experiments could be conducted to see if we can measure the effects of stationary and eye tracked characters in more interactive environments. Where users combine speaking, pointing and eye tracking.

The subjective responses given in the questionnaire indicate that the subjects could readily tell where the avatar was looking. This led them to being confident that they had managed to complete the task fully. When the avatar's eyes did not move, the subjects were not confident of that they had managed to complete the task. The subjects are aware that the expressive eye gaze communication presented by the avatar is missing and is hindering their ability to compete the task. They also felt that the actions of the avatar reflected what the user was doing. These show that the subjects were aware when eye gaze was missing from the avatar, as reflected in their ability to identify the components being viewed by the avatar.

In review, this simple experiment has highlighted how eye gaze can significantly improve how user's can perceive what another avatar is viewing in an immersive virtual environment. This has led to subjects to . Implementing eye gaze within a distributed collaborative environment could lead to improved communication and task performance. In the following section we give an overview of our initial implementation of eye gaze within such a system.

5 Integration into a Distributed Collaborative Environment

Alongside the initial experiment into eye gaze, the hardware was also integrated into an immersive virtual environment. The collaborative virtual environment chosen was ICE [19]. As the eye tracking hardware had already been linked to VRPN, it was now only necessary to make ICE link to the VRPN to pick up the eye tracking values and modify the avatar so that the eyes would be displayed correctly. The avatars within ICE have their head and hand(s) tracked. The avatar code was modified so that the option of reading eye data from VRPN was added. This was then linked to the eye movements of the avatar. So that distributed users would see the local eye movements, it was necessary to share the eye transform data to each of the other users within the system so that they could observe the eye movements. So in addition to the avatar's head and hand movements, the avatar's eye movements are distributed to the other users in the virtual environment.

Formal experiments between other IPTs have not been conducted as none of the other owners of IPTs that we have connected to have eye tracking hardware. Informal sessions have been conducted with three users. One user is at the Barco Trace with eye, head and hand tracking, with the other users located in 4 wall IPT displays with head and hand tracking. Comments such as the avatar appearing to be more *alive* have been illicited. More formal tests are to be undertaken in the future.

Some problems have been noted with use of the eye tracking system within a CVE. Currently we use Stereographics CrystalEyes shutter glasses to obtain a stereo view within the 4 walled IPT displays and the Barco Trace that was used for the experiment. These glasses block approximately 65 per cent of the light that passes through them. Due to the loss of light and the bulky size of the shutter glasses it can make calibration of the eye tracker harder. If the subjects are wearing normal spectacles as well as the

shutter glasses the subject becomes much harder to calibrate due to reflections on the subjects spectacles. This can normally be removed by altering the angle of the reflector, the angle of the eye camera, or the angle of the spectacles the user is wearing (this would involve moving the arms of your spectacles so that they do not rest on the ears but angle down below the ears). These movements are limited due to the user wearing the Stereographics CrystalEyes shutter glasses. We have not yet tested the system within the 4 sided IPT but the user should be able to move around within the IPT with the addition of a long enough cable.

The eye tracking system does require constant attention by an overseer. The system works by shining near infrared light into the eye. The pupil of the eye then appears as a bright circle along with a brighter corneal reflection. Edge detection software is used to find the pupil outline and corneal reflection but the threshold levels for the pupil and corneal reflection edge detection need to be set by the operator of the eye tracker. As the user is wearing the eye tracker the operator must continually monitor the eye tracker for changes as fluctuations in the light level in the room can affect the tracking of the eye. The newest ASL eye tracker does include software that can monitor for these changes but they still recommend that a human operator would be preferable to monitor the eye.

5.1 Choosing an Eye Tracker

The eye tracker used for this experiment was chosen due to its availability, rather than its suitability for the task. In this section we will provide an overview of the main eye tracking technologies available and their appropriateness for the task. The main types of eye tracker are:

- Video oculography pupil only tracking. These systems track the pupil using a camera.

- Video oculography pupil and corneal reflection. Typically an infra red source is shone into the eye to track the pupil and a corneal reflection. This allows for the subjects gaze to be measured on a suitable surface on which calibration points are displayed. Using the pupil and corneal reflection, the system can separate out eye movements from head movements.

- Video oculography dual Purkinje image corneal reflection. These measure 2 corneal reflections and have the benefit of tracking the translational and rotational eye movements.

- Video oculography limbus, iris-sclera boundary. These systems track the iris-scelera boundary by measuring the reflection of infra red light by the area on both sides of the edge between the white sclera and the darker iris.

- Electrooculography electro-potential about the eye. Electrooculography is an electrical method of recording eye movements. Tiny electrodes are attached to the skin at the inner and outer corners of the eye, and as the eye moves an alteration in the potential between these electrodes is recorded.

- Contact lens scleral coil in the eye. A wire coil is placed in the contact lens in the eye and its movements through an electromagnetic field are measured. Other contact lens systems have used reflective phosphors or line diagrams.

The contact lens method of tracking the eyes are obviously to invasive for use in VE. This leaves electrooculography and video oculography. For video oculography a pupil only tracking system would be preferable as typically within VEs the user's head is tracked independently, gaze information is not required from the system as this can be calculated from the head and eye position data. Other factors to take into account when choosing a system are its accuracy, ease of set up, sampling rate, mobility and fit with glasses for stereo viewing.

