# A Preliminary Investigation into Eye Gaze Data in a First Person Shooter Game

Alan Kenny, Hendrik Koesling, Declan Delaney
Department of Computer Science
National University of Ireland, Maynooth
Maynooth, Co. Kildare, Ireland
E-mail: akenny@cs.nuim.ie

Seamus McLoone and Tomás Ward
Department of Electronic Engineering
National University of Ireland, Maynooth
Maynooth, Co. Kildare, Ireland

**KEYWORDS**
Eye Tracking, Networked Games, Psycho-Perceptual Measures

**ABSTACT**

This paper describes a study carried out in which the eye gaze data of several users playing a simple First Person Shooter (FPS) game has been recorded. This work shows the design and implementation of a simple game and how the execution of the game can be synchronized with an eye tracking system. The motivation behind this work is to determine the existence of visual psycho-perceptual phenomena, which may be of some use in developing appropriate information limits for distributed interactive media compression algorithms. Only 2 degrees of the 140 degrees of human vision has a high level of detail. It may be possible to determine the areas of the screen that a user is focusing on and render it in high detail or pay particular attention to its contents so as to set appropriate dead reckoning limits. Our experiment shows that eye tracking may allow for improvements in rendering and new compression algorithms to be created for an online FPS game.

## 1. INTRODUCTION

Networked interactive computer games are an important class of Distributed Interactive Application (DIA). These games generate a large quantity of data that must be communicated across the physical network between participating game nodes to allow the game state to be replicated at remote nodes. However, network bandwidth is a limited resource, being particularly scarce at typical end-user locations. A key research area in distributed computer games is, therefore, concerned with devising new techniques and methodologies that reduce the amount of data that needs to be generated. In this way the available bandwidth is exploited fully, without loss of consistency for the participants. One popular technique introduced with the Distributed Interactive Simulation (DIS) standard (IEEE 1993) and used in games such as Doom and Tribes is dead reckoning (Pantel and Wolf 2002). This compresses user dynamics information by down-sampling the user trajectory. Other techniques employed in DIAs include the hybrid strategy model (Delaney et al. 2003), area of interest management, data compression and dynamic load balancing (Singhal and Zyda 1999).

More recently in distributed media such as video or music, psycho perceptual phenomena have been employed to reduce the quantity of data that needs to be communicated without impacting the quality of the end-user experience. Examples include the mp3 standard for music and the MPEG or DivX standards for video. However, in DIAs such as online computer games the exploitation of psycho-perceptual parameters has received limited attention. This paper is part of a larger body of work that hypothesises that knowledge of player psycho-perception can be exploited to reduce the amount of data that needs to be transmitted between users in distributed games. Visual information processing is one of the most important components of how a user perceives a gaming experience. The contribution of this paper is that it describes how eye movement data can be recorded and synchronized with game play in an action game developed using the Torque games engine.

Eye tracking is one area of research that has yet to be explored in this context, to the best of our knowledge, and has the potential to yield interesting and useful psycho-perceptual results. Eye movements in static scenes, such as a computer screen, are not performed continuously, but consist of "jumps" (saccades) and "rests" (fixations). Eye-tracking results are represented as lists of fixation data. This data contains, for example, the fixation position and duration, the current pupil size and the start and end times of each fixation. Previous work on decision making in soccer match (Koesling and Höner 2003) has shown distinct pre- and post-decisional phases of user action: once a target player has been chosen, the visual field narrows around this target for post-decisional action planning. The underlying action theory for decision-making is known as the Rubicon Theory and the point that separates the pre- and post-decisional phases is known as the Rubicon point (Gollwitzer and Bayer 1999). It is our goal to determine if similar results could be obtained with respect to computer games. This would enable developers to determine entities and regions of the game environment on which players focus their attention. Such information could be used to determine scene rendering priorities, level of rendering detail and load balancing, as well as providing a parameter for assigning priorities to entity extrapolation in compression techniques such as dead reckoning or the hybrid strategy model.

