FOI, Swedish Defence Research Agency, is a mainly assignment-funded agency under the Ministry of Defence. The core activities are research, method and technology development, as well as studies conducted in the interests of Swedish defence and the safety and security of society. The organisation employs approximately 1250 personnel of whom about 900 are scientists. This makes FOI Sweden's largest research institute. FOI gives its customers access to leading-edge expertise in a large number of fields such as security policy studies, defence and security related analyses, the assessment of various types of threat, systems for control and management of crises, protection against and management of hazardous substances, IT security and the potential offered by new sensors.

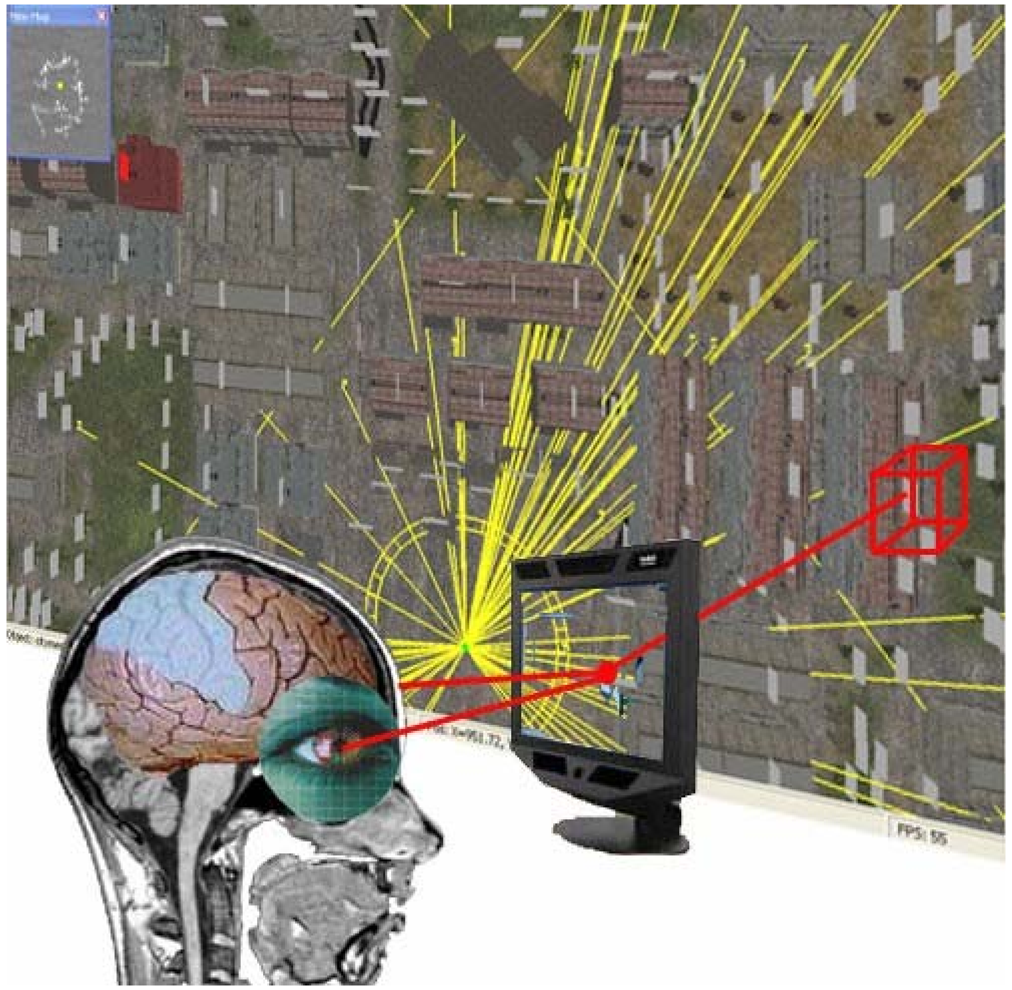
FOI
Defence Research Agency
Command and Control Systems
P.O. Box 1165
SE-581 11 Linköping

Phone: +46 13 37 80 00
Fax: +46 13 37 81 00

www.foi.se

Verification of an Experimental Platform Integrating a Tobii Eyetracking System with the HiFi Game Engine

CHARLOTTE SENNERSTEN, JENS ALFREDSON, MARTIN CASTOR, JOHAN HEDSTRÖM, BJÖRN LINDAHL, CRAIG LINDLEY, ERLAND SVENSSON



FOI-R--2227--SE Methodology report **Command and Control Systems**
ISSN 1650-1942 February 2007

CHARLOTTE SENNERSTEN, JENS ALFREDSON, MARTIN CASTOR, JOHAN HEDSTRÖM, BJÖRN LINDAHL, CRAIG LINDLEY, ERLAND SVENSSON

Verification of an Experimental Platform Integrating a Tobii Eyetracking System with the HiFi Game Engine

FOI-R--2227--SE
ISSN 1650-1942

Command and Control System
Methodology report

February 2007

Utgivare
FOI - Totalförsvarets forskningsinstitut
Ledningssystem
Box 1165
581 11 Linköping

Rapportnummer, ISRN
FOI-R--2227--SE

Klassificering
Metodrapport

Forskningsområde
8. Människa och teknik

Månad, år
Februari 2007

Projektnummer
E7096

Delområde
81 MSI med fysiologi

Delområde 2

Författare/redaktör
Charlotte Sennersten (BTH/HGO)
Jens Alfredson
Martin Castor
Johan Hedström
Björn Lindahl
Craig Lindley (BTH/HGO)
Erland Svensson

Projektledare
Erland Svensson

Godkänd av
Lena Bergvin

Uppdragsgivare/kundbeteckning
FM HKV

Tekniskt och/eller vetenskapligt ansvarig
Erland Svensson

Rapportens titel
Verification of an Experimental Platform Integrating a Tobii Eyetracking System with the HiFi Game Engine

Sammanfattning (högst 200 ord)
Att spela ett kommersiellt PC eller konsoll spel är en mycket visuell aktivitet, oavsett om syftet är spelinteraktion som underhållning eller situerad inlärning vilket diskuteras i `Serious Games´området. Om mer information av spelarens visuella uppmärksamhet kunde spelas in och analyseras på ett lätt sätt, skulle viktig design information kunna utvinnas. En rad olika ögonrörelsemätningsutrustningar existerar på marknaden och har använts i många studier i många år. Trots detta beskriver väldigt få studier den visuella interaktionen hos användaren/spelaren då ett 3D objekt förflyttar sig på datorskärmen. Anledningen är att metoder och mjukvara som är utvecklat för ögonrörelse studier för statiska 2D stimuli inte är tillämpliga för dynamiska 3D stimuli och att manuell analys av dynamisk 3D visuell interaktion är väldigt tidsödande. För att belysa detta så har författarna utvecklat ett mjukvaru gränssnitt mellan Tobii™ ögonrörelsemätningssystem och HiFi spelmotorn för att använda automatisk loggning av dynamiska 3D objekt. Denna rapport beskriver den utförda verifikations studien för att bedöma prestandan av denna integrering mellan ögonrörelsemätningutrustning, loggnings verktyg och spelmotor. Detaljerad analys visar att goda resultat inom de givna förutsättningarna vilket medger såväl småskaliga som de storskaliga studier som behövs för omfattande statistiska analyser. Arbetet som presenteras i rapporten har genomförts i samarbete mellan FOI, Blekinge Tekniska Högskola och Högskolan på Gotland.

Nyckelord
Ögonrörelser, spel, spelinteraktion, design, visuell uppmärksamhet, instrumentering, beteendevetenskap, kognition, mätmetodik

Övriga bibliografiska uppgifter

Språk Engelska

ISSN 1650-1942

Antal sidor: 54 s.

Distribution enligt missiv

Pris: Enligt prislista

Issuing organization
FOI – Swedish Defence Research Agency
Command and Control Systems
P.O. Box 1165
SE-581 11 Linköping

Report number, ISRN
FOI-R--2227--SE

Report type
Methodology report

Programme Areas
8. Human Systems

Month year
February 2007

Project no.
E7096

Subcategories
81 Human Factors and Physiology

Subcategories 2

Author/s (editor/s)
Charlotte Sennersten (BTH/HGO)
Jens Alfredson
Martin Castor
Johan Hedström
Björn Lindahl
Craig Lindley (BTH/HGO)
Erland Svensson

Project manager
Erland Svensson

Approved by
Lena Bergvin

Sponsoring agency
Swedish Armed Forces HQ

Scientifically and technically responsible
Erland Svensson

Report title (In translation)
Verification of an Experimental Platform Integrating a Tobii Eyetracking System with the HiFi Game Engine

Abstract (not more than 200 words)
Playing a commercial PC or consol game is a highly visual activity, regardless of whether the purpose is entertainment or situated learning as discussed in the Serious Games field. If more information about the visual attention of the player can be recorded and easily analysed, important design information can be extracted. A range of different eyetracking equipment exists on the market and has been used in many studies over the years. However, very few studies describe dynamic stimuli involving the visual interaction of the user/player with a moving 3D object displayed on a computer screen. The reasons for this are that methods and software developed for eyetracking studies of static 2D stimuli are inappropriate for dynamic 3D stimuli, and manual analysis of dynamic 3D visual interaction is extremely time consuming. In order to address this, the authors have developed a software interface between the Tobii™ eyetracking system and the HiFi Game Engine for use in automated logging of dynamic 3D objects of gaze attention. This report describes the verification study performed to assess the performance of this integration between the eyetracker, logging tools and game engine. Detailed analysis shows effective results within the derived accuracy range, which is certainly sufficient for studies from a small scale to large scales necessary for extensive statistical analysis. The work presented in the report has been conducted in collaboration between FOI, Blekinge Institute of Technology and Gotland College.

Keywords
Eyetracking, game, gameplay, design, visual attention, instrumentation, methodology, human factors, behavioural science, cognition, measurement

Further bibliographic information

Language English

ISSN 1650-1942

Pages 54 p.

Price acc. to pricelist

Table of Contents

1 Introduction

At the "Serious Games Conference 06" in Washington DC, the content was divided into two major categories: War games and Simulations and `Situated Learning´ in Teaching and Training. The games under consideration provide these situated environments with their tasks and goals, which are provided either with instructions or not, and are of major interest for a broader target group than the entertainment industry.

Serious games are of interest for very specific effects and outcomes on their players in the form of various different kinds of learned skills. The ongoing improvement of the design and effectiveness of serious games requires an increasing understanding of the relationship between game design features, player characteristics, rewards of game play, game play interaction patterns and learning outcomes. Psychophysiological techniques provide a foundation for the detailed collection of data during real time play of games as an empirical foundation for this increasing understanding of play. Psychophysiology is a broad scientific field in which different types of sensors are used to measure the reaction and various states and behaviours of the human body to events in the world (or to experimental stimuli) and use the resulting data to make inferences about the mental state, emotional state and cognitive activity of the human player.

The psychophysiological perspective on game and simulation environments is of great importance in providing empirical foundations for the development of validated theories by which a designer (or an instructor) can make decisions in relation to gameplay design[1]. Of course there are different purposes of gameplay but the foremost is to engage the participant and create strong motivation to play. Very often people ask the question of how much time players can spend on one game and also the question of why they are so immersed in this kind of a learning environment. Gamers might freak out to hear that their games are learning environments, but if the problem solving involved in game play is considered, we can understand that this is the case. The difference between entertainment games and serious games is that entertainment games are designed to reward people for playing as an end in itself, with no consideration of the transferability of the skills involved in other contexts. For serious games, skill transfer is a critical issue and the whole point behind using games for non-entertainment purposes.

The biometrics obtained from psychophysiological studies of game play can be used to provide important knowledge for many reasons. From a design perspective it helps to answer why to use specific design features and arrangements. The purpose can be to control the player's physiological interaction to reach a certain interaction goal or to monitor the state of the player for training and instruction. Generally, the cognitive side of these interactions involves the development of specific skills allowing players to overcome game challenges that may include problem solving, tactics, spatial navigation, reaction time, logical reasoning and many other functions.

Today when situated learning is widely discussed, not just because of its effectiveness in learning but also because of economic efficiencies, the transferability of skills among virtual game environments, simulators and real life operative environments must be considered. To interact with a 2D display and a common keyboard, in relation to a simulator that emphasizes the real setup of devices with visual displays of different kinds and real world operation, can be questioned due to differences in sight/vision and motor interaction between games/simulators and operational environments. Game play biometrics can provide important data contributing to the answers to this question.

[1] Here we use 'gameplay' to designate play forms intended for a game design by its designers, and 'game play' to designate the actual play behaviour of players.

Brown[2] points out three important factors concerning biometrics. First, that game play analysis which is based on biometrics can drive new instructor created missions/scenarios. Second, that increased avatar linkage to actual player state is achievable using biometric measurement technologies as feedback to a game system. Thirdly, the real time dynamics of game play can be adapted; if the time requirements of game interaction are adaptable system variables, then workload involved in play can be either a fun factor or a more severe one depending on the context.

The project reported in the present study contributes to the more efficient use of the biometric technique of eyetracking for the analysis of game play. Previous work by Sennersten (2004) demonstrated the usefulness of eyetracking studies of game play, but encountered the problem of the very long analysis times required; four hours of game play took three months to analyse manually. If the identification of objects under the point of gaze could be automated, however, this basic aspect of analysis could be achieved far more quickly. While the automated identification of objects by image processing of the resulting screen data together with the eye gaze data would be complex, a much simpler solution is possible, i.e. integrating the eyetracking system with the game engine and having the game engine log the objects under the gaze point during play. The implementation of this is conceptually and technically simple, since an (x,y) coordinate representing a gaze point on the screen can be treated in the same way as an (x,y) cursor position when a mouse button is clicked, ray tracing from the (x,y) point to intercept the closest perceived virtual object within the 3-dimensional world synthesised by the game engine. It is then straightforward to modify the game engine code to ensure that the information about the object under the gaze point (i.e. its identity and position) is entered into a log file. In this way automated identification of game objects under the gaze point can be achieved and recorded in an ongoing log of a play session. Object identification under the gaze point then takes place automatically and in real time during play, taking no extra time at all than the play session itself.

This integration of an eyetracking system with a game engine, with real time logging of the object data, has been achieved by the project as reported in this technical report. The eyetracker used is the Tobii™ 1750 eyetracker owned by HGO, while the game engine is the HiFi Engine developed by the Swedish Defence Research Agency (FOI). The verification study reported in this document describes the system and the analysis process used to verify its correct operation. The results of the verification process are described and analysed, and the resulting precision model for the integrated system is presented.

[2] Notes from the presentation "Applying Biometrics to Assess Training In-Game and during AARs" by Randy Brown. (AAR stands for After Action Review) at Serious Games Summit Conference 2006, Washington DC, USA.

2 Background

Designing and interacting with 2D worlds has its own rules and possibilities. Moving on to digital 3D worlds adds many questions in how to design for a view frustum (or viewing volume) potentially having 360 degrees of freedom around three orthogonal axes. Virtual worlds can also be arbitrarily large, so how should we design for a player's attention within a space, which can be experienced as endless and open?

In order to understand how 3D worlds are constructed and perceived, one can refer to existing knowledge from the study of aviation and training simulators. Studies of human factors, situational awareness and workload have revealed a wealth of information on how to optimize and understand the interaction of humans and machines. This approach contrasts with the more prevalent approach to game design from a more general culture studies viewpoint, which has been dominant within international game research for the first few years of the millenium.

For example, in a helicopter there are also 360 degrees of freedom while rotating around the pilot's own shoulder or around their own axes. The instrumentation in a cockpit is both mechanical and digital. Visual Flying Rules (VFR)* are formalized and trained regulations for a pilot. One can say that the operational 3D environments in these cases are primarily situated outside the helicopter. They are external, outside both the physical human body and the physical body of the machine. The rules within the regulations are guidelines for how to meet this physical/external environment and have to be understood and solved by a human pilot. Of course, these environments, especially in aerial work, also include surprises of different kinds like terrain conditions, animals, unpredicted winds, changing light, etc.. Instrumentation Flying Rules (IFR), on the other hand, are those rules the physical human uses within the cockpit in conditions of poor or non-visibility of the external space outside the machine (e.g. in bad weather). Instrumentation within the helicopter can be considered to be read mostly as 2D graphics, as within a game (e.g. staminas, bars etc.), while the pilot's motor actions are carried out through a joystick, thrust control and pedals, that could also be used to play a game although it is more typical for the player of a PC game to use a keyboard and mouse. The theory in this field is strongly connected to visual attention theory, association theory and also theories of neuro-psychophysiological demands on the operator or player. Previc (1998) describes this in an integrated theoretical model of how different brain systems mediate our perceptual-motor interactions in *peripersonal* space (the region immediately surrounding our bodies) and the three major compartments of *extrapersonal* space (the focal, action, and ambient realms).

To operate in a digital 3D space one has to decide where one should be and how to navigate depending on what objects are present and relevant in the digital world. These decision processes are both present in a pilot situation as well for a player in a game play session. To connect this awareness of space, either in a physical reality or in a virtual game reality, cognitive functions try to map (associate) and categorize these things depending on the current task. The mind builds these understandings partly in the "Where" and "What" centers in the brain (Nyberg, 2002). Castor et al., 2003 (p. 112) state that the operating attributes in Military Aircraft Operations includes Flight and Mission Planning as a template for change. This means that the pilot has to plan the flight regarding distance. A central hypothesis of interest to the primary author of this report is that gameplay, depending on genre, is also carried out with decision-making based upon virtual distance, which could follow from real life conventions or vice versa.

It is against this background of the interelationships between aviation operations, situational awareness, workload and computer game play that psychophysiological investigations of game play are of interest,

---

* VFR and IFR are rules for how a pilot can and is allowed to act depending on daylight or darkness, weather conditions, the equipment of the aircraft, ground equipment and certification. Simply expressed, VFR rules indicate how a pilot, with his/her own eyes and the aircraft´s ground equipment, is allowed to perform with the aircraft within areas and aerodromes that have requirements on factors such as least sight distance and distance to clouds and obstacles (sight to ground). The cloud height cannot go under 600Ft and sight not under 800m.
For IFR, the decision height for "precision flight" is 200Ft (60m) or lower for category II or III. The sight circumstance along an IFR path way is allowed to be 0m.

particularly to the primary author of this report.

