Gaze Direction Analysis for the Investigation of Presence in Immersive Virtual Environments

Joel Jordan

Thesis submitted for the degree of
Doctor of Philosophy
of
UCL

Department of Computer Science
University College London
University of London

2010

I, Joel Jordan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

I would like to thank Mel Slater for his continual and enduring support. Without his help and advice it would have been impossible to complete, and I am extremely grateful for him having supervised me and steered me through the long PhD process. I truly appreciate his steadfast efforts to help me improve my attention to detail, as well as his dry sense of humour. I am also very grateful to my second supervisor, Anthony Steed, who was always ready to give me his time, helping me to clarify my objectives and encouraging me on the way. From the Virtual Environments and Computer Graphics group, I am grateful to many for their friendship and encouragement, but especially Manuel Oliveira, Insu Yu, Celine Loscos, and Angus Antley.

There have been a number of people other than my two supervisors who have donated some of their time, energy and expertise to either suggest an idea, teach me something, provide feedback, or help my work in some other way. It would impossible to quantify each of their contributions, but I am grateful to Daniel Alexander, David Barber, Ashwin Beeharee, Fred Brooks, Richard Chandler, Christopher Clack, Stephen Ellis, Manuel Oliveira, Paul Samet, Angela Sasse, John Shawe-Taylor, and David Swapp.

I would also like to express my gratitude to the staff in the Computer Science Department at UCL. One group that comes to mind are those in the Technical Support Group, but in particular Denis Timm who would always promptly help in a technical emergency. Denis, knowing his craft well and being able to locate exotic (or ‘old school’) tools was vital to my work on several occasions. Just as important to the success of my work was John Andrews, to whom I am also grateful. He is not only kind and welcoming to anyone needing help, but extremely knowledgable and conversant with the UCL computer network. For their help, particularly with the logistics in having equipment sent back and forth across the Atlantic, I thank JJ Giwa and Patricia Fenoy, Dawn Bailey and Wendy Richards, as well as all the remaining administrative staff.

I shall never forget the kindness shown to me by each and every one of the people named above all of whom helped me in some way or another, not just to complete the work itself, but more importantly, to encourage me along the journey.

I have also enjoyed continual support from my friends and my family. I thank in particular my father, Ed, who helped in proof-reading, my mother, Ilda, my brother Mark, my parents-in-law Roger and Camilla Harley, and our close friends Siam and Lisa Bhayro, and Jonathan Kitchen.

Finally, I owe a huge debt of gratitude to my wonderful wife, Ruth. She has been unwaveringly patient, stoic, supportive, and hard-working for such a long time. I imagine that it shall take quite a

while to repay this debt, but every moment that I can be a blessing to her is a moment well spent.

Several years before I began my research I coincidentally stumbled upon the immersive-virtual-environment researcher Fred Brooks’ famous book ‘The Mythical Man-Month’. (It was later on that I happened to briefly meet the Professor on two occasions whilst doing my research. This was a great pleasure for me, as he and his wife are such a friendly and warm couple.) As I read the preface I was surprised to see the words he signed off with, and I respectfully take my hat off to him and humbly echo his words, Soli Deo gloria — To God alone be glory.

Abstract

The aim of this thesis is to investigate the idea that the direction of gaze may be used as a device to detect a sense-of-presence in Immersive Virtual Environments (IVE) in some contexts. This method would enable the evaluation of presence in IVEs without the use of auxiliary systems or devices in many cases, as the necessary elements are already intrinsic to IVE systems. This method which involves recording the direction of gaze as a sequence of fixated regions of space (a ’gaze scanpattern’), is evaluated in three ways:

Firstly, the method is used to probe an open question within the presence-research field, that of the existence of minimal cues. This theory proposes that there is a cue threshold above which a sense of presence is attained, and after which further cues do not necessarily increase the sense of presence. An experiment is performed where visual cues are slowly introduced to eventually form a complete IVE designed to provoke a stress response. Evidence supporting the existence of minimal cues and the potential use of the gaze scanpattern is presented, by showing a simultaneous and sudden change in gaze-scanpattern entropy and a physiological (stress) response indicating presence before the environment is fully complete.

Secondly, the method is shown to be useful even without eye-tracking for estimating the direction of gaze, using only head-tracking. An evaluation is performed of the entropy and the mutual information between data from eye-tracking and head-tracking, and head-tracking alone. From this, along with a re-analysis of the aforementioned experiment using head-tracking data alone, it is concluded that head-tracking data alone can provide a useful approximation of the direction of gaze.

Thirdly, the method is used to make quantitative comparisons between IVEs and a real world environment, demonstrating that IVE gaze appears similar to that when viewing the real world when there are sufficient visual cues. This is an important feature of the method as it shows that the gaze scanpattern is founded on a real-world response, and that it is to some degree idiosyncratic with respect to an environment. It also demonstrates that improvements in visual quality do not always lead to a more realistic response in terms of gaze behaviour, as predicted by the minimal cues theory.