The remainder of this paper is structured as follows: section 2 describes the experiment and how the data acquisition was achieved. The data is analysed and discussed in detail in section 3. Finally conclusions are made along with some proposals for future work in section 4.

## 2. METHODOLOGY

The system consisted of an SR Research EyeLink®2 eye tracker, a control Personal Computer (PC) and a display PC. The EyeLink®2 is a binocular video-based eye-tracking system with a sampling rate of up to 500Hz. During experiments, subjects wear the eye-tracker headset while they view stimuli on a computer monitor screen (Figure 1). Small infrared (IR) cameras on the headset transmit information about the subjects' head and pupil positions to the eye tracker. From this data the eye tracker calculates and records the positions subjects look at on the screen. The control PC hosts the eye tracker operational software and is used to control and calibrate the eye tracker. The experiment is run on the display PC, which is connected to the control PC via an Ethernet connection. The display PC for these experiments was a laptop with an external monitor, mouse and keyboard. Four IR markers were attached to the external monitor in order for the eye tracker to compensate for head movements. The advantage of using an external monitor, keyboard and mouse is that it provides the operator with real-time control over the experiment, so that the eye tracker software can be started and stopped without interfering with the subjects controls.

Figures 1: Experiment Set-up

A First Person Shooter game was created using the Torque Game Engine (http://www.garagegames.com). Torque is an industrial game engine that has been used to produce many games (Marshall et al. 2004) including the award winning Tribes 2. The objective of the game was to get as high a score as possible. A user could increase their score by destroying enemy bots or by collecting tokens, but in order to achieve a high score more difficult bots had to be attacked. Each game lasted two minutes, with a brief period for synchronisation at the start and end of the game. Six volunteers participated, each playing the game twice; the first run was considered a learning experience. This resulted in a total of twelve data sets. Through the use of questionnaires it was discovered that each volunteer had a good level of computer expertise. Three volunteers had keen interest in video games and were quite proficient at them; the remainder had relatively little experience.

One of the greatest challenges encountered in performing the experiment was the synchronisation between the eye tracker software and the FPS game. These software packages do not have easily accessible interfaces and the communication and processing of data is dictated by the real-time constraints of the experiment. Once the eye tracker program is initialised and the calibration procedures have finished, the external Torque application is executed. A “STARTING GAME” message is sent to the eye tracker file so that the recorded eye data can be synchronized with the application time. In addition, just prior to the game play commencing and finishing, a set of animated points are displayed on screen, each lasting approximately 900ms. The user is asked to fixate on these points. By locating this pattern in the eye tracker data, the data relating to the start and end of game play can be isolated.

Three sets of data were recorded for each trial: the eye tracker data, the Torque recording and a custom Torque log file. Eye tracker data was recorded every 4ms. The Torque recording is a playback of game events based on network samples and each recording was later recorded as an AVI file using FRAPS. FRAPS is a utility specially designed for recording game footage (further information on FRAPS can be found at http://www.fraps.com/). The Torque log records entity position every 500ms in addition to all events associated with items, weapons, damage and kills.

## 3. ANALYSIS

In Figure 2 a sample image is shown, which superimposes the computer game footage with markers representing eye fixations. The darker cross shows the left eye fixation point and the dark triangle shows the right eye fixation point. In order to generate such superimposed video, the eye data relating to the game play session had to be extracted by hand using the synchronization mechanism mentioned in section 2. Perl and Matlab® scripts were then developed to strip out and recombine the information from the two separate data.


Figure 2: Eye Fixation Data

Observations of initial results from this experiment are summarized in the following paragraphs.