2.1 Reason for doing this project – the Research Project Context

As noted in the introduction, during a previos study on eye movements in an action game tutorial (Sennersten, 2004) the need for more efficient methods of data analysis became evident. Already at Lund University while working on a master's thesis in Cognitive Science, Sennersten turned to FOI wondering about the eyetracking methods used in previous studies. The focus on 2D environments in Lund did not meet Sennersten's needs in saying anything about player´s interaction patterns in 3D game worlds. The game worlds Sennersten was thinking of were particularly First Person Shooters (FPS) and Role Playing Games (RPG´s). What could help would be to be able to have an automatically obtained relation between x and y gaze coordinates and logging of both objects (identification and location) and events while playing to improve the efficiency of the analysis process, which otherwise is very time consuming. It is also a problem when making judgements by eye about the objects that are under gaze points identified by eyetracking, subjective interpretation is an unavoidable factor. "Clean" data is needed. The use of eyetracking to study game play is part of Sennersten's ongoing research, which will be greatly facilitated by automating lower level aspects of the analysis process.

The Swedish Defence Research Agency (FOI) has over recent years conducted several studies in game related areas (Hasewinkel & Lindoff, 2002; Kylesten & Söderberg, 2001; Lindoff & Hasewinkel, 2004; Rencrantz, 2003; Wikberg, Hasewinkel, Lindoff, Eriksson, Stjernberger, & Persson, 2003). These studies have mainly studied the usefulness of game consoles and games in training as a basis for choice of learning methods. Another study from FOI by Svensson et al (1997) has shown that the frequencies of shorter fixation times (head-up, HU) and frequencies of longer fixation times (head-down, HD) increased as a function of the avarage information load on a Tactical Situation Display of two JA37 aircraft simulators. One aspect that could be studied more thouroghly as a foundation for assessing simulation-based training is motor and visual interaction across computer-, simulator- and real life missions: can there be gaze and behavioural reinforcements in between these different environments and what differences do they create in humans when interacting? When bringing in game related environments and a generation with pre-learned individual gameplay patterns, one also has to consider pre-established game play behaviours that could be more or less efficient in different contexts.

The foremost reason for doing this project has been to develop the integrated eyetracker/game engine system and to see how accurate this kind of system can be for application within the research in the areas described above. In general, a lot of effort goes into biometric measurement, but our most important organ, in relation to perceiving the environment around us and generating the highest perceptual data input, is the eye, which to date has been underused in biometry. Without distributed visual attention it is impossible to say a great deal about the meaning of other kinds of physiological or psychological measurements in relation to responses to an environment in which visual data is significant. A peak in heart rate variability, for example, with no reference to audiovisual stimulus can be hard, if not impossible, to interpret.

2.2 Review of eyetracking technology

Eyetracking is a method that has been used commonly over the last 40 years and given insight into how (visual) attention is distributed when carrying out different tasks. The technique involves tracking eye movements and recording the distribution of gaze over time. Eyetracking techniques have been used in many fields including neurology, cognition, linguistics, branding, advertisement in newspapers and the like, car design, art, media communication, security of different kinds, etc..

There are four broad categories of eye movement measurement methodologies (Duchowski, 2003):

electro-oculography (EOG), scleral contact lens/search coil, photo-oculography (POG) or video-oculography (VOG), and video-based combined pupil and corneal reflection. Today the latter measurement is the most widely used because of ecological validity and ease of use.

EOG relies on measurement of the skin´s electric potential differences, using electrodes placed around the eye. Scleral contact lenses and search coils provide the most precise eye movement measurements. The search coil is embedded in the contact lens so the electromagnetic field frames can be placed directly onto the eye (Eye and Vision at Bernadotte Laboratory at Karolinska Institutet has these devices). POG measures the eyes under rotation/translation (the shape of the pupil) using corneal reflections of a close directed infrared light. The video-based technique with corneal and pupil reflection demands the head to be steady so the eye´s position in relation to the head and the point of regard coincide. This last technique has also been integrated into a variety of devices including the head-mounted eye tracker, which does not just track the screen but also gaze behaviour ouside the screen area.

It is important to distinguish between techniques for measuring eye movements and techniques for measuring eye-point-of-gaze (EPOG), also called "point of regard". Eye movements may be measured to determine where someone is looking. Only if the absolute orientation of the eye in relation to the surrounding objects is known can the EPOG be determined. Eye movements could, however, be measured for other purposes than determining the EPOG. Often it is not only interesting to know *where* a person is looking, but also *how* the person is looking. For instance, can eye-movements tell us something about the mental state of an operator, such as the mental workload? For this purpose it is not only interesting to measure the orientation of the eye, but also measures such as blinks and pupil diameter.

A user uses both foveal vision and peripheral vision to gather visual information. With foveal vision, which is about two degrees of visual angle, the user can get a detailed look at a certain point, the eye point of gaze. The movements of the eye that maintain fixations are called fixational eye movements. To make the image of a constant moving object stay in the fovea, pursuit movements are made by the eye. Movements of the eye that are jumping from one point to another are called saccadic movements. Vistibuloocular eye movements are those compensating for the movements of the head. Optokinetic eye movements are used when a large part of the visual field is rotating. Finally, vergence or disjunctive eye movements are used, for example, to follow a target that is getting closer or further away, by moving the eyes in opposite directions to one another.

The time that the eye is not moving, but staying in the same position, is called fixation time or dwell time. Fixation times are used to evaluate computer displays or other interfaces. However, it is not always obvious how to interpret the meaning of a long fixation time. For example, important information can make the user fixate upon the information for a long time, as well as badly displayed and confusing information. By comparing the fixation times to other data it can sometimes be possible to draw valuable conclusions from the data.

In active vision or overt attention, sampling is achieved by a fixation-move-fixation rythm and this pattern can be found in the vision of humans, most other verbrates and some invertebrates (Land, 1995; Land and Nilsson, 2002). Passive vision is also referred to as peripheral vision or covert attention, with the meaning of attending without looking. Peripheral vision increases in degree outside the 2-degree high-resolution foveal centre. Findlay and Gilchrist (2003) point out that overt attention, via the fovea, plays the major role in visual attentional selectivity.

It has long been believed that the eyes and eye-movements can tell us some things about the human mind; when looking at a stranger for the first time, it is possible to see whether he or she is frightened, sad, tired etcetera simply by looking at the persons eyes. In recent years, eyetracking technology has been developed to provide competent aids in determining user interaction with technology. Computer interfaces have been evaluated with eye movement data in several studies. Graf and Krueger (1989) used eye movement data as a performance measure for evaluation of an alphanumeric display, as well as a measure of the cognitive load on the user. Lankford, Shannon, Beling, McLaughlin, Israelski, Ellis, and Hutchinson (1997) have used eye gaze tracking in the design of telecommunications software, arguing that "...graphical interfaces can benefit substantially if eye tracking data on a user's visual interaction with the software is considered as a part of the design process. This eye tracking data would supplement the more traditional human performance data (e.g. timing and errors gathered during usability testing)". The combination of eye movement data with other data sources, to get a more complete picture of what is happening, is a promising area deserving much more investigation.

Eye movement measures can be used both in the design of a new system and in the later evaluation of the same system. Thus, the purpose of using eye movement data to evaluate interfaces could be a concern both for the better design of new interfaces and for ensuring that interface designs are really effective by later testing.

2.2.1 Comparison of specific eyetracking systems

Different eyetrackers have different qualities suited to different purposes. The sampling rate, for example, can be a major issue when choosing equipment for a particular study. A relative assessment of various systems was made by researchers at FOI to get an overview of their pros and cons in relation to purpose and goals for different studies. The table below is derived from the technical report presenting the results of this assessment (Alfredson, Nählinder & Castor, 2004), where five different systems/techniques, all of which FOI has experience with using, are compared. In this report the properties of the Tobii eyetracker used in this study are added to the table.

	GazeTracker	EOG	JAZZ	Smart Eye	Video based	Tobii 1750
Measurement principle	Differential corneal reflection and electromagnetic head-tracking	Corneo-retinal electrical potential	IR-pupil tracking	Optical feature recognition	Optical video	Optical feature recognition
Maturity	High	High	Low	Medium	High	Medium
Accuracy	High	NA	NA	High	Low-medium	High
Max range horizontal	Medium. Limited by eye movements	Medium	Medium	High. Limited by head movements	Unlimited for head	High. Limited by head movements
Max range vertical	Medium. Limited by eye movements	Medium	Medium	Medium. Limited by head movements	Unlimited for head	Medium. Limited by head movements
Head movements	Yes	No	Yes	Yes	Yes	Yes
EPOG	Yes	No	No	Yes	Yes	Yes
Time resolution	60 Hz	Optional	1000 Hz	30-60 Hz	Low	10-50 Hz
Fixation duration	Yes	Yes	Yes	Yes	No	Yes
Scan paths	Yes	No	No	Yes	Yes, but very time consuming	Yes
Saccade velocity	Yes	No	Yes	No, not implemented	No	No
Eye blink detection	Yes	Yes	Yes	Yes	Yes	Yes
Pupil diameter	Yes	No	No	No, not implemented	No	Yes
Eye blink	Yes, optional	Yes, optional	No, not implemented	Yes	Yes	No
Eye activity (EME)	Yes	Yes	Yes	Yes	Yes	Yes
Costs	High	Medium	Medium	High	Low, but high analysis costs	Medium

Reliability	Medium	Medium	Medium	Medium	Medium	Medium
Intrusiveness	High	Medium	Medium	Low	Low	Low
Type of data	Eye movement and EPOG	Eye movement	Eye movement (Saccades)	Eye movement and EPOG	EPOG	Eye movement and EPOG
Interference from surroundings	High (Electro-magnetically)	Low (Muscle activity)	Medium (Infra red illumination)	Medium (Illumination and picture background)	Low (Illumination)	Medium (Illumination and picture background)
Ease of use	Low	Medium	High	Medium	High	Medium
Preparation time/calibration time	High	Low	Low	Low. A couple of hour's installation if moved to new setting. Low calibration time.	Low	Low
Analysis time/cost	Medium	Medium	Medium	Medium	High (Proportional to amount of data)	Medium

The "High", "Low" and "Medium" values in the table do not have any scientific thresholds; they express qualitative judgements based upon the experience of the FOI researchers and authors of this paper, and could possibly be ranked differently by another group. The maturity of experience in different studies nevertheless gives the table a high trust value as overview information.

3 Experimental platform specification

The project that this document describes uses the Tobii eyetracking system since it has EPOG, blink detection, is robust against head movement, tracks both eyes for higher reliability, can be calibrated quickly, has comparatively good spatial accuracy, an adequate sampling rate, and an API for the server that supports convenient integration with the game engine.

The HiFi Engine was selected since the source code is accessible and FOI own the engine and are therefore not constrained in its use. Features and performance of the engine are very good and it is an engine for first-person shooter games, which are of interest to the researchers involved. Since FOI developed the engine, familiarity with the engine also makes it very easy to modify, thus facilitating software integration with the eyetracker.

3.1 Technical architecture HiFi Engine

The platform used in the experiment is based on the HiFi Engine (HFE), a game/simulation engine developed at Man-System Interaction - Swedish Defense Research Agency (FOI) in Linköping, Sweden.

3.1.1 Open source components

HFE is developed around several Open Source components that take care of specific tasks such as physics, sound, network communication, etc.. Besides wrapping in several open source components, there is a core of functionality that glues everything together and adds some specific functionality to the simulation engine. This is functionality that is specific for the simulation area that HFE is mostly used for at FOI.

HFE itself is developed as a simulation component that could be used for different applications. It could be used for pure simulation purposes like simulating a specific vehicle, or for more game-related simulations such as the context of this report, or in a completely different type of application such as a command and control application. In the latter case HFE can take care of everything from visualization to symbol handling, including move, add and remove symbols, send symbols to other systems, logging etc. (e.g. see the report for the project “Cognitive overview” where CoMap, Cognitive Map is described, which is a C&C application for urban combat, Kylesten, 2004).

With HFE it is possible to run one stand-alone client or to create a bigger simulation with many clients. This could be very useful when looking at, e.g., effects of solving situations in a cooperative manner. HFE can also enhance simulations with its excellent visualization and in-depth vehicle physics simulations etc.[3]

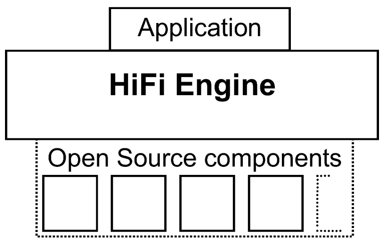


Figure 1. The HFE architectural hierarchy. An application is built using HFE, which in turn uses open source components that handle some functionality in the simulation engine.

[3] For further information about HiFi Engine, contact Johan Hedström, johhed@foi.se.

3.1.2 The Test Application

A basic FPS (First Person Shooter) game has been built for use as the verification stimulus game for the present study that utilizes some of the advantages that HFE provides. Some of the things implemented in the FPS are basic weapon handling, picking up ammo boxes and similar triggers to pick up health-increasing objects.

3.1.3 Logging

During execution of the experiment, interesting perceptual objects within the game world, as indicated by player gaze direction, are logged in a file. The logging data consists of structures with one entry per gaze sample, where a gaze sample is taken from the eyetracker server with every frame generated by the game engine, corresponding also to a simulation time tick. The data logged includes:

- object identifier, consisting of name and (x,y) position of the object model origin upon the virtual ground
- time of the gaze sample

3.2 TET Server and TOBII eyetracker

The Tobii Eye Tracker 1750 runs on a PC with the Windows Operating system (Windows 2000 service pack 3 or Windows XP service pack 1 or 2). The system can either be used in single or double computer setups. In this study a game application is running at the same time as eyetracking, which is most likely too much processing to do on one computer, hence in this study we use the double computer and double screen setup. One computer is running the eyetracking software ClearView (2.6.3) and the TET eyetracking server. The other computer is running the game application. The two are connected via TCP/IP. The TET server gets the data from the eyetracker and sends it to the ClearView software. For calibration of the eyes of the player, the operator switches the eyetracker display over to the second screen of the operating computer that runs the Clearview software, for display of the calibration pattern (hence a dual head graphics card is needed on the operating computer for the double display setup). The output from the Game Computer goes through a splitter where two video outputs are created, one going to the video capture card on the operator computer, where the captured video data is input to ClearView, while the other output goes to a switch that selects either the game display or the operator computer second monitor display to send to the eyetracker display. The routing of the video display is illustrated in Figure 2.

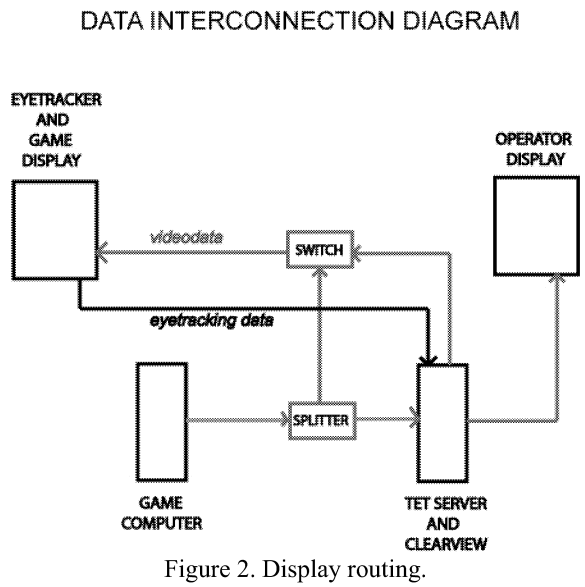


Figure 2. Display routing.

A typical play session using this system involves the game player playing the game in front of the eyetracker. After briefing the player with any instructions relevant to the study, the calibration process is run, calibration patterns being displayed on the eyetracker screen while Clearview uses the known positions of elements in the display to calibrate eyetracking for that specific player. Then the eyetracker screen is switched over to the game display via the routing switch. The operator must start game object logging within the game engine using a command line instruction, at which a synchronisation message is sent by the game engine via the TET server to Clearview to commence eyetracking. Synchronisation is necessary to simplify time correlation of object log entries on the game computer with video frames captured by the operating computer. At an instruction from the operator, the player then commences playing while eyetracking and game engine object logging are running.

During game play, with each simulation tick or game display frame the HiFi (game) engine requests a gaze data sample from the Clearview software via the TET server. The game computer then traces a ray from the (x,y) coordinates of the gaze point to intercept the first collision mesh within the game world under that (x,y) point, corresponding with the object being looked upon by the player. Within the game engine, a log entry is generated that includes the name of the 3D object model appended with the (x,y) coordinates of the model instance on the horizontal plane of the 3D game world, ensuring that each object instance is uniquely identified, together with a time stamp generated by a trigger signal from the eyetracker collected with a high-resolution timer, based on Windows multimedia timer. The log record is then appended by the game engine to the object log file stored on the gaming computer. When the goals of the session have been met, the operator asks the player to stop playing. A command line instruction on the gaming computer is then used to turn off the object log and eyetracking.

The Tobii eyetracking system detects and collects eye gaze data at a rate of 50 samples per second, with frame grabbing of the corresponding game display at a lower but selectable frame rate. The eyetracking system monitors the eyes of the subject/participant based upon reflected infrared light patterns and calculates gaze positions automatically, with hardware in combination with advanced software algorithms. The system has head motion compensation and low drift effects, with binocular tracking, which yields higher accuracy then just a monocular method. Another advantage of this system is that if the participant happens to move, especially the head, tracking is not lost if only one eye cannot be tracked during the movement. (With other systems, the participants may have helmets, headrests or markers, with different respective freedoms and constraints of movement.) For ecological validity, the eyetracking camera is implemented at the bottom part of a normal sized screen. The Tobii 1750 system can only track what is presented on the screen and not anything outside the screen area. The calibration time is only 3 minutes if no problems occur due to eye disorders, blinking, etc.. The recommended eye distance to screen is approximately 60 cm. The accuracy of an eyetracker is measured in degrees. The *accuracy* or *bias error* of the Tobii 1750 system has been tested over a set of individuals to 0,5 degrees of visual angle when using standard accuracy measurement principles for eye trackers. One degree of accuracy corresponds to an average error of about 1 cm between the measured and intended gaze point at 60 cm distance from the user to the screen. The tolerance of head movements is about 10 cm sec^{-1}.

The light condition requirements for eyetracking are generally the same as when using an ordinary computer display. Near Infrared light (NIR) such as components of sunlight are not recommended to be present while tracking the eyes, since this can cause disturbances in the tracking. The eyes and especially the pupils are very light sensitive, so it is also best to avoid direct and/or strong light from the study environment.

For both eyes (left and right) the following data is available derived from the eye tracker data processed by the Clearview software and TET Server:

- Time (The timestamp is in microseconds),
- Screen X (horizontal position),
- Screen Y (vertical position),
- Cam X (horizontal location of the pupil in the camera image),
- Cam Y (vertical location of the pupil in the camera image),
- Distance (distance from camera to eye),
- Pupil (Size of pupil in mm) and
- Code (validity of gaze data).