The eye tracker used for the experiment used a pupil and corneal reflection tracker providing gaze coordinates rather than eye position. We are currently in the process of choosing an eye tracker to suit the needs of developing this research.

6 Summary

This paper has outlined the need to research the utility of eye gaze in IPTs for collaborative working. Eye gaze has been shown to be a key interaction resource in collaboration and we have argued that it is not well supported by today's current communication technology. IPTs allow the recreation of a virtual environment where users can interact and perform tasks even though they can be located at geographically diverse locations. Current systems typically support the tracking of the user's head and hands. Within this paper we have performed a preliminary experiment to examine the benefits of the inclusion of eye gaze information into an avatar. The results indicate that eye gaze information could be extremely beneficial for communication within immersive collaborative virtual environments. The experiment performed showed that the subjects were able to distinguish what components avatars were looking at within the environment. The eye tracker and display were sufficient for subjects to accurately identify the viewed components. When eye gaze information was missing the subjects could not identify the object the avatar was looking at. The addition of eye gaze also made the avatars appear more natural and realitic to the subjects.

We have also integrated the eye tracking system into a CVE, namely ICE to support eye gaze in a distributed team. However at present we only have one eye tracker. The system has been tested by linking the eye tracked user at a Barco Trace with two others users at 4 sided CAVE style displays, one located in Salford and the other in Reading.

Eye gaze is an important communication resource for collaboration and is required for many varying forms of communication from directing visual attention, prompting people and handling objects. The work here has shown that eye gaze is required for users to correctly identify users gaze direction. Further research needs to be conducted to examine the utility of eye gaze within IPTs for collaborative tasks. These will identify where eye gaze is important, for what tasks it is important and where it is critical for collaboration at a distance. Also to be examined is the suitability of the currently available commercial eye trackers for use within immersive virtual environments.

References

[1] M. Argyle. *Bodily Communication*. Methuen and Co, London, 1975.

[2] M. Argyle. *Bodily Communication*. Methuen and Co, London, 2nd edition, 1988.

[3] M. Argyle and J. Graham. The Central Europe Experiment - looking at persons and looking at things. *Journal of Environmental Psychology and Nonverbal Behaviour*, pages 6–16, 1977.

[4] M. Garau, M. Slater, S. Bee, and M. A. Sasse. The Impact of Eye Gaze on Communication using Humanoid Avatars. In *Proceedings of the SIG-CHI conference on Human factors in computing systems*, 2001.

[5] M. Garau, M. Slater, V. Vinayagamoorhty, A. Brogni, A. Steed, and M. A. Sasse. The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. In *Proceedings of the SIG-CHI conference on Human factors in computing systems*, 2003.

[6] C. Goodwin. Action and Embodiment Within Situated Human Interaction. *Journal of Pragmatics*, pages 1489–1522, 2000.

[7] R. M. T. II, T. C. Hudson, A. Seeger, H. Weber, J. Juliano, and A. T. Helser. VRPN: A Device-Independent, Network-Transparent VR Peripheral System. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 2001.

[8] A. Kendon. Some Functions of Gaze-Direction in Social Interaction. *Acta Psychologica*, 26:22–63, 1967.

[9] S. H. Lee, J. B. Badler, and N. I. Badler. Eyes Alive. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pages 637–644, 2002.

[10] G. Lerner. Selecting next speaker: the context-sensitive operation of a context free organization. *Language in Society*, pages 177–201, 2003.

[11] G. Markus and S. Wurmlin. Blue-c: A Spatially Immersive Display and 3D Video Portal for Telepresence. In *Proceedings ACM SIGGRAPH*, 2003.

[12] N. Murray, J. Y. Goulermas, and T. Fernando. Visual Tracking for a Virtual Environment. In *Proceedings of HCI International*, volume 1, pages 1198–1202, 2003.

[13] J. P. Rae. Organizing Participation in Interaction: Doing Participation Framework. *Research on Language and Social Interaction*, pages 253–278, 2001.

[14] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays. In *Proceedings ACM SIGGRAPH*, 1998.

[15] D. J. Roberts, R. Wolff, O. Otto, A. Steed, D. Kranzlmueller, and C. Anthes. Supporting Social Human Communication between distributed walk-in displays. In *Proceedings of the ACM International Symposium on Virtual Reality, Systems and Technology*, 2004.

[16] J. Streek. How to Do Things With Things. *Human Studies*, pages 365–384, 1996.

[17] R. Tanger, P. Kauff, and O. Schreer. Immersive Meeting Point (im.point) - An Approach Towards Immersive Media Portals. In *Proceedings of Pacific-Rim Conference on Multimedia*, 2004.

[18] R. Vertegaal, G. van der Veer, and H. Vons. Effects of Gaze on Multiparty Mediated Communication. In *Proceedings of Graphics Interface*, 2000.

[19] R. Wolff. Reconfigurable event handling for cve systems. Technical Report Salford-CVE-Roberts-Wolff-TR-2005-07-29, Centre for Virtual Environments, University of Salford, Sept. 2005.