The crosshair in a FPS game effectively creates a natural fixation point. This coupled with the fact that the user effectively controls the worldview with the crosshair, results in the majority of fixations taking place near the centre of the screen, where 'near centre' is taken to be the inner 400x300 rectangle from the 800x600 resolution screen. In essence the control of the natural fixation point around the point of most activity in the game results in the user using this as a virtual eye. Nearly 88% of all fixations fall within this region as shown in Figure 3. This results in the users spending 86% of fixation time and 82% of the game time within the near centre region. Interestingly the regions of the screen representing the user's health, message box and score received very little eye fixations, only 2% of all fixations fall within these regions of which one user contributed 53%.

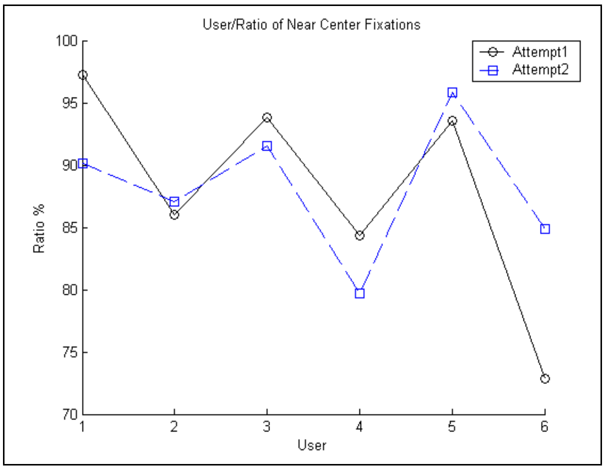


Figure 3: Ratio of Near Centre Fixations to Total Fixations for the Twelve Data Sets

This phenomenon could be exploited by graphics rendering algorithms (O' Sullivan and Dingliana 2001). Areas within the near centre region could be rendered in high quality with the periphery being rendered in a lower quality, as the human peripheral vision is less aware of the lower quality graphics. It has been shown in previous research that the human eye only focuses on a small area of a screen in high detail. This result, based on human perception capabilities, can be used to further compress real time MPEG video streams, which have appropriately-paced content and require high resolution gain in the eye gaze window (Komogortsev and Khan 2004). Research in level of detail (LOD) management has also shown that a user's subjective opinion of a scene is not hampered by lowering the detail of non-important objects in that scene (Brown et al. 2003). This would reduce processing requirements in graphically intense FPS titles, such as Doom 3 and Half-life 2, allowing developers to create graphically detailed titles that reach a wider audience.

Additionally, the human eye is more likely to notice inconsistencies (i.e. due to latency, jitter and packet loss) that occur around the points of fixations. As most fixations take place within this near centre region for a FPS game it is likely that any inconsistencies that occur within this region are more likely to be noticed than those that occur outside this region. In networked online games this could be exploited by "area-of-interest" algorithms to reduce the update packets that need to be exchanged between participants.

Previous work on a set of soccer videos has shown that there exists up to 600ms of delay from when a person makes a decision about executing a task as indicated by eye movement and when the task is actually executed (Koesling et al. 2001). In those experiments users had only to make a single decision while viewing a series of six-second video clips. However, in the experiments described here the nature of a FPS game means that the time between user decisions can be very short. This makes it difficult to consistently identify the Rubicon point. As a consequence, it may be difficult to utilize eye data in order to pre send information during online multiplayer games. However, other game genres that do not require twitch actions may be more suitable to using eye data for predicting users' actions. Further experiments to investigate the decision-making process in games would have to be conducted to test this hypothesis.

Our results also indicate that, in general, users exhibit fixation durations that are comparable to those of other media (Sibert and Jacob 2000). In Figure 4 the average left eye fixation duration for each session is shown (a similar graph can be derived from the right eye fixations). As can be seen the average fixation duration for most users falls within the range of 300ms and 750ms, with only two outliers. These outliers belong to an extremely experienced FPS player, which may explain their occurrence. The results appear to contradict the information displayed in the