Validity codes ranging from 0-4 and are logged for each eye with every gaze data point. 0 is the highest validity value and means that all relevant data for the particular eye is correct.

The frame rate and timing are crucial when this system is connected with a game engine. The game engine has its own tick-by-tick simulation update rate, the eye tracker has a different sampling rate and the real time video capture has another time rate again. The sampling rate of the 1750 eyetracker, representing the rate of gaze point data collection by the system, is a constant rate of 50Hz. The display frame update rate of the game engine varies depending upon the complexity of the visual scene, generally being between 25-35 ms. Video data capture was set at 10 fps (frames per second) for the present study. These different rates must be correlated during analysis in order to obtain the most accurate correspondences between gaze data, game frames and video frames, although due to variable latencies in the system the correlations are never perfect. This is analysed in detail in later sections of this document. Latency is defined as the time taken from the generation of a signal to its reception, in this case including time taken for eyetracking camera exposure, transfer to the TET server, calculation and delays in the server, transfer to Clearview, request generation from the game engine, processing time in the TET server, and response propagation back to the game engine.

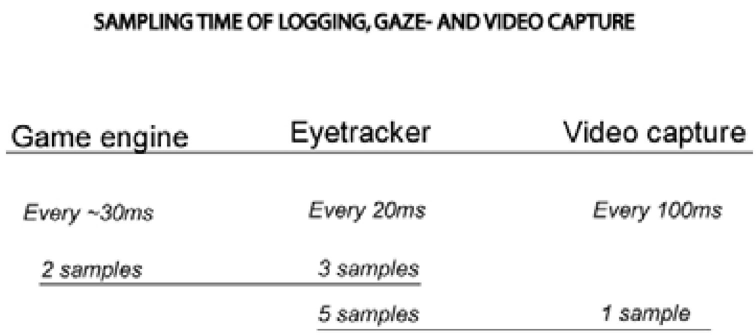


Figure 3. The start time from the Game Engine has to be synchronized with the Real Time Video Capture due its delay. The figures above show the internal relationships in this study.

4 Description of the Verification Procedure

The verification study reported in this document concerns verification of the effectiveness and accuracy of the integration of the Tobii 1750 eyetracker with the HiFi game engine (HFE). This concerns the basic success of the software methods used to perform the integration as well as the effects of various potential sources of error in both spatial accuracy and temporal accuracy. The essential method of the verification procedure is to conduct a study using the integrated system, then to conduct an analysis of the objects under the gaze point by manual interpretation of the captured game video data with superimposed gaze data, and then to compare the results of this visual analysis with the object log produced by HFE. Detailed analysis of the comparative data is then used to obtain an accuracy model of the integrated system that provides the basis for interpreting the accuracy of automated log data obtained from ongoing studies based upon the integrated system. The accuracy model is the significant analystical result of this project.

Note that the first verification procedure showed that the integrated system did not work due to faults that were later found in the implementation of ray tracing within HFE. These faults were quickly fixed and the verification procedure was repeated to undertake the analysis presented here.

Various software packages are used for the different stages of the process. They are:

BattleCraft™	- Stimulus
ClearView™	- Videocapture and Video output generation with superimposed gaze data
Virtualdub™	- Analysis/Transcription
Excel™	- Statistics/Transcription

The rest of this section presents an overview of the steps of the verification procedure. These steps are then exemplified in the subsequent section, followed by a section containing a detailed discussion of issues observed within the verification process and data. The final section then presents the accuracy analysis and model based upon the data and the verification process.

4.1 Steps in the verification process

A. Prepare Stimulus

The 3D stimulus is created using a commercial, off the shelf game level editor. Most popular 3D games (e.g. HalfLife, Oblivion) have these editing tools to allow players to create and play their own game levels. ‘Mods’ are also possible, allowing players to create their own game content (media assets), worlds and simulations. Integrating a game engine with an eyetracker as described in this study, however, requires access to the source code of the game engine. Source code is typically not freely available unless it is an open source game engine. For the project reported here, FOI has access to the HFE code and HFE can import game levels from a commercial level editor, making the editor a convenient tool for stimulus development.

B. Select Player Participant(s)

The criteria for selecting players depends upon the study being conducted, the hypotheses of interest, etc.. It may be important to consider player characteristics, since variability of play style preferences, tastes, etc. can have a major impact on study results. One early example of a system for categorizing players is that developed by Bartle (1996). Bartle´s four main player categories are: Achievers, Explorers, Socialisers and Killers, each reflecting differences of in-game behaviour. Also there are numerous psychological tests that one can consider for categorising players to find character and personality differences that may correlate with statistical differences in game play patterns. The latter tests may have to be carried out by authorized and certified psychologists in some countries if they need to be approved by ethical authorities. Different demographic categories of players could also make a difference to their gaze behaviour during game play, including: Age, Gender, Novice/Average/Skilled player, How long the participant has been playing games, What games the participant has been playing, Certain tactics or strategies habitually used by the player, etc. A standardized eye test should preferably also being carried on participants since eye conditions can influence the results in eyetracking.

In a military context, when there are often homogenous trained groups, the emphasis of interaction testing can be on to see how a particular task is carried out or on the usefulness of specific equipment. In a non-homogenous group the categorisation of player types is more important, so a baseline can be created. A baseline is a so-called neutral interaction pattern, which can be related to inividual differentiations among players.

Despite these concerns with player characterization in experimental studies, the verification procedure is not concerned with the analysis of play patterns as such. Hence no questionnaires or observations that would be used in a full-scale experiment were used in this verification study. Similarly, biometric measures can say very much about temporal states within a player, one purpose with eyetracking being to relate biometric responses to what is represented on the display, similarly to traditional stimuli-response-result learning (e.g. Pavlov´s saliva test, *conditioned reflexes*, Eysenck, 2000). However, in this verification study only eyetracking data is considered.

C. Log Data

The sequence of operations for undertaking data logging is as follows:

i) Operator: Start Clearview software

ii) Operator: Start game using HiFi engine.

iii) Operator: Set player start point within game level (HiFi command line instruction).

iv) Operator: Spawn player character within game level (HiFi command line instruction).

v) Operator: Initiate logging by game engine (HiFi then starts eyetracking via message to TET server; HiFi command line instruction)

vi) Player: Start playing, play to objective.

vii) Operator: When player reaches objective, end logging and eytracking via game engine (HiFi command instruction)

The resulting logs and data include:

- Object log from the game engine
- Clearview eyetracking data
- Captured video of the play session with superimposed gaze data, also generated by Clearview

Figure 4 shows an example of data from the object log created by the HiFi engine. For the verification study, the HiFi engine logged in total 13285 rows in a 5 minutes play session.

179	Name	berlin_galler_m1_820_905
180	Name	berlin_galler_m1_820_905
181	Name	berlin_galler_m1_820_907
182	Name	berlin_galler_m1_820_907
183	Name	berlin_galler_m1_820_907
184	Name	berlin_galler_m1_820_907
185	Name	berlin_galler_m1_820_907
186	Name	berlin_galler_m1_819_908
187	Name	berlin_galler_m1_819_908
188	Name	suburbhouse_2_closed_m1_800_943
189	Name	suburbhouse_2_closed_m1_800_943
190	Name	suburbhouse_2_closed_m1_800_943
191	Name	suburbhouse_2_closed_m1_800_943
192	Name	suburbhouse_2_closed_m1_800_943
193	Name	suburbhouse_2_closed_m1_800_943
194	Name	suburbhouse_2_closed_m1_800_943
195	Name	germansoldier_808_935
196	Name	germansoldier_808_935
197	Name	germansoldier_808_935
198	Name	suburbhouse_2_closed_m1_800_943
199	Name	germansoldier_808_935
200	Name	suburbhouse_2_closed_m1_800_943
201	Name	suburbhouse_2_closed_m1_800_943

Figure 4. Object log by HiFi Engine.

D. Create Transcription

A transcription here means a textual description of what objects are judged by the analyst as being visually fixated in the captured video of the play session. This requires the analyst going through the video frame by frame and making a transcript record for each frame. The video is displayed using the VirtualDub™ software with a screenshot shown in figure 5, which allows each frame to be accurately examined.

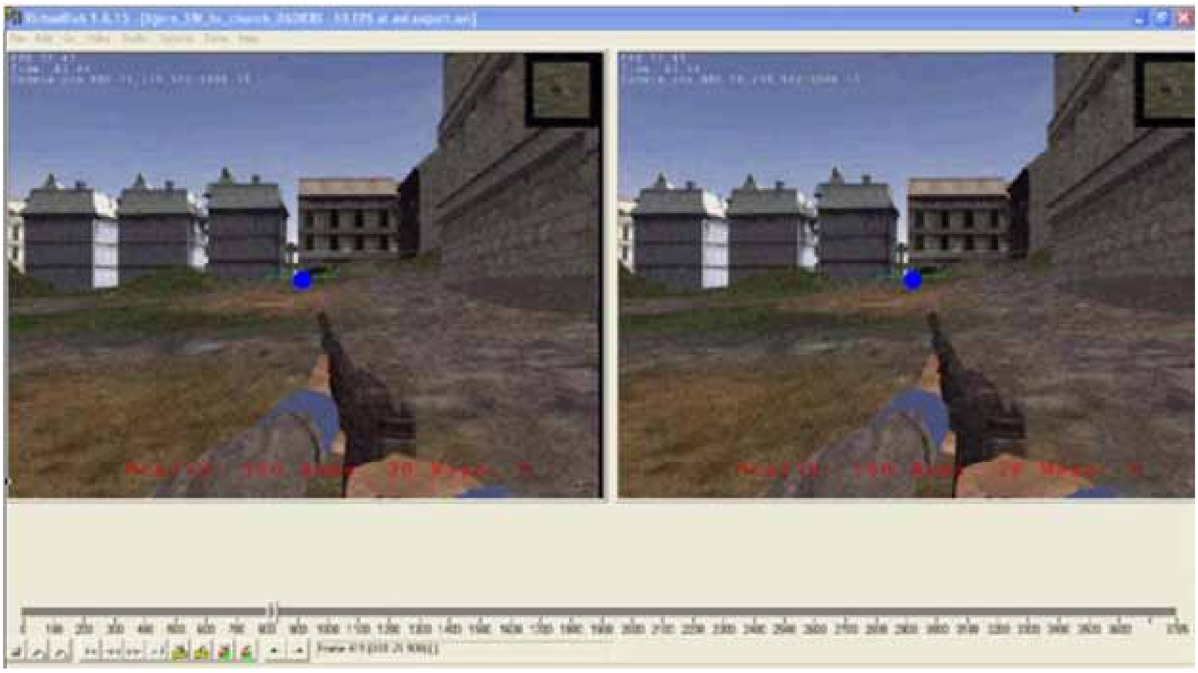


Figure 5. The software VirtualDub™ version 1.6.15 used for manual transcription.

A transcript record includes:

- Time of video frame
- Object name
- Gaze behaviour (e.g. new fixation, extended fixation)
- Object(s) interpreted by the analyst as being under the gaze point

These records are entered into a spreadsheet to facilitate further analysis. Figure 6 shows an example of the information from the avi file transcribed into the spreadsheet (note that data in column H is not used for the analysis).

E	F	G	H	I
1,000	terrainobject	new fixation	2,52	
1,100	terrainobject	fixation grows	2,62	The dot is at SW
1,200	terrainobject	fixation grows	2,72	The dot is at SW
1,300	terrainobject	fixation grows	2,88	The dot is at SW
1,400	terrainobject	fixation grows	2,92	The dot is at SW
1,500	TR_Ammo_829_948	fixation grows	3,07	The dot covers the
1,600	TR_Ammo_829_948	fixation grows	3,12	The dot is covering
1,700	TR_Ammo_829_948	fixation grows	3,23	The dot covers the
1,800	suburbhouse_2_closed_m1_839_961	shift fixation grov	3,31	The dot is mainly
1,900	suburbhouse_2_closed_m1_839_977	new fixation	3,43	
2,000	suburbhouse_2_closed_m1_839_977+suburbhous	new fixation	3,52	Hard to determine
2,100	suburbhouse_2_closed_m1_839_977+suburbhous	fixation grows	3,57	
2,200	suburbhouse_2_closed_m1_839_977+suburbhous	fixation grows	3,68	
2,300	suburbhouse_2_closed_m1_839_977 +ground	fixation grows	3,83	
2,400	suburbhouse_2_closed_m1_839_977 +ground	new fixation	3,93	
2,500	suburbhouse_2_closed_m1_839_977 +ground	fixation grows	4,03	
2,600	suburbhouse_2_closed_m1_839_977 +ground+an	fixation grows	4,12	Hard to determine
2,700	TR_Ammo_829_948+suburbhouse_2_closed_m1_	fixation grows	4,22	Aiming cross
2,800	TR_Ammo_829_948+suburbhouse_2_closed_m1_	fixation grows	4,32	Aiming cross
2,900	TR_Ammo_829_948+suburbhouse_2_closed_m1_	fixation grows	4,42	Aiming cross
3,000	TR_Ammo_829_948+suburbhouse_2_closed_m1_	fixation grows	4,52	Aiming cross
3,100	TR_Ammo_829_948+suburbhouse_2_closed_m1_	fixation grows	4,62	Aiming cross
3,200	TR_Ammo_829_948+suburbhouse_2_closed_m1_	fixation grows	4,72	Aiming cross

Figure 6. An examplification of transcription in the spreadsheet, see column "F".

Object identifiers are first annotated by the analyst on a printed top-down view of the game level obtained from the level editor or from within the editor on a second screen to help identify objects on the video and to work out the camera orientation within the game world (Figure 7). The identifiers have to be marked out manually because there is no automation of this.

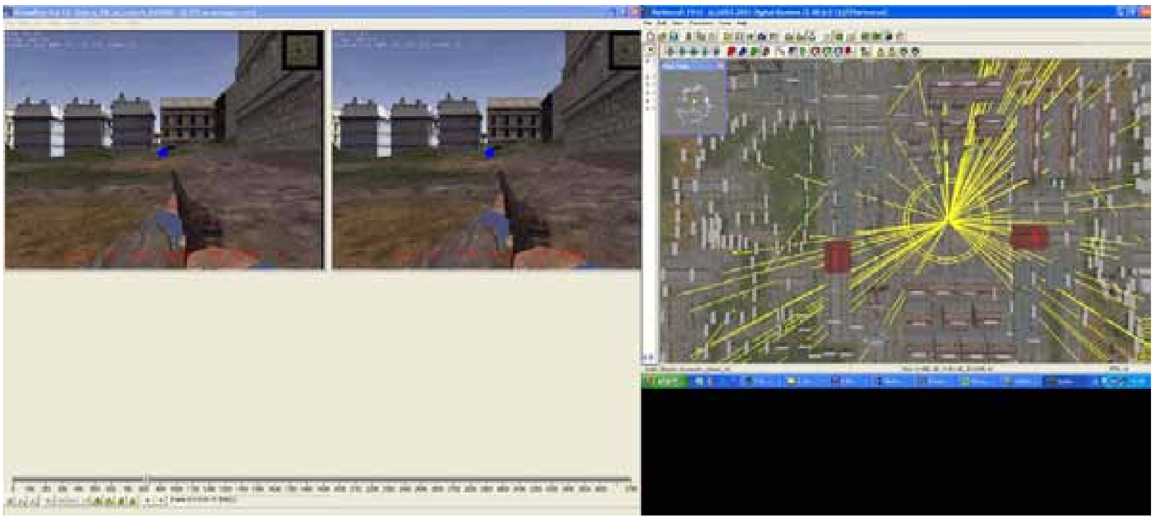


Figure 7. A "print screen" of dual screen mode while transcribing from VirtualDub. The first-person view to the left and the top-down-view to the right. Source: VirtualDub and Editor "BattleCraft".

E. Synchronise record times from the Game Engine Object Log and Video Transcription
The objects in the object log created by the game engine need to be correlated with the corresponding records in the transcribed information from the avi file. The procedure used to do this is:

- Insert the object log entries in the same spreadsheet as the video transcription

- Search through the video to find frames representing distinctive new objects of fixation

- Find the corresponding log entry changes in the object log data

- For those correspondences that are found, look at the time difference for each of those records between the object log and the video transcript

- Average those time differences to obtain a best estimate of the starting time differences between the object logging and the video data capture

- Create a new (hypothetically) synchronized object log time column by modifying the frame times in the original log data by the offset time value found in the previous step

For the verification procedure reported here, 7 match points were selected for obtaining an average value as a basis for time synchronisation. The 7 locations that were picked for the study are objects that the player participant has not looked at before until that point in the game session. The 7 different match points used, with the corresponding time differences, are:

1st match
AviFrame:4,000 wdfence2_fen1_m1_820_896
GameEngine_Start: 5.515037
Time difference: 1,515037

2nd match
AviFrame:8,100 frwall 784_947 / hangar1_m1-hangar1interior_0_0
GameEngine_Start: 9.674996
Time difference: 1,574996

3rd match
AviFrame:172,500 euwindmill_m1-euwindmillinterior_0_0
GameEngine_Start: 174,02664
Time difference: 1,526642

4th match
AviFrame:234,900 landmineprojectile_992_1227
GameEngine_Start: 236,446518
Time difference: 1,546518

5th match
AviFrame:248,200 citymesh2_m1-eubooksh_m1_2_6
GameEngine_Start: 248,395844
Time difference: 0,195844

6th match
AviFrame:284,400 eu_church_m1-eu_churchinterior_0_0
GameEngine_Start: 286,175049

Time difference: 1,775049

7th match

AviFrame:360,700 citymesh3_m1_1078_1175

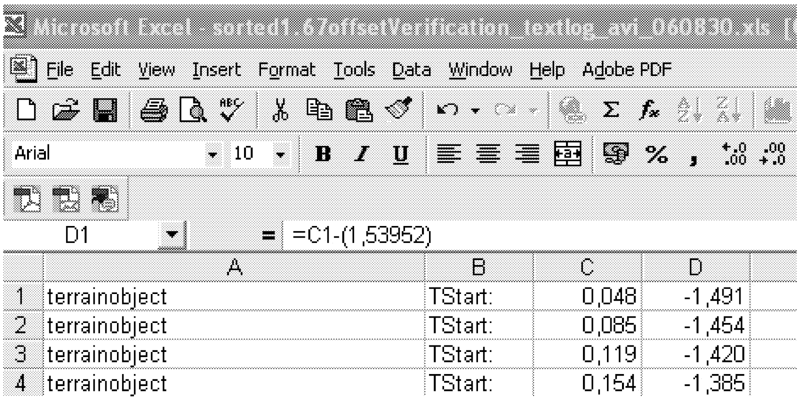
GameEngine_Start: 362,234406

Time difference: 1,534406

The average value used for synchronization is taken from the 1st-4th match values and the 7th value. The two other values are taken away because the records are far enough away in time from the other values to be regarded as aberrant. The average time difference works out to be 1,539520 s. This positive number represents the time period for which the object log was active before video capture commenced. Hence it must be *subtracted* from object log times in order to synchronise the object log entries with the video transcipt entries.