This novel technique should prove useful as a tool for IVE presence-research. At the moment, there is no known objective presence indicator that can be consistently applied to such a wide range of environments. The use of gaze scanpatterns would be useful for the construction of IVEs as they can be created effectively and objectively. Effectively, because a minimal level of visual cues required to support a specific perception of an IVE may be tested for. Since the efficacy of visual elements in promoting presence can be objectively tested, by using these methods in some cases, designers of IVEs will no longer have to rely on subjective assessment alone in order to make decisions as to what attributes might induce a sense of presence. In addition, it should be possible, in some cases, to present the required cues with less design-effort and computation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The term *virtual environment* (VE) is generally understood to mean an environment that is described in three dimensions to be presented on a computer display. They are perhaps most commonly encountered in computer games, but are also found in research, simulation, training and design — particularly architectural design. Even the most technologically pedestrian of us are likely to have stumbled across them in the sensationalist news reports that occasionally surface[1] [2] [3].

When we use the more qualified term *Immersive Virtual Environment* (IVE) in this thesis we refer to VEs that are displayed using equipment that produces an ego-centric view, allowing the view position and direction to be changed by moving the head and body in a natural way (as first theorised by Sutherland, 1965.) Today this may be achieved through the use of a six degrees-of-freedom spatial tracker and a stereoscopic head-mounted display (HMD). When using an IVE one can attain a sense that one is actually present within the virtual environment that is displayed, and 'presence' has in fact been identified as a key feature for their general use (Held and Durlach, 1992), or even their defining factor in terms of the human experience (Steuer, 1992). This 'sense of presence' in an IVE is the ultimate focus of this thesis.

In order that we may operationalise the 'sense of presence' for the purposes of investigation, we define it as something that occurs when a person responds to an IVE as though it were a real environment. This has been called 'Response-as-if-real', abbreviated RAIR (Slater, 2009). There are of course numerous ways in which such responses can occur and be measured, and in this thesis we shall investigate the sense of presence strictly under this definition. We shall do this by computing measures from the path that the direction of gaze follows, and compare this to the actual or expected gaze path in a real environment. The path that gaze follows has been referred to as a searchpath (Choi et al., 1995), scanpath (Noton and Stark, 1971), or scanpattern (Henderson, 2003) depending upon the context. We shall adopt the term *scanpattern*, the use of which has been encouraged as it is "theory neutral" (Henderson, 2003).

1.1 Research Motivation

In order to investigate presence in an IVE we require some way of measuring it. Many methods for measuring presence have been developed, and the most widely used has been the administration of ques-

---

[1] "Second Life affair leads to real life divorce", Guardian (online), UK, 13th November 2008.

[2] "On the trail of Manhunt 2", "most controversial video game in history", BBC News (online), UK, 8th February 2008.

[3] "Whitehall defends 'fantasy world"', BBC News (online), UK, 19th March 2009.

tionnaires. However, in the last few years particularly, sole reliability upon questionnaires has been the subject of criticism. This is in part because they are (of course) subjective, that there has been little standardisation or agreement over the questions used, and there has been confusion regarding how responses ought to be analysed (Slater, 1999, 2004; Slater and Garau, 2007; Gardner and Martin, 2007). Other methods have therefore appeared, particularly more objective ones that rely upon the RAIR concept. Unfortunately the vast majority of the objective measures or indicators rely upon methods that are not easily applied across a wide range of environments. For instance, stress may be measured using relatively simple physiological recording. These have been employed to indicate when a person feels present in an IVE that would cause a stress-response if the IVE was their actual (real) environment. Indeed, this indicator is one that we ourselves use in this thesis (Experiment I, Chapter 4). However, clearly, not all IVEs should be or can be designed to induce a stress-response, so this method of presence detection is limited in its application domain. We describe a number of such objective presence measures, including that one, in Section 2.4.4.

Perhaps the ideal way to objectively measure presence would be to somehow look within a person's brain to 'see' the change take place as they leave reality and enter the virtual world. Certainly the combination of presence-in-an-IVE and fMRI (functional Magnetic Resonance Imaging) has been considered together (Hoffman et al., 2003). Such research is still in its infancy though, and it currently requires highly restricted movement, and utilises expensive and complex equipment.

For the above reasons, it seems that there is scope for new methods that could be applied to a wide range of contexts, that are objective, that rely upon as few technical devices as possible, and that are simple to understand and operate.

The majority of IVEs are primarily visually based, and interaction with the environment is achieved via movements of the body that change the displayed view that the user can see (that is, the user moves their actual head and body to look around the virtual environment). Since such interactions are grounded in the visual perception of one's environment, a person's gaze scanpattern, which has been shown to reflect perception in past studies (Buswell, 1935; Yarbus, 1967; Ellis and Stark, 1978, 1979), would seem a potential avenue for investigating what an IVE observer perceives over a particular period of time. Although movement of the direction of gaze has been considered (Renaud et al., 2007) before in a presence-research context, our approach considers not just how an IVE observer views the environment but ties this to *what* they look at, and *what they perceive*.