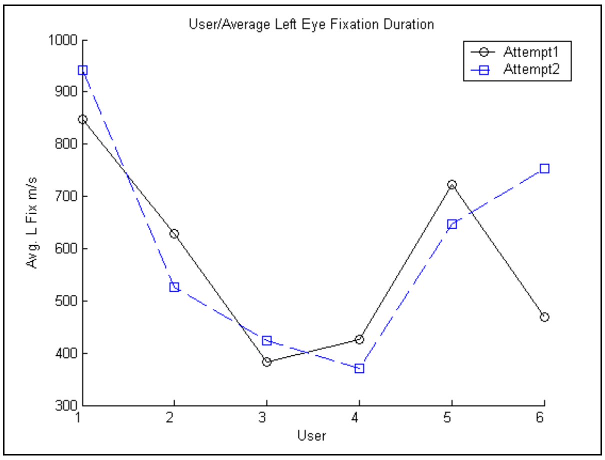


Figure 4: Average Left Eye Fixation Duration

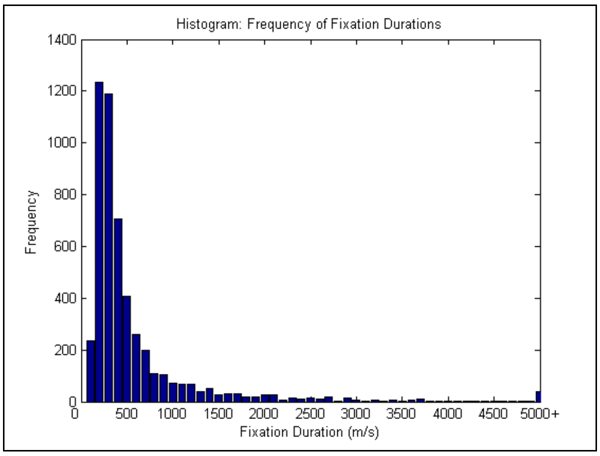


Figure 5: Frequency Distribution of Fixation Durations

superimposed videos, where users seem to have relatively long fixation durations (fixations of up to ten seconds being measured). This can be explained by the fact that most users (except the previously noted outliers) have more fixations under 300ms than between 300ms and 600ms or over 600ms as seen in figure 5. The average fixation duration minus the fixations under 300ms is nearly 1150ms, which more accurately represents the data that was observed in the videos.

Another point of interest is the tendency for higher game scores to result in longer fixation durations and consequently a lower number of fixations as shown in figure 6. It can also be seen that players didn't necessarily perform better on their second attempt. While the scoring mechanism used within the game may not perfectly represent task proficiency all subjects exhibited higher scores for games with longer average fixation durations. More accurate measures of task proficiency would have to be developed in order to verify this. There may also be a correlation between the high percentages of near centre fixations, the duration of the average fixation and the user's score. Using a player's score as an indication of their abilities, the three best players had an average of 91% near centre fixations and average fixation duration of just over 700ms. Additionally the worst players had an average near centre fixation of 84% and average fixation duration of just less than 470ms. Eye movement research has already suggested that high frequency of fixations in a region coupled with long durations indicates high task difficulty and can be taken as an indication of information complexity, similarly a smaller number of saccades indicates higher mental load (Pan et al. 2004). In relation to our FPS game this may be explained as the more proficient players having greater control over the virtual eye. As the task of scoring involved disabling enemies and this is accomplished by focusing the crosshair on the target it may indicate that the better a player becomes at supplementing their own eye movement with virtual eye movement the more accurately they are able to focus on the activity around the crosshairs. Further research would have to be carried out to determine if this is the case.