Once this average is obtained, a new column is created in the spreadsheet with values representing the original object log times modified by this value as the offset representing the difference in start time between object logging and video capture. The new object log values represent the best estimate of times since the same starting time as the video capture.

This is illustrated in Figure 8, where column C is the original object log time and column D is the start time synchronised with video capture. Note that the negative values in column D indicate that object logging commenced prior to video capture, indicating a delay from when the start of eyetracking message is issued by the game engine to actual start of eyetracking and video capture by the Tobii system.



	A	B	C	D
1	terrainobject	TStart:	0,048	-1,491
2	terrainobject	TStart:	0,085	-1,454
3	terrainobject	TStart:	0,119	-1,420
4	terrainobject	TStart:	0,154	-1,385

Figure 8. Column C is original time and column D is new time concerning start delay for the video capture.

F. Run Sorting Macro

The spreadsheet resulting from step 5 above includes the object log data alongside the video transcript data, including the offset time values for the object log data. For the sake of comparing the object log data with the video transcript data, it is then very convenient to sort these two data sets into a single time ordered list. This can be done automatically by running a macro written for this purpose in the spreadsheet. An example of the result is shown in Figure 9 below. Since the object log samples are taken every 25-30 ms and the video frame rate gives a frame interval of 100 ms, there are typically three object log entries between each video transcript entry. Sorting in the verification study resulted in a total of 17064 rows of object log and transcript information. Since this is the sum of the number of object log rows and video transcript rows (3780 + 13285), it may be inferred that the calculated times never exactly match (i.e. there is no single row, corresponding to a time from start, at which there is both an object log entry and a video transcript entry).

187	TR_Ammo_829_948	TStart:	5,157	3,617	
188	suburbhouse_2_closed_m1_839_977	TStart:	5,192	3,653	
189	terrainobject	TStart:	5,227	3,687	
190					3,700 terrainobject
191	terrainobject	TStart:	5,260	3,720	
192	terrainobject	TStart:	5,294	3,754	
193	terrainobject	TStart:	5,327	3,787	
194					3,800 terrainobject
195	terrainobject	TStart:	5,357	3,817	
196	terrainobject	TStart:	5,390	3,851	
197	terrainobject	TStart:	5,422	3,882	
198					3,900 berlin_galler_m1_820_905
199	terrainobject	TStart:	5,450	3,910	
200	berlin_galler_m1_820_905	TStart:	5,481	3,941	
201	wdfence2_fen1_m1_820_896	TStart:	5,515	3,976	
202					4,000 wdfence2_fen1_m1_820_896
203	wdfence2_fen1_m1_820_896	TStart:	5,543	4,004	
204	[illegible]	[illegible]	[illegible]	[illegible]	

Figure 9. The sorted Object Log and the AVI frame transcription.

G. Detailed analysis of sorted data

The sorted list is inspected to identify three categories of correlation information of the object log with the video transcript that describe the accuracy of the system. Classifications in these categories are entered into three new columns added to the spreadsheet, as shown in Figure 10. The categories are:

- the number of ambiguous objects under the representation of the eye point of gaze observed in the video transcript (column E on Figure 10). The figures derived from this are used to obtain a probabilistic model of spatial accuracy in the accuracy model developed later in this document.

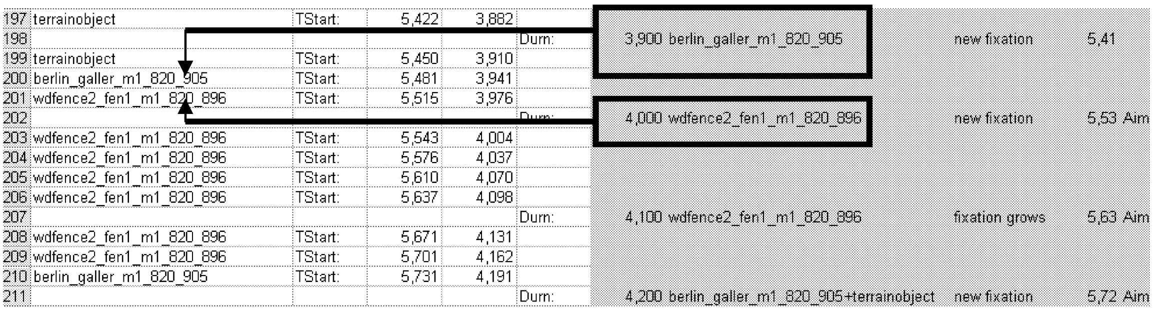
- 'number of records apart in time' of objects that correlate (column F on Figure 10). Since the object logging time is variable, object log and video transcript records do not occur at exactly correlating times. The analyst therefore records the number of records separating log entries and the apparently corresponding video transcript entries. A value of '0' means that the object is logged both before and after the video transcript record in which it occurs. '-1' means that the object log entry occurs one record later than the transcript entry, '+1' one record earlier than the transcript entry, etc.. This illustrated on Figure 11. The figures derived from this analysis are used to obtain a probabilistic model of timing accuracy in the accuracy model developed later in this document.

- a description of specific circumstances of correlation or reasons for error, for example that a logged object lies under a saccade (valid) or that an object was logged when the raytracing in the game engine intercepted its bounding volume when the bounding volume extended beyond the visual edge of the object (so the actual gaze object was behind this transparent area covered by the bounding volume) (column F on Figure 10).

I62 =

	A	B	C	D	E	F	G	H
48					poss	n rows	comment	0,100 terrainobject
49	terrainobject	TStart:	1,668	0,129	objects			
50	terrainobject	TStart:	1,703	0,163				
51	terrainobject	TStart:	1,734	0,194				

Figure 10. The three implemented categories in the spread sheet.



197	terrainobject	TStart:	5,422	3,882				
198					Durn:	3,900 berlin_galler_m1_820_905	new fixation	5,41
199	terrainobject	TStart:	5,450	3,910				
200	berlin_galler_m1_820_905	TStart:	5,481	3,941				
201	wdfence2_fen1_m1_820_896	TStart:	5,515	3,976				
202					Durn:	4,000 wdfence2_fen1_m1_820_896	new fixation	5,53 Aim
203	wdfence2_fen1_m1_820_896	TStart:	5,543	4,004				
204	wdfence2_fen1_m1_820_896	TStart:	5,576	4,037				
205	wdfence2_fen1_m1_820_896	TStart:	5,610	4,070				
206	wdfence2_fen1_m1_820_896	TStart:	5,637	4,098				
207					Durn:	4,100 wdfence2_fen1_m1_820_896	fixation grows	5,63 Aim
208	wdfence2_fen1_m1_820_896	TStart:	5,671	4,131				
209	wdfence2_fen1_m1_820_896	TStart:	5,701	4,162				
210	berlin_galler_m1_820_905	TStart:	5,731	4,191				
211					Durn:	4,200 berlin_galler_m1_820_905+terrainobject	new fixation	5,72 Aim

Figure11. Here the berlin_galler_m1_820_905 is matching 2 rows down so that represents "-2" in the inserted column. If its instead 1 or 2 rows up it is represented with a "+" before digit. If one object is embedded between two identical objects its represented with "0".

H. Derive Accuracy Model

The first two categories under point 7 above provide the basis for statistical characterization of the temporal and spatial accuracy of the system. The accuracy model is then the derived accuracy for the Tobii eyetracking system integrated with the HiFi game engine. This is described in detail below, both in terms of frequencies and an abstracted probabilistic model (in the section 'Analysis: Resulting Accuracy/Precision Model').

4.2 Summary of Data Quantities in the Verification Procedure

The verification study documented here covers logged data from a player in the `SW´, `NW´ and minor parts of the `NE´ zones of the 3D environment stimulus described below. The generated text data from the Game Engine is logged every 25-30 ms. The real-time captured AVI (Audio Video Interleaved) video file including superimposed gaze data has a frame rate of 10 fps corresponding to an inter-frame time of 100 ms. So 3-4 rows of game engine log data correlates on average with one row of AVI information in this case. The total object log entries generated by the game engine for the session used for the verification study is 13 285 rows, while the AVI file includes 3780 video frames resulting in 3780 rows of transcribed AVI information.

4.3 Detailed Description of the Stimulus Used in the Verification Procedure

The first author of this report developed the stimulus (i.e. the game world) using the game level editing environment BattleCraft™. BattleCraft is a level editor for the game "Battlefield 1942". Before starting level editing, the scenario, created by one of the coauthors at FOI, was described as a foundation for the layout of the game world. The description of the stimulus is as follows.

4.4 Scenario for the Verification of Automated Gaze Object Logging

The player plays a soldier alone in a city environment. The task is to navigate through a city labyrinth to find the rest of their group. The level of challenge might be increased by increasing the time pressure to find the group, e.g. using the scenario that an enemy artillery attack is expected within x minutes.

In order to be able to find the group, the player can use a mini map; the use of the minimap (i.e. how often and how long the player looks at it) is an example where eyetracking can provide useful design information.

In the game environment several "navigational clues" can be provided as to where the player's group can be found (e.g. tracks in the terrain, road signs, wheel tracks from their tank, "red ribbons around branches", lit torches, signals/arrows painted on walls, dead bodies, bullet holes, exploded holes in walls, blown up cars, discarded weapons, etc.). It is also possible to have allied soldiers in the town that the player can run up to and gain clues from (e.g. by dialog text such as "I saw them by the church").

Threats/objects to avoid in the terrain include remotely deployed antipersonnel mines ("fjärrutlagda" mines, FASCAM mines that are deployed by artillery, which means that they are openly exposed on the ground and easy to see), improvised explosive devices (IEDs, booby-traps, which are hard to see), enemy vehicles and soldiers that patrol the area. Game testing will be needed to tune the severity of damage from these threats.

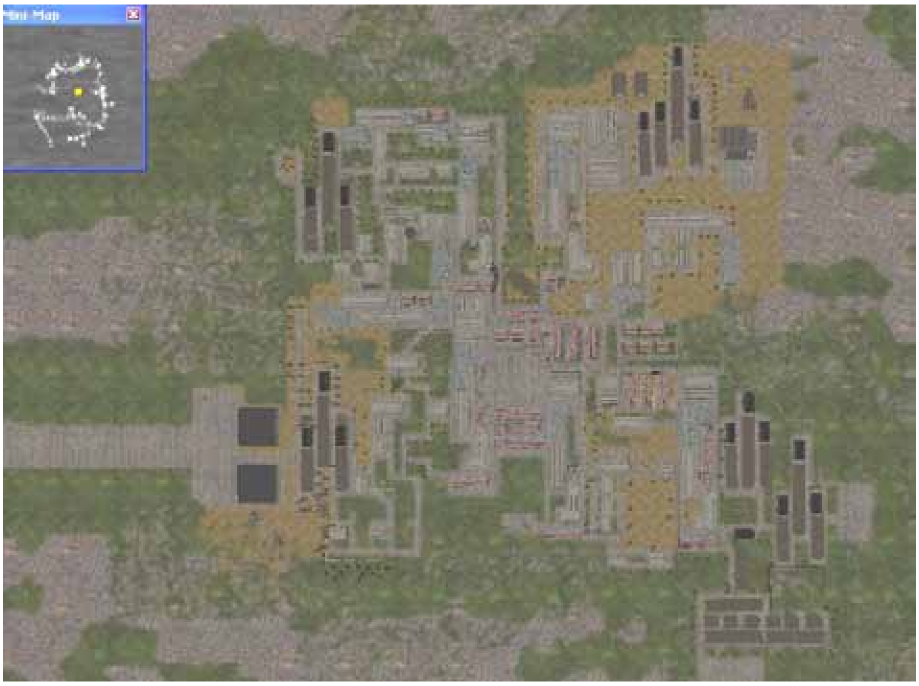


Figure 12. An overview of the level with 4 starting points,
SW (South West), NW, NE and SE.

A number of "power-ups" and objects to collect are located in the game world. The player has a secondary task to collect ammunition for the machine gun since the group is low on ammunition.

All of the above mentioned objects and mines have to be graphically large enough to be certain of being able to detect gazes upon them from the eyetracker.

With the following scenario in mind, the first author designed the maps presented in figure 12 to 16, which are images from the level editor (these high level displays are not available to the player within the game itself). The map shall be possible to copy so it is possible to play the same map several times but entering it from a different direction. This creates a labyrinth of four quadrants entered from each compass direction and having approximetely the same difficulty.

The stimulus has met the requirements mentioned above except for the provision of clues. This is a major element for gameplay but because of time limitations this was not implemented for the verification study, where the logging of the objects is the priority, and not the gameplay itself.

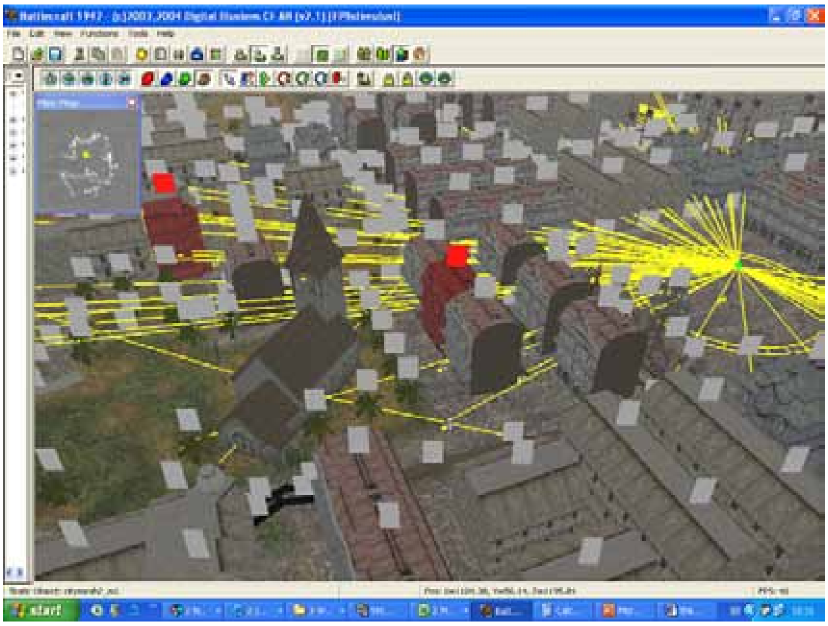


Figure 13. The objective of the player is to reach the church. The player is told that their group can be found near the square. The number of opponents increases as the player gets closer to the squareand to the church.

Figure 14. A top-down-view of the environment (North West corner), the game objects and their positions.

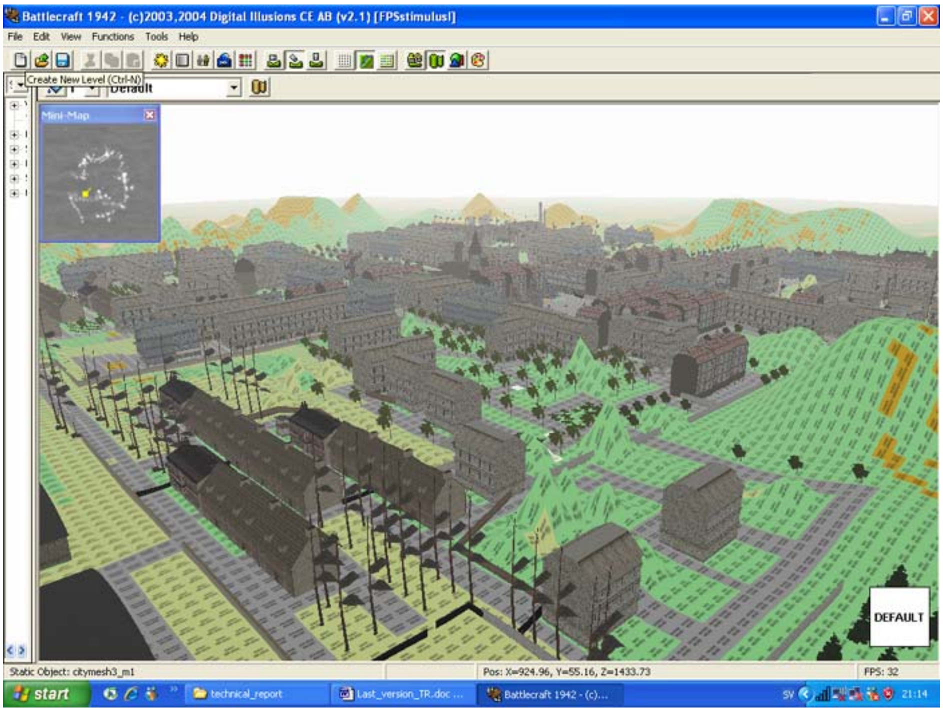


Figure 15. The BattleCraft “Camera Fly Mode” and the “Material Mapper” view.

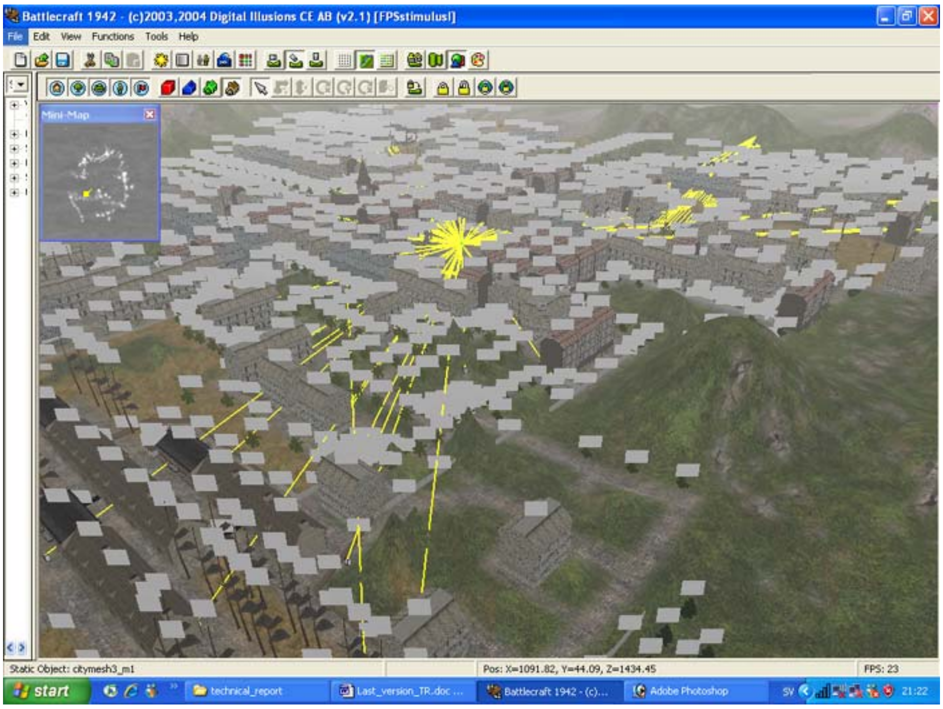


Figure 16. The BattleCraft “Camera Fly Mode” and the “Object Mapper” view.
Spawn points are added to design the combat opportunities.