Table 1: Statistics Summary

| User | Score | Near Centre Fix. % | Time in Near Centre Region % | Avg. L. Fix. ms | Num. Fix. Under 300 ms | Num. Fix. Between 300 ms – 600 ms | Num. Fix. Over 600 ms |
|---|---|---|---|---|---|---|---|
| 1.1 | 3350 | 97 | 99 | 847 | 62 | 100 | 125 |
| 2.1 | 3700 | 85 | 95 | 628 | 194 | 93 | 84 |
| 3.1 | 1200 | 93 | 96 | 383 | 368 | 131 | 83 |
| 4.1 | 2100 | 84 | 88 | 426 | 302 | 155 | 91 |
| 5.1 | 2500 | 93 | 97 | 723 | 148 | 76 | 103 |
| 6.1 | 1100 | 72 | 86 | 468 | 249 | 141 | 78 |
| 1.2 | 5800 | 90 | 98 | 941 | 97 | 60 | 97 |
| 2.2 | 2800 | 87 | 92 | 526 | 220 | 140 | 87 |
| 3.2 | 1500 | 91 | 96 | 424 | 321 | 144 | 79 |
| 4.2 | 1200 | 79 | 85 | 370 | 368 | 167 | 81 |
| 5.2 | 2150 | 95 | 96 | 647 | 157 | 113 | 91 |
| 6.2 | 1900 | 85 | 90 | 754 | 165 | 58 | 88 |

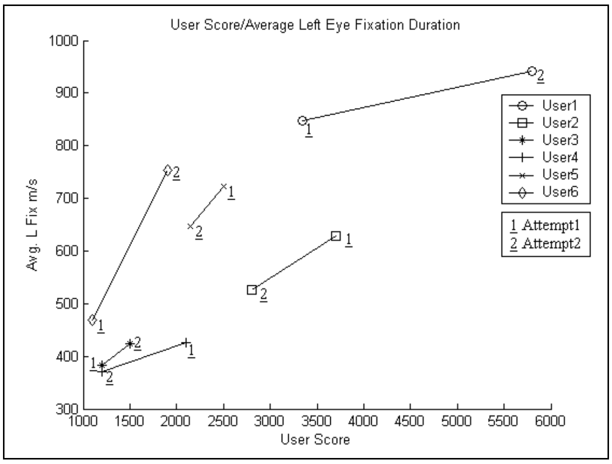


Figure 6: Average Left Eye Fixation Versus Score

## 4. CONCLUSIONS AND FUTURE WORK

In this paper a system for synchronising an eye tracker with a commercial game development engine has been described. Preliminary results suggest that the system successfully co-registers eye gaze information with the visual scene presented to the user during their game play. Analysis of the data indicates that users of a FPS spend an extremely large proportion of time focused around the centre of the screen, that there is some indication of the Rubicon theory taking place within games and that there may be a correlation between task proficiency and the duration of eye movement.

Future work will investigate psycho-perceptual phenomenon that can be used to filter the information that needs to be transmitted in DIAs such as networked computer games to maintain an adequately consistent global shared state. Specifically if the point of the Rubicon can be accurately and continuously determined then some information about user actions could be sent before they take place. While our results have shown that it may be difficult to achieve this in a fast action FPS game, other game genres with slower interactions or an FPS with more controlled interactions may prove easier to analyse. Future experiments will aim to more accurately highlight the Rubicon effect within games.

## REFERENCES


Brown, R., Copper, L. and Pham, B. 2003. “Visual Attention-based Polygon Level of Detail Management.” *First International Conference on Computer Graphics and Interactive Techniques, Melbourne, Australia.*

Delaney, D., Ward, T. and McLoone, S. 2003. “Reducing Update Packets in Distributed Interactive Applications using a Hybrid Model.” *16th International Conference on Parallel and Distributed Computing Systems*, Reno, USA.

Gollwitzer, P.M., and Bayer, U. 1999. “Deliberative versus implemental mindsets in the control of action.” *Dual-process theories in social psychology (pp. 403-422)*. New York: Guilford.

IEEE 1993. IEEE Standard for Distributed Interactive Simulation - Application Protocols, in IEEE Std 1278-1993, I.C. Society, Editor. 1993, IEEE: New York.

Koesling, H. and Honer, O. 2003. “Bend it like Beckham – Mindsets and Visual Attention in Decision-Making in Soccer.” *Proc. 12th European Conference on Eye Movements*, Dundee, UK.