4.5 Discussion of Observations Made During the Verification Process

This section presents commentary and examples of accuracies and inaccuracies observed in the object logging while creating the video transcript and cross-reference to the HiFi object log. Undertaking this analysis requires detailed investigation of the relationships within and between close rows through all 17,064 of the sorted data table, with reference also to the display of the video frames in the video file.

Note that some instances are discussed primarily in the body of the text while others are only discussed within figure captions. Examples of observed errors and issues of interpretation are given in this section, while error frequencies are presented in the section ‘Analysis: Resulting Accuracy and Precision Model’ below.

4.5.1 Picking and raytrace errors

Picking errors occur when the engine is not logging the raytraced object correctly within the virtual world. The raytrace is a linear pointer and the target object has to be hit by that line. Raytrace errors happen when a raytrace goes through the ‘correct’ object and the engine picks out another object behind it. An examplification of a raytrace error follows below.

3667	terrainobject	TStart:	83,284	81,744					
3668	terrainobject	TStart:	83,312	81,773					
3669					3,00	-1,00		81,800	germansoldier_965_1052
3670	germansoldier_965_1052	TStart:	83,343	81,804					
3671	terrainobject	TStart:	83,371	81,831					
3672	terrainobject	TStart:	83,397	81,857					
3673	terrainobject	TStart:	83,428	81,888					
3674					1,00	999,00	rayerror	81,900	citymesh4_closed_m1_976_1042
3675	citymesh4_closed_m1_1083_1054	TStart:	83,456	81,916					
3676	terrainobject	TStart:	83,484	81,944					
3677	citymesh4_closed_m1_1083_1054	TStart:	83,512	81,973					
3678					4,00	999,00	ocular	82,000	germansoldier_965_1052
3679	citymesh4_closed_m1_1083_1054	TStart:	83,542	82,003					
3680	citymesh4_closed_m1_1083_1054	TStart:	83,569	82,029					
3681	citymesh1_closed_m1_977_1059	TStart:	83,596	82,056					
3682	citymesh1_closed_m1_977_1059	TStart:	83,625	82,086					
3683					4,00	999,00	ocular	82,100	germansoldier_965_1052
3684	citymesh1_closed_m1_977_1059	TStart:	83,666	82,127					

Figure 17. On row 3,674 one can see the transcribed object on the right side and on the next row (3,675) shows the logged object by the HiFi Engine.

Figure 18. Player´s gaze is on corner of building citymesh4_closed_m1[976_1042].

In frame 819 the avi and the transcription state that the gaze-point-dot-representation represents a gaze point on citymesh_4 _976_1042 (row 3,674), but the object log says citymesh_4_1083_1054 (row 3,675). See Figure 17, 18 and 19.

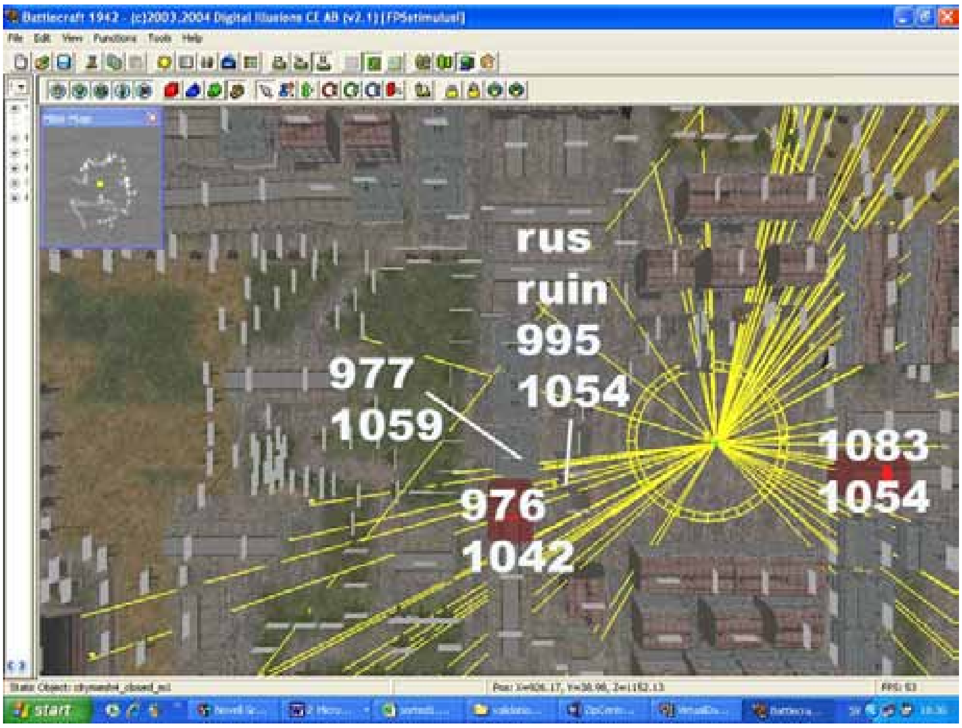


Figure 19. The player is on the left side watching building citymesh4_closed_m1 [976_1042] but the Object Log (HiFi Engine) logs citymesh4_closed_m1 [1083_1054].

In row 3,679, it could be that the house on the other side of the square is correctly picked. Here is an ambiguity though, see figure 17 and 19. Between two houses the germansoldier_[965_1052] stands and behind him there is a ruin with open windows (figure 20). Row 3,679 is either a raytrace error through the german soldier or the visual interpretation is wrong. The fixation marker suggests that it is a visual error.

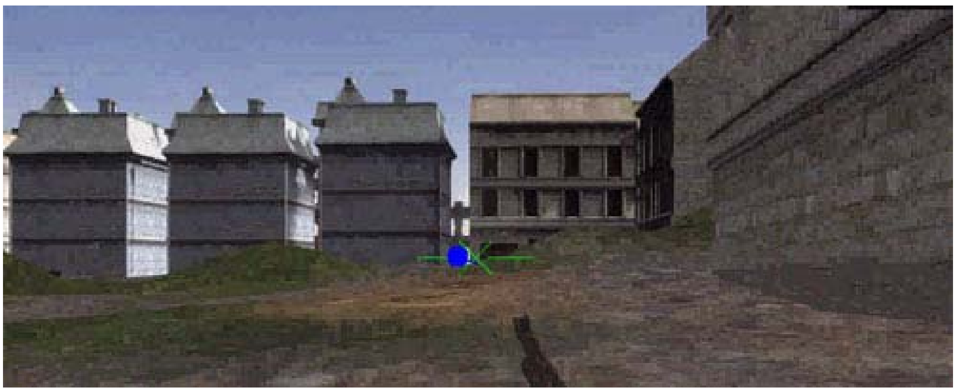


Figure 20. “Print Screen” from situation row 3,679. Soldier is behind blue marker.

Figure 21. Both highlighted soldiers were fixated many times, but only three log entries for the right hand soldier were generated.

Neither of the soldiers in figure 21 has been logged even if the fixation has been placed on central parts of the body representations. The soldier on the right has had only three of its fixations logged, while the left soldier appears not to be logged at all. However, the wdfence in the background is logged all the time.

5680	terrainobject
5681	citymesh1_closed_m1_976_1092
5682	
5683	germansoldier_972_1103
5684	germansoldier_972_1103
5685	germansoldier_972_1103
5686	
5687	citymesh1_closed_m1_976_1092
5688	citymesh1_closed_m1_976_1092

Figure 22. On row 5,683 the right soldier is logged three times in a row.

Below is an example of when the player is inside citymesh2_m1_interior and the raytrace goes through the wall.

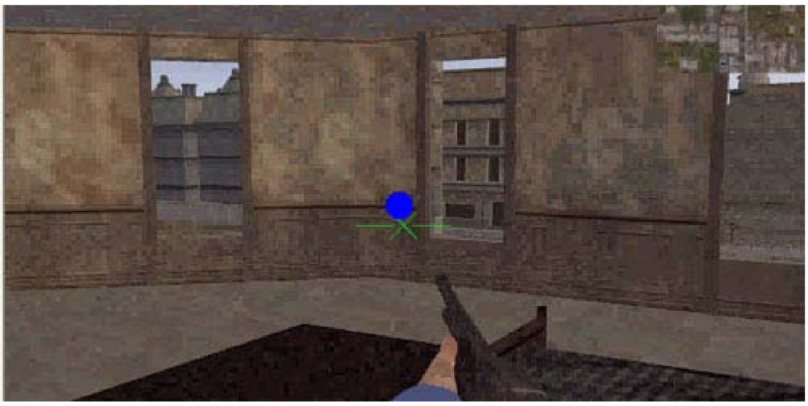


Figure 23. The player is inside the building and fixations are distributed mainly on the indoor walls. This happens from row 11,859-11,864 /frame 2605-7.

11859	citymesh2_m1-citymesh2interior_0_0
11860	citymesh4_closed_m1_1081_1254
11861	eu_strtlight_m1_1065_1245
11862	
11863	citymesh4_closed_m1_1081_1254
11864	citymesh2_m1-citymesh2interior_0_0
11865	citymesh2_m1-citymesh2interior_0_0

Figure 24. This shows what is logged while the player is inside Citymesh 2. In figure 23 the blue marker is on the wall but the raytrace seems to go through the wall to pick eu_strtlight instead. The eu_strtlight and citymesh4_1081_1254 are outside in the same direction and behind the blue marker and the wall. The house though is more likely to be picked up because it is a bigger object and could be logged while player is sweeping their view across the window.

.

Another example of raytrace error is presented in figure 25.

Figure 25. Here the ray-trace goes through the finger graphic and picks the house object in the prolonged direction behind it [row 14,144 / frame 3,129].

Figure 26. The same problem occurs but the ray-trace hits the ground instead of the weapon or the right hand [row 14,211 / frame 3,144].

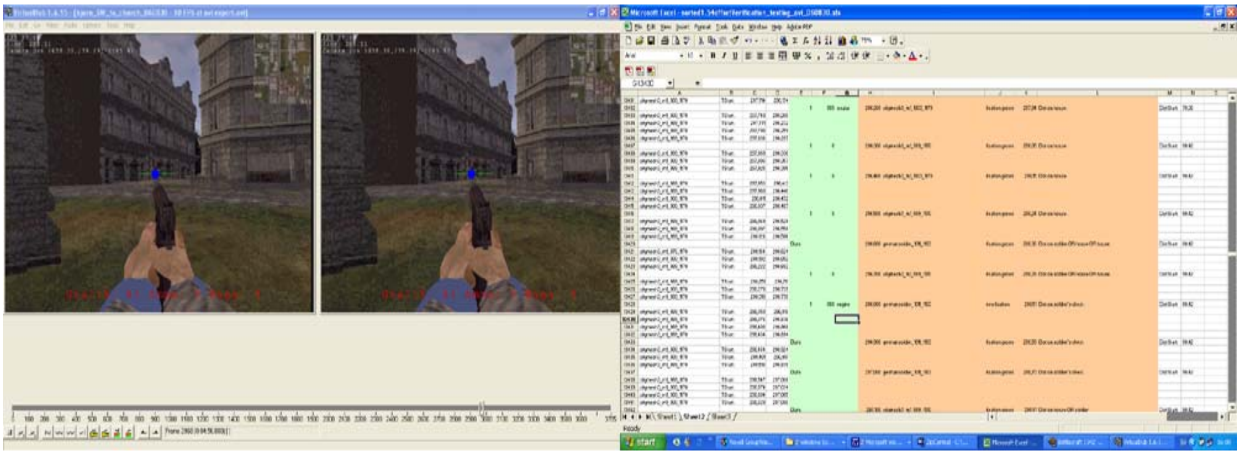


Figure 27. The engine seems to have difficulty logging the soldiers. Here is an obvious case where a soldier should be logged and it is [row 14,085 / frame3,116] but most of the time the building behind is logged. See also Figure 28.
.

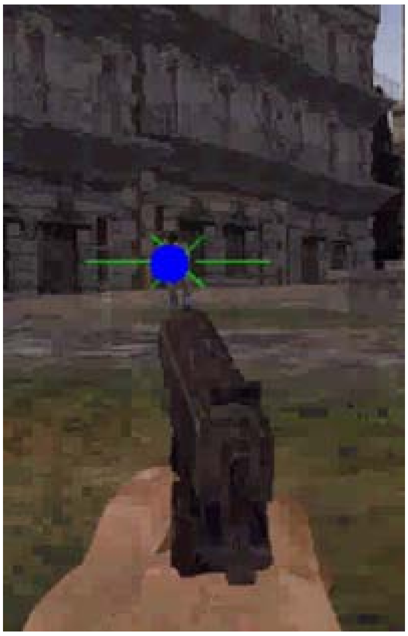


Figure 28. An enlargement from Figure 27.

In figure 28 we can without doubt see that the fixation is directed at the centre of soldier´s chest, but most of the time the Object Log records the house in the back-ground. Raytracing goes through the object in these cases.

4.5.2 Saccades and logging

Saccades and fixation histories over the past 100 ms are represented on the AVI file, to ensure no data loss since 100 ms is the AVI inter-frame time. Sometimes, however, the saccades do not show.

Figure 29. Saccades can cover 2 objects and more, so what is under a saccade is tracked and logged, but maybe not shown as a fixation. See row 1,304.

suburbhouse_2_closed_m1_839_977	TStart:	31,615	30,076	5	-4 saccade	30,100 ussoldier-bar1918complex_0_0
suburbhouse_2_closed_m1_839_977	TStart:	31,640	30,100			

Figure 30. The spreadsheet entry corresponding with the image in Figure 29 shows how the house is logged instead of the actual fixated object as shown in the AVI file. The fixation itself is logged later (with a -4 row lag).

4.5.3 Rendering problems

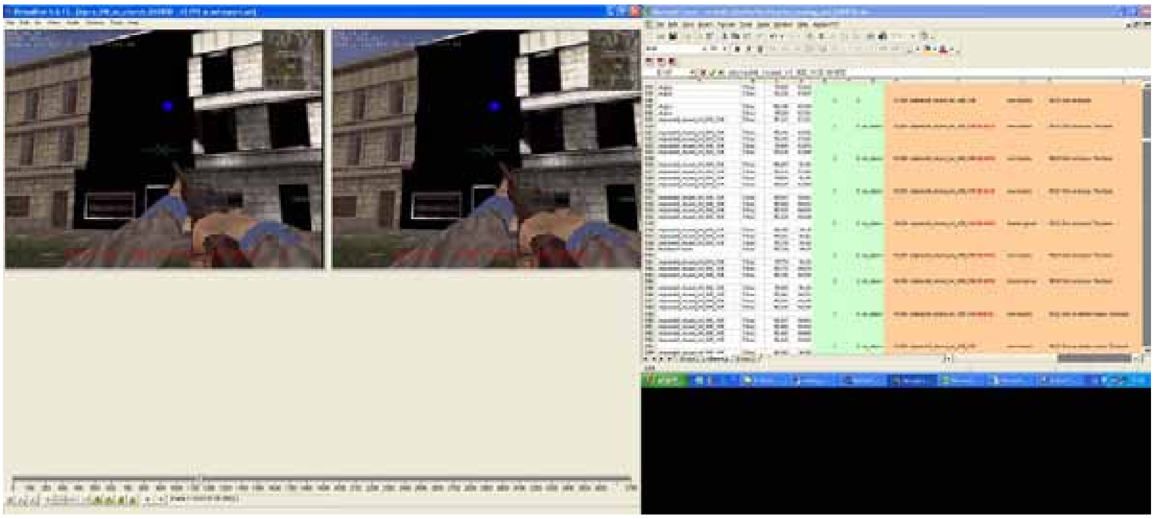


Figure 31. Here a problem with the logging occurs and it seems to be a rendering problem. A house of this type has a “window-door-window-door” sequence at the bottom level. The black and white part of the same house is logged here as two different objects. The left black part of the house, which has not been rendered correctly, seems to be logged as the house on its left side, Row 5,119. When the fixation has been on the white side and goes back to the black side, it is then logged correctly again.

4.5.4 Obscuration problems

The blue marker representing the gaze point increases in radius over time to represent gaze duration (i.e. fixation time) on the same object(s). Unfortunately the marker is opaque, so precise positions and jerky eye movements under the marker cannot be analyzed in detail. Since one fixation can have a duration of over 29 frames, this is a major disadvantage when analyzing detailed visual content from the AVI file.

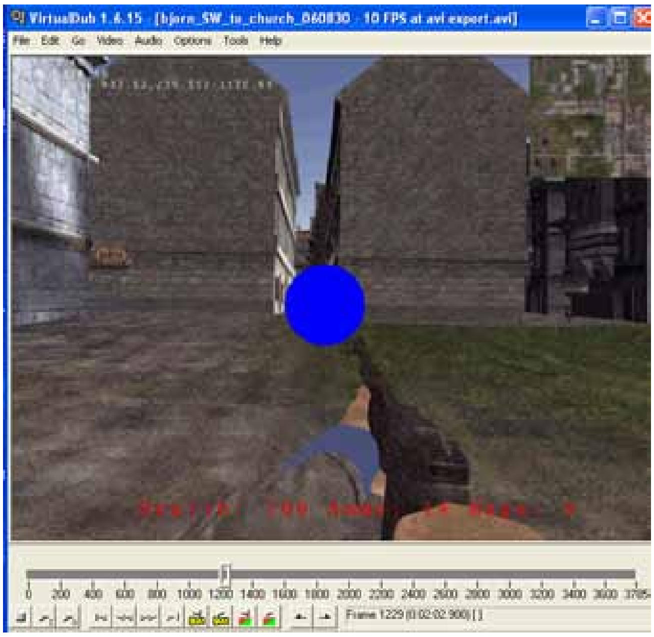


Figure 32. A typical case when a fixation is targetting one object and the marker representing it increases over time to obscure the detail beneath. If the marker is turned off, gaze time information is lost.

5467	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5468					3,00	999,00	ocular	121,200	germansoldier_972_1103	WDFENCE	fixation grows
5469	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5470	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5471	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5472					3,00	999,00	ocular	121,300	germansoldier_972_1103	WDFENCE	fixation grows
5473	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5474	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5475	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5476					3,00	999,00	ocular	121,400	germansoldier_972_1103	WDFENCE	fixation grows
5477	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5478	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5479	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5480					3,00	999,00	ocular	121,500	germansoldier_972_1103	WDFENCE	fixation grows
5481	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5482	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5483	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5484					3,00	999,00	ocular	121,600	germansoldier_972_1103	WDFENCE	fixation grows
5485	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5486	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5487	wdfence1_fen1_m1_1012_1123	TSt	##	##							
5488					3,00	999,00	ocular	121,700	germansoldier_972_1103	WDFENCE	fixation grows
5489	wdfence1_fen1_m1_1012_1123	TSt	##	##							

Figure 33. This is an example of how visual interpretation problems occur due to the opaque fixation marker increasing over time. The transcription says [germansoldier_972_1103] in column "I" and in column "J", while the transcription says "fixation grows". In column "A" the Object Log [wdfence_1042_1123] has logged differently. The reason for this is the increasing size of marker, which is not transparent.

4.5.5 Collision detection errors

Graphical objects within a virtual game world have a bounding volume that is referred to by the physics system to detect collisions between that object and other objects having bounding volumes. When adding these objects to a graphical model, the volumes can have different shapes from the model shape depending on what kind of triggers or actions one wants to implement for these objects, and depending upon the required accuracy of collision detection (in relation to the shape of the graphical mesh). If there is no bounding volume at all around an object, then it does not exist as far as the physics system is concerned. Since the integrated HFE/eyetracking system uses the collision of a ray trace with a bounding volume to identify the object under a gaze point, an object having no bounding volume will not be logged by the system as an object within the world.

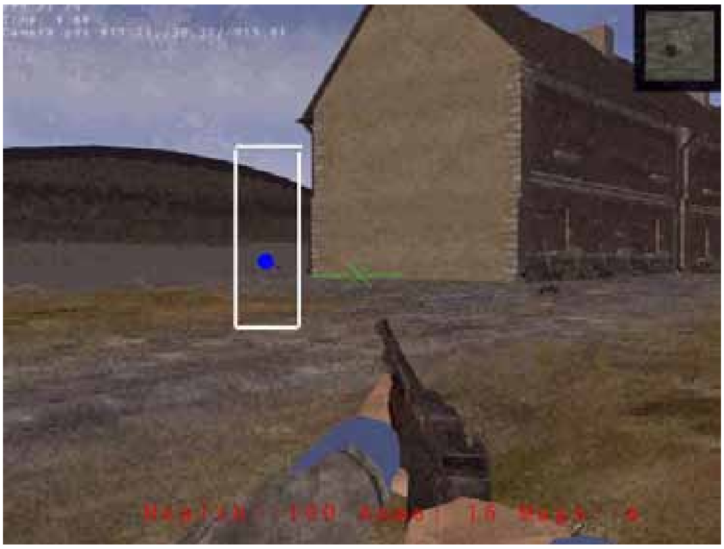


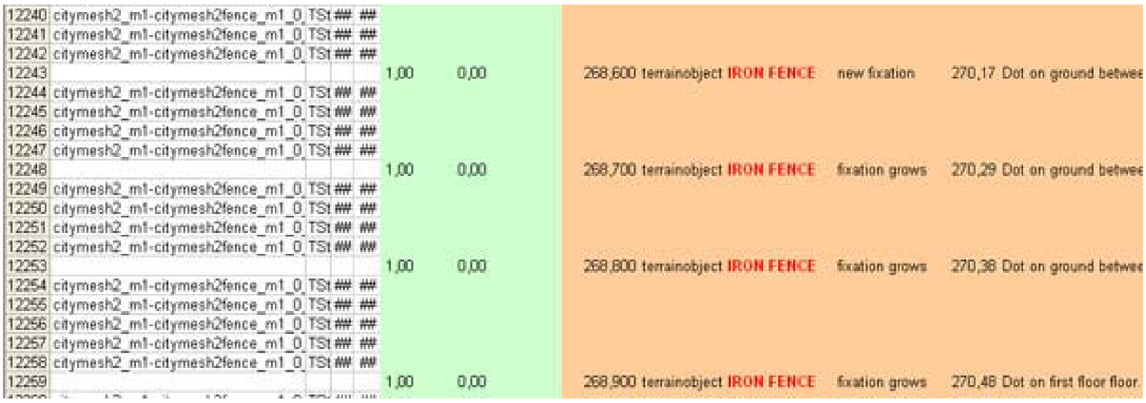
Figure 34. In this case the wall/fence does not have any collision detection and therefore, since the object under the gaze point is detected by the engine using a collision volume/surface, the hangar is logged instead.

Figure 35.

Figure 36.

In this case the soldier sinks into the building after being shot. However, this is a collision detection error that is not relevant to verification of the eyetracker integration per se.

4.5.6 Collision mesh accuracy



12240	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12241	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12242	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12243					1,00	0,00	268,600	terrainobject	IRON FENCE	new fixation	270,17	Dot on ground betwee
12244	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12245	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12246	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12247	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12248					1,00	0,00	268,700	terrainobject	IRON FENCE	fixation grows	270,29	Dot on ground betwee
12249	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12250	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12251	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12252	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12253					1,00	0,00	268,800	terrainobject	IRON FENCE	fixation grows	270,38	Dot on ground betwee
12254	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12255	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12256	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12257	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12258	citymesh2_m1-citymesh2fence_m1_0	TSt	##	##								
12259					1,00	0,00	268,900	terrainobject	IRON FENCE	fixation grows	270,48	Dot on first floor floor.

Figure 37.

Figure 38.

Figures 37 and 38 show how ambiguous it can be to estimate what the gaze is actually directed at when collision detection is carried out on transparent parts that may be looked through [row12,224 / frame2682-6]. It would be possible in principle to create collision meshes corresponding exactly with all graphical boundaries. However, other spatial inaccuracies in the system (e.g. the spatial accuracy of eyetracking) suggest a limit to the usefulness of this, although it is also dependent upon virtual distance to the object: from far away, the spatial accuracy of the eyetracker is the dominant limiting accuracy factor, while very close up the accuracy of the collision volume in relation to the graphical object is the dominant factor.

4.5.7 Spatial Ambiguity due to distance

The problem of ambiguity due to the small-scale spatial inaccuracy of eyetracking, together with obscuration by the gaze point graphic, also applies for objects that are close together and have a far virtual distance from the player character viewpoint. Small scale, natural jerky eye movements also become potentially significant in size for distant objects that are close together.

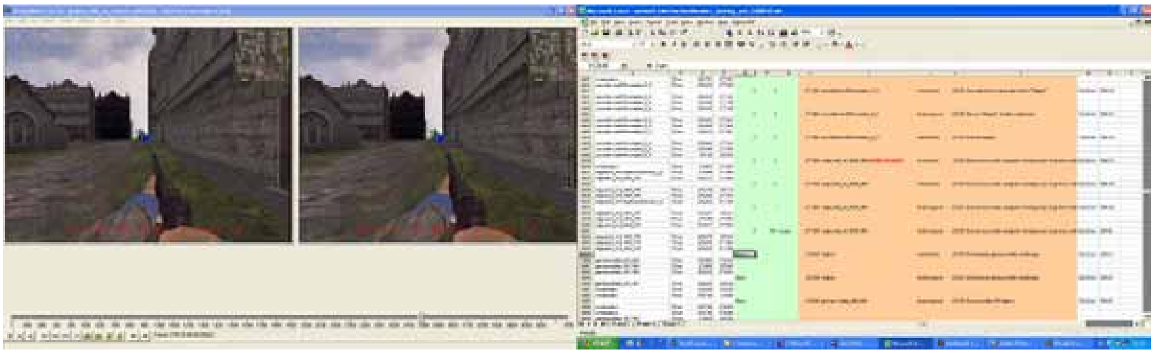


Figure 39.

In figures 39, 40 and 41 it is hard to interpret what is being fixated when the fixation dot is centered over a cluster of distant objects, especially when the dot itself is increasing over time. The problem can be illustrated more clearly in close up images.

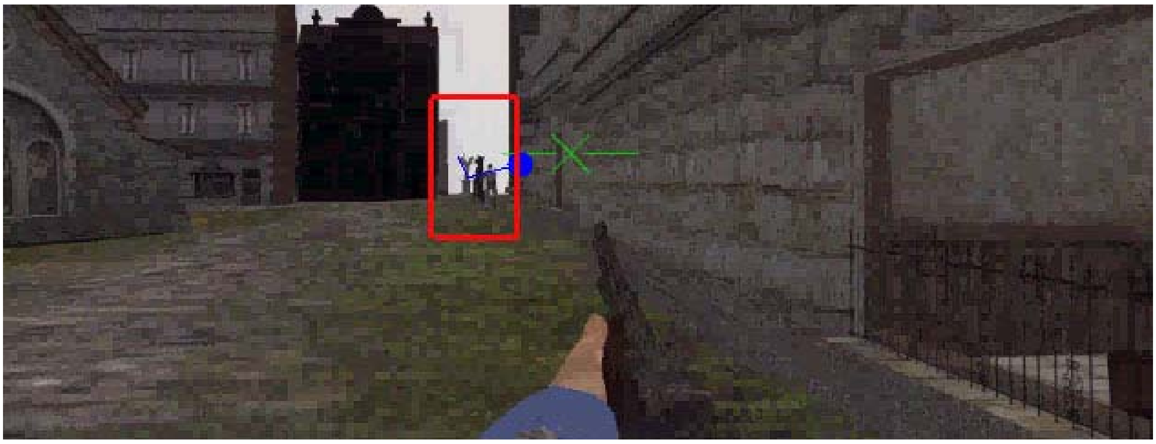


Figure 40.

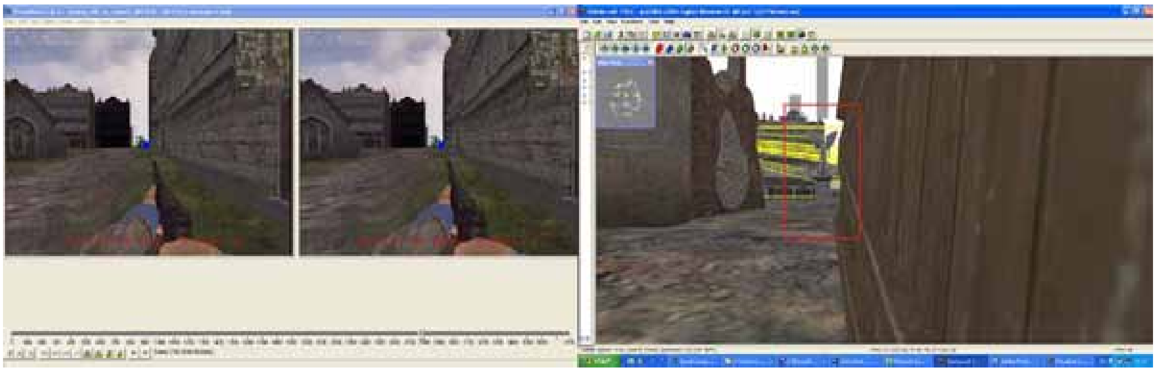


Figure 41.

From left there is an angel_sculpture, in the middle a fence in profile that consists of three individual models in depth, and to the right there is a soldier. The scene is shown from a distance in the left of figure 42.

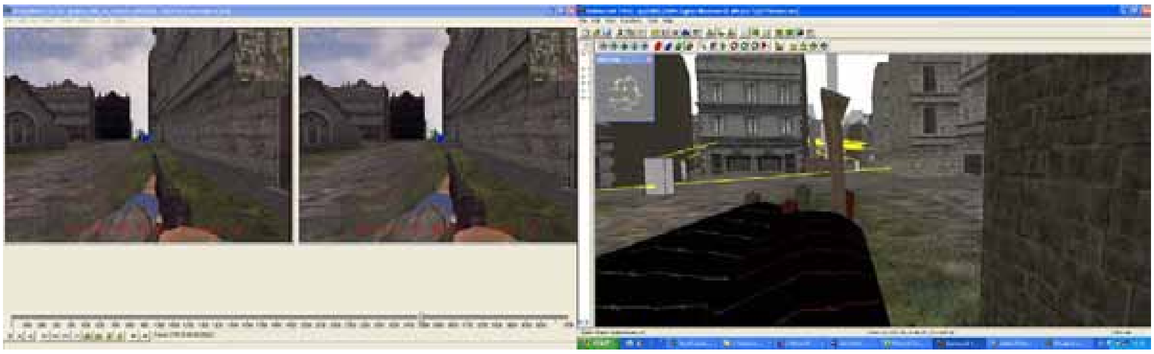


Figure 42.

The fence has two more pieces behind it that are not viewable from this angle (figure 43).

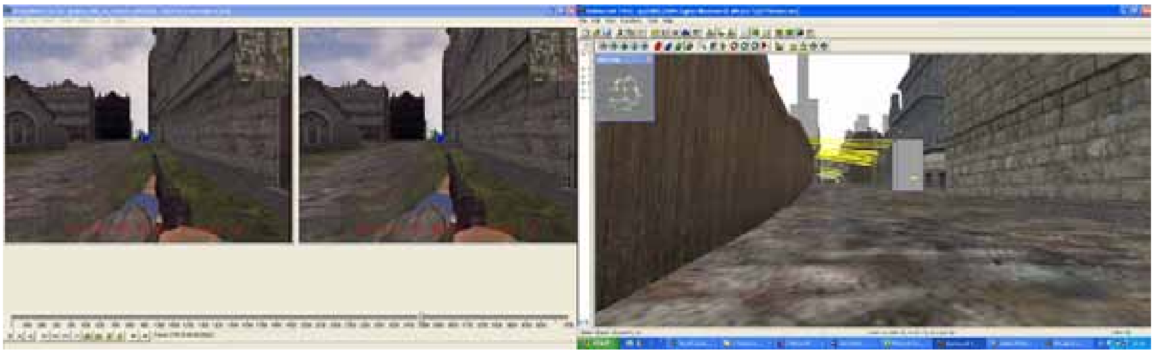


Figure 43.

On the right there is a soldier (the grey box) and the fence is also visible from the right side (figure 43).

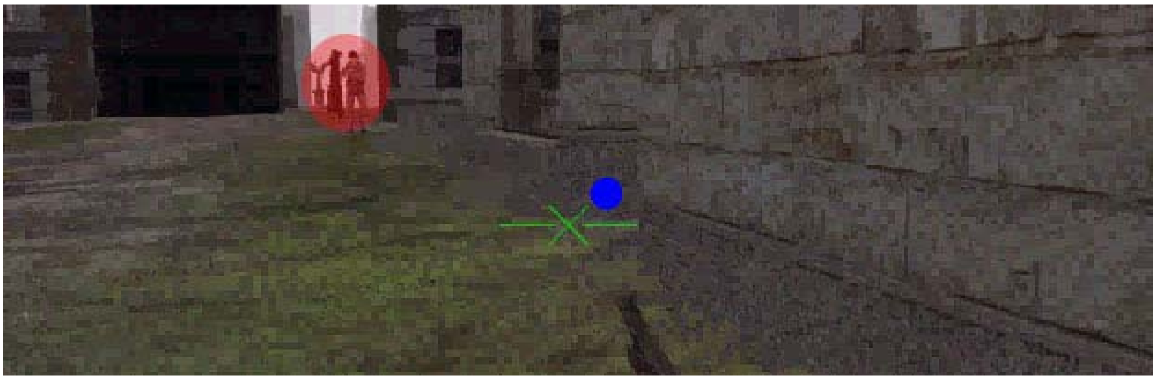


Figure 44.

In fig.44 a transparent circle is added and placed over the three objects to illustrate the disadvantage of the opaque (blue) marker currently available within the Clearview software, which can be the same size as the circle, obscuring all of the relevant detail. The system could alternatively show just the circle boundary so the transcriber can see what is behind the circle. A third alternative would be a cross, centred upon the fixation and with a size depending upon fixation time.

4.5.8 Physics errors

Physics errors other than raytracing collision errors are not immediately relevant to the verification of the integrated eyetracker/HFE system, but may have a bearing upon interpretation of the data provided by the system in later studies. This is because any such errors that are visually observable by players may represent novel events that attract the attention of a player by virtue of their unusualness and unexpectedness. The resulting gaze behaviour will then not be related to visual attention patterns that characterise task-related decision processes.

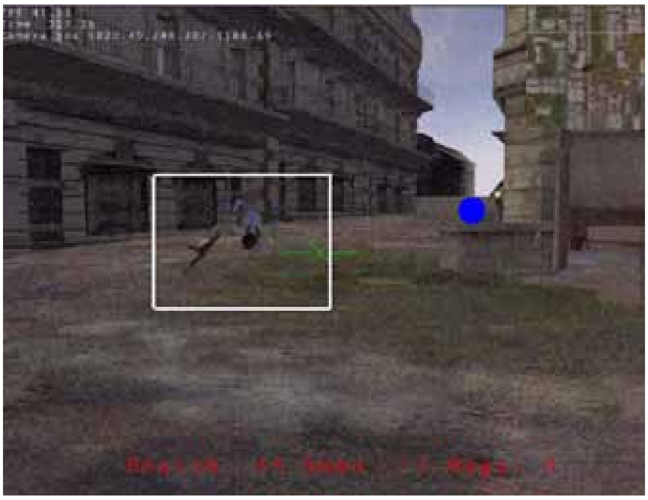


Figure 45. An enemy soldier floats in the air when killed!

4.5.9 Other engine errors

Specified “distance logging”, referring to the virtual distance from the player character observation point to the object that is the target of each fixation within the virtual world, is not carried out by the version of the HFE software used for the verification.

5 Analysis: Resulting Accuracy/Precision Model

5.1 Characterising Accuracy and Precision

The verification of the integrated eyetracker/HFE system reported here has been undertaken by comparison of the eye gaze data logged by the HiFi engine with captured video data of the history of game play with eye gaze data superimposed over the image in the form of dots representing fixation points and lines representing saccades. For analytical clarity and for transfer of the results of this study to the interpretaion of data in experimental (rather than verification) studies based upon the use of the integrated eyetracker/HFE system, it is necessary to derive an abstracted and transferable model of the performance of the system from the detailed data used in and derived from the verification analysis. Such a model represents the conceptual result of the verification study.

It is useful and appropriate for this conceptual result to take the form of an empirically derived model of the accuracy and precision of the system. In general, it may be stated that:

> “In the fields of science, engineering, industry and statistics, *accuracy* is the degree of conformity of a measured or calculated quantity to its actual (true) value. Accuracy is closely related to *precision*, also called reproducibility or repeatability, the degree to which further measurements or calculations will show the same or similar results.
>
> “The results of calculations or a measurement can be accurate but not precise; precise but not accurate; neither; or both. A result is called *valid* if it is both *accurate* and *precise*.”
>
> (from http://en.wikipedia.org/wiki/Accuracy_and_precision).