Komogortsev, O., and Khan, J. 2004. “Predictive Perceptual Compression for Real Time Video Communication.” *Proc. 12th ACM International Conference on Media, New York, USA.*

Marshall, D., Delaney, D., McLoone, S. and Ward, T. 2004. “Exploring the Spatial Density of Strategy Models in a Realistic Distributed Interactive Application.” *International Symposium on Distributed Simulation and Real-Time Applications*, Budapest, Hungary.

O’Sulllivan, C. and Dingliana, J. 2001. “Collisions and Perception.” *ACM Transactions on Graphics, Volume 20, Issue 3.*

Pan, B. et al. 2004. “The Determinants of Web Page Viewing Behaviour: An Eye Tracking Study.” *Proc. Eye Tracking Research and Applications Symposium, San Antonio, Texas.*

Pantel, L. and Wolf, L. 2002. “On the Suitability of Dead Reckoning Schemes for Games.” *First Workshop on Network and System Support for Games*, Braunschweig, Germany.

Sibert, L. and Jacob, R. 2000. “Evaluation of Eye Gaze Interaction.” *Proc. SIGCHI conference on Human Factors, The Hague ,Netherlands* .

Singhal, S. and Zyda, M. 1999. “Networked Virtual Environments: Design and Implementation.” *ACM Press SIGGRAPH Series*. Addison-Wesley, Reading, Massachusetts.


## AUTHOR BIOGRAPHIES

**ALAN KENNY** was born in Mullingar, Ireland and went to NUI, Maynooth, where he studied computer science and software engineering and obtained his degree in 2004. Having spent a short stint working as a web developer he returned to the University to study a PhD in the area of psycho-perceptual measures in networked games. His e-mail address is: akenny@cs.nuim.com and his Web page can be found at http://www.cs.may.ie/~akenny.

**HENDRIK KOESLING** graduated in computer science and physics (M.Sc.) in 1997 and received a PhD in computer science from Bielefeld University, Germany, in 2003. He worked as a lecturer and researcher in the Neuroinformatics Group and the CRC360 "Situated Artificial Communicators" at Bielefeld University until 2004. He currently holds a postdoctoral research position at the Computer Science Department at NUI Maynooth. His research interests are in visual perception/attention and cognitive processes in humans, based on the analysis and modeling of eye movements.

**DECLAN DELANEY** received his electronic engineering degree from University College Dublin in

1991 and his Masters degree in 1999. He has worked abroad as both a software engineer and project manager. At present he lectures in the Computer Science Department in NUI Maynooth. His current research interests focus on techniques for reducing latency in distributed networked applications and on the use of small group PBL in software engineering education. He is a chartered member of the IEI.

**SEAMUS McLOONE** received his B.Eng. honours degree in Electrical and Electronic Engineering from Queen's University Belfast in 1996 and his Ph.D., entitled Nonlinear Identification using Local Model Networks, in September 2000. At present, he is employed as a lecturer in the Department of Electronic Engineering in the National University of Ireland (NUI), Maynooth. His main research interests include multiple model approaches to nonlinear system identification and reducing and masking the effects of latency in distributed interactive applications (DIAs).

**TOMAS WARD** received the B.E., the M.Eng.Sc. and Ph.D degrees from University College, Dublin, Ireland in 1994, 1996 and 1999 respectively. He worked as Research Coordinator in the Rehabilitation Engineering Department at the National Rehabilitation Hospital from 1996-1999. From 1999-2000 he was an Assistant Lecturer in the Department of Computer Science at the National University of Ireland, Maynooth. In 2000 he became a lecturer in the Department of Electronic Engineering at NUI Maynooth. His main research interests lie in the areas of near-infrared optical and terahertz medical imaging, the development of therapeutic tools in medical rehabilitation, emotion synthesis in artificial speech and distributed environments for multi-user interaction.