In other words, *accuracy* is the degree of veracity (or closeness to the truth) of a measure, while *precision* is the reproducibility of the measure of accuracy.

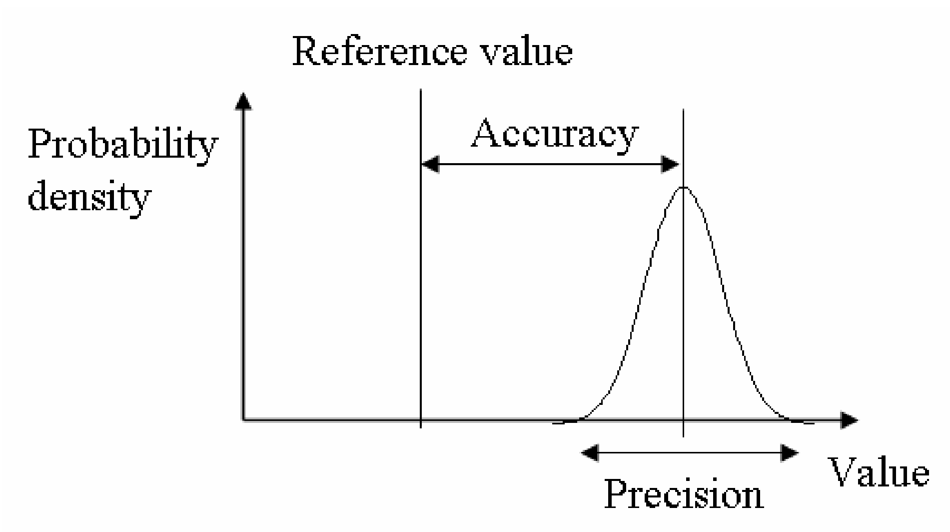


Figure 46. A graphical representation of the concepts of accuracy and precision (from http://en.wikipedia.org/wiki/Accuracy_and_precision).

In the current verification study, data can be extracted from the sorted and analysed object log and video transcript table to consider accuracy and precision in both the spatial and the temporal performance of the system. Considering accuracy and precision specifically relating to the integrated system (and not its two primary component subsystems, HFE and the eyetracking system) these terms can be applied in the following ways:

Spatial:

Accuracy:

Each object log entry reports one object under the gaze point. Accuracy can be assessed by comparison with a known actual number of objects within the region of accuracy of the eyetracking system. The result can be expressed in terms of: for a given record for an object within the object log, what is the probability that there were actually n objects under the gaze point, for an arbitrary +ve integer n?

Precision:

Given a characterisation of spatial accuracy as described above, spatial precision can then be assessed in terms of the repeatability of the probabilities obtained for different values of n objects under the gaze point.

Temporal:

Accuracy:

Each object log entry reports the time at which a specific object was under the gaze point during one simulation time cycle of the game engine. Accuracy can be assessed by comparison with a known actual time of that specific object under the gaze point during one simulation time cycle of the game engine. The result can be expressed in terms of: for a given record for an object within the object log occurring at a time t_{log}, what is the probability that the actual time at which the gaze occurred is t_{log} +/- t_{offset}, for t_{offset} in some known real number range of seconds?

Precision:

Given a characterisation of temporal accuracy as described above, temporal precision can then be assessed in terms of the repeatability of the probabilities obtained for different values of t_{offset} from the time at which an arbitrary object is recorded as being under the gaze point.

Repeatability in both cases could be assessed by looking at probabilities derived: a) from different subsets of data within a single verification test run, or b) from data across multiple verification test runs.

In developing accuracy and precision models in detail below, other aspects of accuracy and precision will be considered in addition to the above components specifically relating to the integrated system. This results in the following categories of the components of spatial and temporal accuracy and precision:

- the visual system of the player
- the HFE system
- the Tobii eyetracking system

These categories will be considered in turn below, first for spatial accuracy and precision and then for temporal accuracy and precision.

5.2 Detailed Model of Spatial Accuracy and Precision

5.2.1 The Visual System of the Player

It is useful to briefly summarise the accuracy or precision of the human visual system, especially in relation to the other systems. Foveal resolution refers to the visual angular spatial resolution of the most sensitive, high resolution area of the retina within the eye. The highest resolution of the human eye (i.e. highest perceivable spatial frequency cycle) has been characterized to be around 30 arc seconds. At the optimal viewing distance of the Tobii eyetracking system of 60 cm (and hence the distance of the game screen from the player's eye in this study), this represents a distance of 4.6e-7 m on the screen. The Tobii screen measures 0.34 x 0.27 m, which at a pixel resolution of 1024 x 728 gives pixel dimensions of 3.3e-4 x 3.7e-4. Clearly, foveal resolution is insignificant compared to the pixel size on the screen.

Eye pointing accuracy can be considered in terms of muscular tremor, which has a variation of 20-40 arc seconds. Since this is around the same value as the 30 arc seconds of foveal resolution, it can similarly be disregarded as a factor compared with apparent pixel size.

A potentially significant factor in the accuracy of vision is the role of peripheral vision. This is dependent upon the task, the task environment and the learned visual competence of the observer within the environment and so should be studied within that context.

5.2.2 The HFE System

The HFE system and the game environment are sources of basic limitations in spatial accuracy. If the game engine is subjected to high processing demand during game play, it could slow down to a point of introducing noticeable visual delays that might interfere with the mapping of screen (x,y) coordinates onto virtual visual objects. However, this was neither observed nor specifically tested for in the study. Graphical discrimination due to factors of hue, saturation, lighting, materials and textures can make a difference to the perceivability of objects (e.g. whether a player perceives a boundary between objects or not). These factors are dependent upon the basic graphics and rendering functions of the system. Again, these factors were not specifically tested for in the study.

A factor that did have an observed impact upon the accuracy of the system was the correspondence of specified object bounding volumes within the game design with the graphical objects that they are associated with. This factor is subsumed within the human/machine error statistics described below.

Maximum screen resolution, with its influence on the position of edges, is a function both of screen capacity and rendering capacity. Since this was dependent in this study upon the Tobii eyetracker screen, it is considered in the next section.

Additional game engine errors having a bearing upon the study include rendering errors (eg. an object is not visible when it should be), raytracing errors (concerning engine functions added for integration with the eyetracker, amounting to failure to intercept objects under the point of gaze even though the gaze coordinates have been successfully retrieved from the eyetracker) and errors in the position of objects. These factors are subsumed within the human/machine error statistics described below.

5.2.3 The Tobii Eyetracking System

As noted above, the recommended screen resolution of the Tobii eyetracker is 1024 x 768 pixels, resulting in pixel dimensions of 3.3e-4 x 3.7e-4 m. The Tobii eyetracker has a stated accuracy of 30 pixels, which can be interpreted as a diameter of about 1 cm at the recommended viewing distance of 60 cm. This is a significant factor in spatial accuracy.

Nyquist (or Shannon's) sampling theorem states that a signal must be sampled at least at the same rate as its highest frequency component in order to avoid aliasing (i.e. reading higher frequencies as lower frequencies). 1 cm accuracy of the eyetracker (corresponding to 100 spatial wavelengths per metre) means that in principle, the highest spatial frequency that can accurately be reproduced is 50 cycles per metre, or 2 cm. Hence periodic spatial frequency variations of a higher frequency than that cannot be accuractely discriminated as gaze objects.

Regarding screen resolution, sampling theorem allows spatial frequencies greater than or equal to 2 pixels in size to be reproduced without spatial aliasing, amounting to a visual size of 6.6e-4 m horizontally and 7.4e-4 m vertically. Clearly, the eyetracker spatial frequency limits are a far more significant factor than screen resolution in determining the spatial accuracy of the system. Hence screen resolution can be disregarded.

Graphical discrimination (as affected by hue, saturation, light, material and texture) is a function of the game design, rendering engine and the Tobii screen itself. This factor is not considered in the study.

Calibration of the eye point of gaze is a critica factor in determining the spatial accuracy of eyetracking. However, the quality of calibration data is assessed automatically (the calibration procedure can be repeated in the case of poor calibrations), with no quantified indication of the meaning of a good calibration. Hence this factor is not considered in the study, other than by using only what the ClearView system regards as 'good' calibrations.

For the verification procedure, the spatial accuracy of video capture and the representation of gaze position over the captured video for interpretation are significant factors. The recommended screen resolution of captured video is 640 x 480 pixels, much lower than that of the Tobii screen itself. The impact of this is significant, as manually assessed and reflected in the figures for machine versus visual interpretation error presented below.

Gaze data is represented in the video file as a coloured line for saccades and as an opaque dot for fixation points. The line is thin enough to be disregarded as a limiting spatial accuracy factor. The fixation point dot can be set with a radius of between 10 and 100 pixels, with a default value of 30. Since the eyetracker has a rated accuracy of 30 pixels, a setting of 15 pixels radius corresponds to the spatial accuracy of eyetracking. If the video is analysed with a dot of this size, ambiguity due to the dot obscuring object detail should be an accurate representation of the basic spatial accuracy limits of the eyetracker. Using a larger dot size can obscure image detail and reduce the accuracy of visual interpretation of objects under the represented point of gaze to a level less accurate than that due to the accuracy limits of the eyetracker. For the study, a radius of 30 pixels was used, with the resulting ambiguities being taken into account in the summary accuracy model (below). As discussed below, this is not regarded as a critical issue, since object ambiguity under the point of gaze is a function of game level design and player tasks, making this aspect of spatial accuracy difficult to generalise into a precision model across level designs and players.

5.3 Resulting Spatial Accuracy and Precision Model

5.3.1 Spatial Accuracy

As noted above, the spatial accuracy of the integrated HFE/Tobii system can be expressed as the probability for a given record for a specific object within the object log that there were actually n objects under the gaze point, for an arbitrary +ve integer n.

Examination of the captured video data showing the game screen with superimposed gaze data revealed many cases where visual examination could not disambiguate the number of objects under the gaze point, especially within the 30 pixel diameter accuracy limits of the eyetracking system. The following table shows the frequency of the occurrence of n ambiguous objects under the gaze point, together with unrelated errors in the final column.

n	0	1	2	3	4	5	6	7	error
count	80	3331	235	112	12	3	1	0	3

Figure 47.

In Figure 47 these figures are depicted as probabilities of n Objects under the gaze point. The meaning of this probability distribution is that if the integrated HFE/Tobii system is used in a study, for each entry in the object log these is actually a probability of not 1 but n objects being under the gaze point, where the probability of there being n objects, P(n), is as shown on Figure 48.

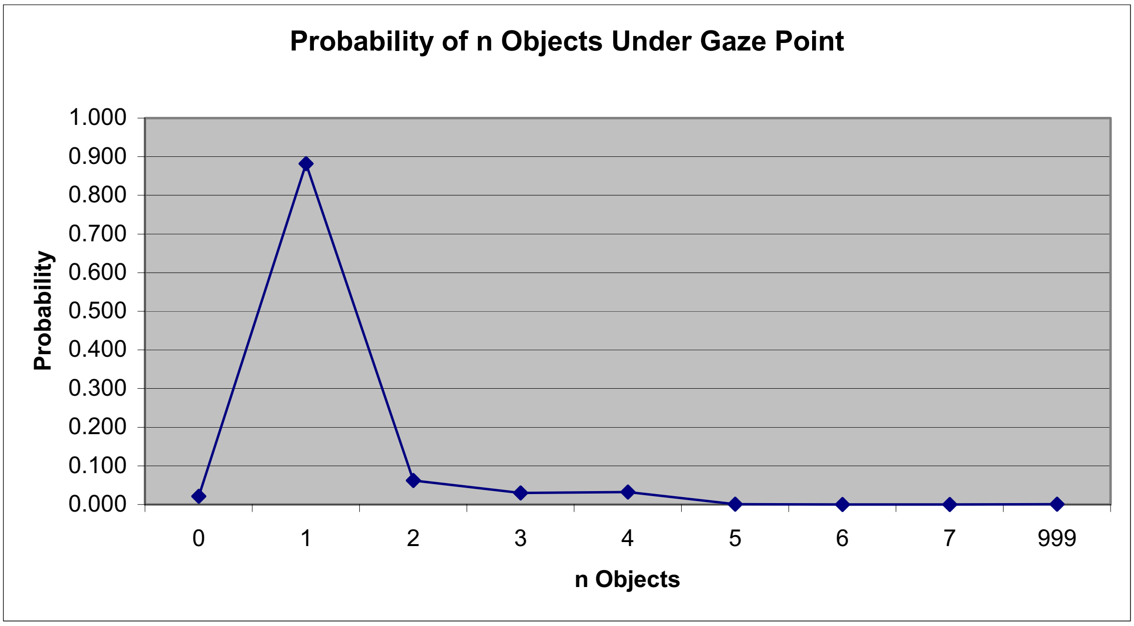


Figure 48. Probability of n Objects under the gaze point.

5.3.2 Spatial Precision

The spatial precision of the integrated HFE/Tobii system is a matter of how the accuracy derived above might generalise to other studies. No detailed spatial precision model has been derived. In general it is doubtful whether such a model can be derived, since a spatial accuracy model derived using the method above may vary a great deal depending upon stimulus design and the deailed task(s) given to a player under study. For example, a stimulus game level could be constructed with a small number of very simple objects (e.g. buildings) within the game level design. In this case, there may be few instances where the player experiences overlapping objects within a small visual field and hence a very low frequency of highly ambiguous perceptual situations. On the other hand, a stimulus could use a very large number of densely located objects (e.g. using many small plants and artefacts very close together). In this case there may be very many situations in which many small objects fall within the 30 pixel accuracy limits of the eyetracker. In both cases, the nature of the tasks given to a player may draw attention towards or away from spatially ambiguous areas of the game. For instance, a task to search for a small object may lead the player to look at areas where many plants are close together, providing good hiding places. Alternatively a large scale navigation task may lead the player to disregard clusters of small plants while looking more at large scale navigational features, such as buildings.

5.4 Detailed Model of Temporal Accuracy and Precision

5.4.1 The Visual System of the Player

The frequency of high frequency tremor of the eye is generally within a range of 30 to 90 Hz. However, since the spatial consequences of this are insignificant, this factor is disregarded.

The general sampling rate of the eye for sharp edges in the visual field, due to high frequency tremor, is 90 Hz. Sampling rates due to saccadic motion can vary within a wide range, depending upon the visual stimulus, the task, the person and the occasion. This is not considered directly here in terms of general saccadic movement, but it may be noted that the Tobii sampling rate of 50 Hz limits accurately sampled saccadic movement frequency components to those having frequencies of less than or equal to 25 Hz (see below).

Spontaneous Eyeblinking Rate (SEBR) generally occurs at 2 to 10 sec intervals, having a duration of 0.2 to 0.4 sec per blink. This is significant and results in corresponding loss of tracking by the Tobii system. However, this was not found to have any significant analytical consequences and so is disregarded in this study.

5.4.2 The HFE System

The HFE frame rate varies depending upon transient processing loads, with an average frame interval of about 30 ms; this is the same as the simulation update rate (more precisely, the game engine update rate appears to vary between 28.6Hz and 50Hz, with an average of 30 Hz). According to sampling theory, this update results in the unaliased synthesis of time frequency components less than or equal to approximately 15 Hz. The effects of this will be dependent upon the time frequencies built into a game level design, e.g. frequency components of animation sequences.

The more immediate and intrinsic effects of the variable HFE frame rate for the integrated HFE/Tobii system occurin terms of time offsets of object log entries compared to captured video frames. This variability subsumes lag, effectively encompassing the effects of both lag and frame rate variations in comparison with gaze data log data obtained from the Tobii system.

5.4.3 The Tobii Eyetracking System

The Tobii eyetracking system has a sample rate of 50 Hz, and hence a sample interval of 20ms. By sampling theory, this means that only eye movements having a temporal frequency of less than or equal to 25 Hz can be reproduced without aliasing. This is a fundamental limitation of the eyetracking system, so no higher frequency information about eye movement is available for this study. Since this frequency is nevertheless higher than that of the HiFi game engine and the frame grabber, the Tobii gaze movement data must be used as the basis from which accuracy and precision characterisations are obtained. Temporal accuracy therefore refers to the time accuracy of correlation between the eyetracker and the game engine object logging.

Video screen capture is conducted at a rate of 10 Hz, or 10 fps (hence with an inter-frame time of 100ms). Using a gaze history time of 100 ms results in all captured gaze data from one frame to the next being represented in the video output generated by the Clearview software. By sampling theory, this frame capture rate means accurate capture of frequency components having a frequency of less than or equal to 5 Hz. The relatively small scale of changes between frames as a function of the game design and its dynamics makes this acceptable as a basis for interpreting the objects under the gaze point within each frame, especially given the 30 pixel resolution of the eyetracker (since faster changes tend to occur within smaller spatial regions). Variations due to the engine frame rate are analysed in relation to the relative difference in time between when a frame is grabbed and when the corresponding object log record is made by the HiFi engine.

5.5 Resulting Temporal Accuracy and Precision Model

5.5.1 Temporal Accuracy

As noted above, temporal accuracy of a given record for an object within the object log occurring at a time t_{log} can be expressed in terms of the probability that the actual time at which the gaze occurred is t_{log} +/- t_{offset}, for t_{offset} in some known real number range of seconds.

If the game engine had a fixed frame rate that is a multiple of 20 ms and the engine and the eyetracker were precisely synchronised, every engine frame would correspond with a subsample of every n-th gaze sample taken by the game engine.

However, the game engine frame rate and lag in the interface with the eyetracker are both variable. For the verification study, the only possible reference for 'the actual time at which the gaze occurred' is the video file with overlayed gaze data. The timing of video frame capture *is* consistent, but at a much slower frame rate of 10 Hz, compared to the eyetracker sample rate of 50 Hz. With the gaze history set to 100 ms, a video frame and its gaze data represent the best available representation of where the gaze point was at the time of frame capture and for 100 ms prior to and leading up to that time.

During 100 ms the video image changes; this change is small and was not quantified during the study, but is assumed to be small enough for the frame image to represent a reasonable basis for interpreting the objects under the gaze point through the whole 100 ms gaze history represented on the frame image (no more accurate data than this is available).

Variability in the frame rate and lag in the connection with the eyetracker result in object log entries having highly variable sample timing, having an average frame interval, and hence sample interval, of 28.6 ms (with a population standard deviation of 6 ms). The object log entry is stamped with a time stamp from the game computer. Hence the log entry time value includes the variable lag in the request and reception of a gaze data point from the TET eyetracking server.

Given these uncertainties, in order to obtain an approximate time accuracy characterisation, it is assumed to be a reasonable approximation to equate time variability with relative entry position in the sorted log/transcript file. That is, from the time of a given captured video frame, the temporally closest corresponding object log file entry is found and the number of entries ahead or behind the video transcript in the sorted log file is taken to reflect its time accuracy, based upon the 28.6 ms average frame rate.

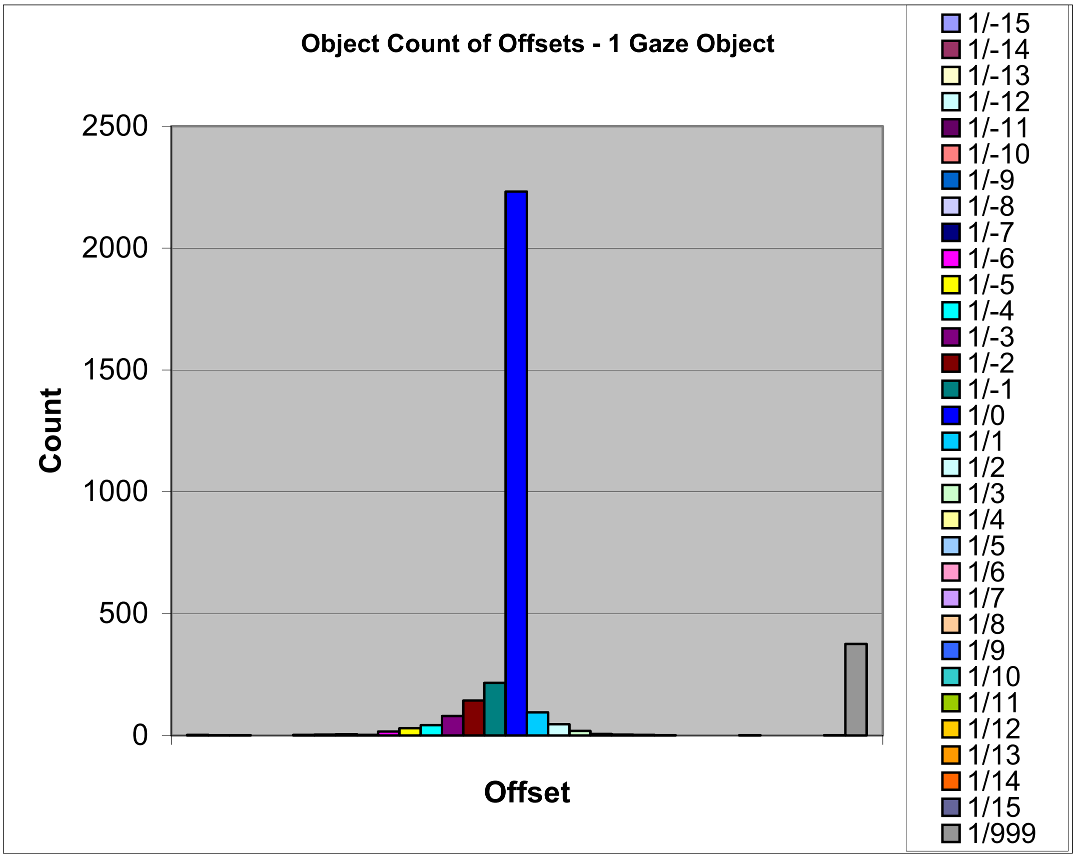


Figure 49. Frequency count of offsets of object log entries from object visibility in the video transcript.

The count of offset values is then given in the following tables. In the first the case, one unambiguous object under the gaze point is considered (i.e. the object log record corresponds with the only object under the gaze point in the video transcript). The resulting tables are as follows:

Negative Offsets (-):

-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
2	1	1	0	0	3	4	5	3	16	30	42	80	144	216

0 Offset:

1
2232

Positive Offsets (+):

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
95	46	19	6	4	3	1	0	0	0	1	0	0	0	1

These values are shown on the chart on Figure 49. Note that the ‘total match’ condition occurs when the video transcript entry time exactly matches the time of a corresponding object log entry (actually there are 0 cases of this), or when a the video transcript entry is flanked in the sorted time list by two corresponding object log entries. The chart on Figure 50 shows the same frequency distribution using a logarithmic scale.

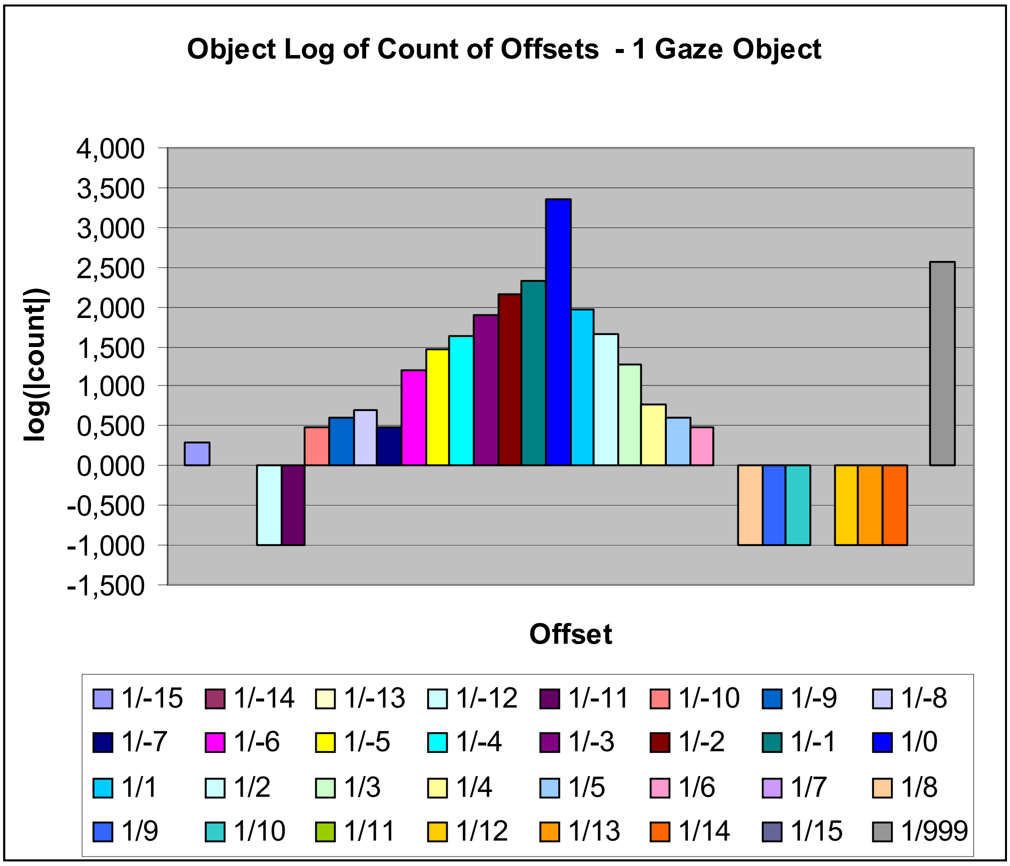


Figure 50. Frequency count of offets of object log entries from object visibility in the video transcript, using a logarithmic scale.

For the case of any number of objects under the gaze point (i.e. the object log record corresponds with at least one object out of any number that are under the gaze point in the video transcript), the resulting tables are as follows:

Negative Offsets (-):

-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
2	2	1	0	0	3	5	6	4	18	36	45	103	183	247

0 Offset:

1
2352

Positive Offsets (+):

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
101	60	26	6	5	6	2	0	0	1	1	0	1	0	1

Figure 51 shows probabilities of offsets based upon these frequency counts, both for single objects and for multiple objects under the gaze point. Hence for a given object log entry, a reasonable heuristic is that the actual time of the entry has a probability of being the stated time +/- 28.6 P(offset), where *offset* has a range from –15 to 15 and P(offset) is as depicted on Figure 51.

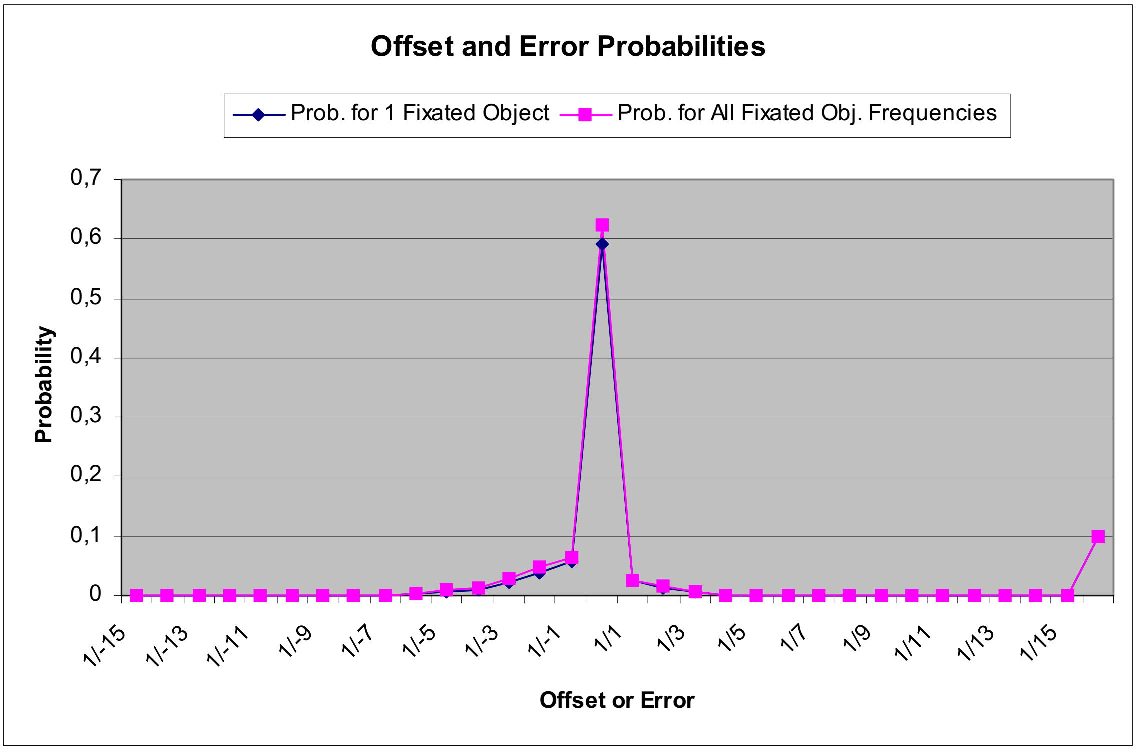


Figure 51. Probabilities of time offsets, for both a single object and for multiple objects under the gaze point.

Since a video frame covers 100 ms and the object log entries have an average inter-frame rate of 28.6 ms, it can be considered that a log entry falls within an acceptable range if it is within 100/28.6 = 3.50 entries of the video transcript within the sorted log file. There are 3780 transcribed video frames. There are 3072 objects log entries that fall within +/- 3 frames of the corresponding video frame time. Hence the overall temporal accuracy of object logging might be said to be 3072/3780 = ~81%. Note that this follows from the sampling rate represented by the captured video file of 100 msec per sample (frame), acting as an accurate sample rate only for sampled data periods greater than or equal to 200 msec. +/- 3 frames amounts to a 6 frame interval, equivalent to 200 msec but rounded down to the integer 6 multiple of average log entry durations, thereby representing a time span of 6 x 28.6 = 171.6 msec. In effect, interpretation of the video frame capture rate in terms of corresponding object log entries within +/- 3 log entries, functions as a temporal filter on the object log data, filtering out small scale variations in line with the temporal resolution of video frame grabbing.

5.5.2 Temporal Precision

Temporal precision has not been derived by considering a series of measurements over different trials. However, unlike the case of spatial precision, since the accuracy as analysed above is derived from system characteristics, it is a reasonable assumption that this represents a good characterisation of temporal precision across many stimuli. Further testing as described in this report will be necessary to verify this assumption.

5.6 Additional Sources of Error Qualify the Accuracy Models

The probabilistic accuracy models proposed above use the overall count of object log entries. This count includes a number of errors that are attributable neither to time variations nor spatial ambiguities. Instead, these are errors either within the HFE system or made by the human analyst in interpreting the video data. HFE system errors include rendering errors, collision detection errors and raytracing errors, as discussed earlier in this document. Human errors include wrong interpretations of fixated objects and saccades due to fatigue, pixelation, distance, or a dot representing a gaze point that is growing over time and obscuring what is behind it. The overall error count is broken down into percentages by type in Figure 52. Note that raytracing errors could most likely be fixed with further software development, and collision detection errors could be eliminated by careful construction of collision volumes to match as closely as possible their corresponding graphical meshes.

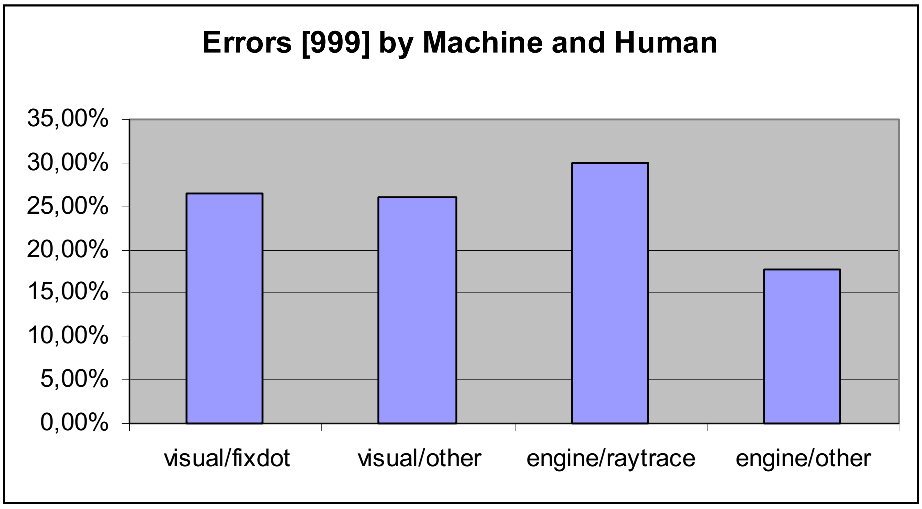


Fig. 52. The distribution of errors within the total of category 999. The total amount of errors includes 376 occasions of one gazed object and also if more than one object has been included due fixation size (ratio of 1 cm). In total, on 528 occasions

6 Conclusion

The verification study reported in this document has shown that the integrated HFE/Tobii eyetracking system performs well, with the accuracies described above. Such accuracy figures are well within what is regarded as constituting a useful system, especially for characterising gaze behaviour and variations in gaze behaviour for statistically significant numbers of test subjects. The availability of this system makes it feasible to consider experiments using large numbers of subjects and long test sessions, generating data that would involve impractical amounts of time to analyse if done purely manually. The system is now regarded as being a highly valuable tool suitable for ongoing use in such empirical studies.

7 Acknowledgements

The primary author would like to thank Alf Svensson for discussions and advice regarding helicopter flight operations, VFR and IFR, as well as many engaging conversations and much support over the years.

References

Alfredson, J. (2001). *Aspects of situational awareness and its measures in an aircraft simulation context*. LiU-Tek-Lic-20001:2, Linköping University, Sweden.

Bartle, R. (1996). *Hearts, Clubs, Diamonds, Spades: Players who suit MUDS* http://www.mud.co.uk/richard/hcds.htm

Castor, M., Hanson, E., Svensson, E., Nählinder, S., LeBlaye, P., MacLeod, I., Wright, N., Alfredson, J., Ågren, L., Berggren, P., Juppet, V., & Ohlsson, K. (2003). *GARTEUR Handbook of Mental Workload Measurement. Group of Aeronautical Research and Technology in Europe.* GARTEUR Technical Paper 145. FOI-S--0989--SE.

Duchowski, A. (2003). *Eye Tracking Methodology – Theory and Practice*. Springer Verlag. London.

Eysenck, M. (1998/swedish: 2000). *Psychology –an intergrated perspective*, Student Litterature, Lund, Sweden.

Findlay, J., & Gilchrist, I. (2003). *Active Vision: The Psychology of Looking and Seeing*, Oxford University Press.

Hasewinkel, H., & Lindoff, J. (2002). *Modelling, simulation and COTS: Using commercial PC games in command and control studies*. In proceedings of the first Swedish-American workshop on modeling and simulation, SAWMAS. Orlando, Fl, 8p. FOI-S--0703--SE.

Hedstrom, J., & Lindahl, B. (2003). *Användning av förstapersonspel vid militär simulering av strid i bebyggelse*, LiTH-ISY-EX-3337-2003.

JAA Administrative & Guidance Material, Section Four: Operations, Part Three: Temporary Guidance Leaflet. LEAFLET NO. 34: NIGHT VISION IMAGING SYSTEM (NVIS) OPERATIONS. Attachment 2, Ground Training –Areas of Instruction. Section 4/Part 3 (JAR-OPS) 01.06.03.

Kylesten, B., & Söderberg, H. (2001). *Erfarenheter av att använda kommersiella förstapersonsspel för taktisk ledning- och beslutsträning*. FOI-R--0228--SE.

Kylesten, B. (2004). *Kognitiv Lägespresentation*. FOI-R--1436--SE.

Lindoff, J. & Hasewinkel, H. (2004). *Warefare training with game consoles – Training without getting your hands dirty*. In proceedings of the second Swedish-American workshop on modeling and simulation, SAWMAS. pp. 83-88. FOI-S--1309--SE.

Nyberg, L. (2002). *Cognitive Neuroscience –studies of the collaboration between brain activity and mental processes,* Previc. F. 1998. *The Neuropsychology of 3-D Space*, (Brooks Air Force Base).

Rencrantz, C. (2003). *Gamebased training and education. Testing the usefulness of gameconsols in the Swedish Armed Forces*. FOI-R--1086--SE.

Sennersten, C. (2004). *Eye movements in an Action Game Tutorial*, Masters Thesis, Department of Cognitive Science, Lund University, Sweden.

Svensson, E., Angelborg-Thanderz, M., Sjöberg, L., & Olsson, S. (1997). *Information complexity - mental workload and performance in combat aircraft*. Ergonomics, 40 (3), p. 362-380.

Tobii Technology (2005). *User Manual: Tobii Eye Tracker and ClearView analysis software.*

Wikberg, P., Hasewinkel, H., Lindoff, J., Eriksson, L., Stjernberger, J., & Persson, A. (2003). *Kommersiella PC-spel som försöks- och övningsmiljö: Prövning av jägarbataljons ledningsmetod vid tillgång till realtidsinformation*. FOI Metodrapport, FOI-R--0989--SE.