In order to develop and test this approach we require a problem to train it upon. One of the greatest problems in the field of IVEs is that of the efficient production and presentation of environments. This is due to the complexity of the real-world environments that they endeavour to describe and mimic. As an example, consider the nature of a typical room, which as one moves closer to almost any object within it one finds there is further visual detail. Such depth of detail is normally difficult if not impossible to reproduce.

It is therefore fortunate that the brain is easily fooled in such a way that a person may feel as though present in an IVE with far less detail than that which is available in a real-world environment as explained by Stark (1995). He states that when using an IVE an illusion is created that the environment is complete, and this has lead to the idea that there must be a minimal set(s) of cues required to represent each possible environment (Slater, 2002). It would be of great interest and practical importance to determine *when* such necessary minimal cues are present in an IVE. This would not only help us develop our basic understanding of what is really going on when IVEs are viewed and interacted with, but enable us to be pro-active in the design of virtual environments by ensuring we include the most important aspects of them for presence. This can then inform future development of technologies and design methods, ideally leading to optimal presence induction.

Therefore, in this work we endeavour to provide evidence to link gaze, perception, and presence (under the operational definition RAIR). We investigate the thesis that the path of one's gaze may be used as a perceptual response indicator of RAIR presence in an IVE, that the gaze scanpattern actually relates to behaviour in the real world, and that there is a demonstrable minimal visual cues threshold.

1.2 Research Scope

As described above, an IVE is a virtual environment displayed using immersive[4] technologies. In order to carry out any scientific investigation it is necessary to limit the scope appropriately so as to make the objectives achievable in a finite length of time. To this end, the research described in this thesis is constrained as follows:

- We shall only consider visual modes of display. In effect, this means that all stimuli will be presented as images on graphically capable screens. This is achieved solely with head-mounted displays (HMDs).

- Throughout this thesis, our operational definition of 'being present' in an IVE is that a person responds as though the IVE were a real place (RAIR, see Section 1). While there are many potential responses, which range from autonomic reflexes to highly-cognitive and volitional ones, we will only concern ourselves with a limited set of measures of a person's gaze response.

- Concerning the detection of the 'sense of presence' in an IVE: we will not be considering real-time instantaneous detection of a person's state, but rather the sense of presence that an IVE has the potential to induce, which is expressed by observers (experimental subjects) in general over some period of time.

- With respect to the actual virtual environments that have been used for experiments, there have also been inevitable restrictions made on their design. For instance, investigating both indoor and outdoor scenes in addition to the objectives set would have expanded this thesis' scope dramatically. In light of this we have limited the research scope such that all environments, virtual or otherwise, are 'indoors' (where such a term is meaningful), and have been spatially constrained in their dimensions, being set within a volume of ten metres cubed ($10m^3$).

---

[4] Immersion is discussed in Section 2.3, and for this work we are using the definition that has been called 'system immersion' according to Slater (1999).

1.3 Research Questions

In order to carry out our investigation we set out the following research questions:

Question 1: Can the gaze scanpattern be used to detect RAIR in IVE users?

From past studies it has been found that the point of gaze follows repetitive and idiosyncratic paths over a stimulus (image) (Buswell, 1935; Yarbus, 1967; Noton and Stark, 1971; Choi et al., 1995). These gaze paths have not only been demonstrated to be idiosyncratic with respect to the particular stimulus, but even to the specific percept that they evoke (Ellis and Stark, 1978, 1979).

In a separate discussion Gregory (1977) suggests that when viewing a stimulus, if the mind does not obtain a superior hypothesis as to what is perceived it will not settle upon any perception of the stimulus. In this case, the above findings would suggest that the lack of a stable percept would result in gaze paths that would also be less stable, in terms of them being less repetitive and idiosyncratic. We therefore conjecture that the path of one's gaze would then have a lower entropy (all other things being equal.)

We investigate whether these ideas can be leveraged to discern between a person viewing a meaningful IVE and one that is difficult to perceive as meaningful. If a scene is not viewed as being a meaningful environment, then this will not be (as) efficacious for inducing presence.

Because we normally perceive our everyday environments as being meaningful we shall expect concomitant gaze paths to be repetitive and idiosyncratic (having a relatively low gaze entropy). But we may simulate a non-meaningful scene using random polygons in three-dimensional space (which we conjecture will lead to relatively higher gaze entropy.) Thus we may be able to use this simulated scene as a control, enabling us to test whether an IVE observer perceives a scene as meaningful or not. This would constitute another novel test for RAIR presence.

This question is therefore investigated in the first experiment, which is described in Chapter 4.

Question 2: Does a visual cues threshold for the inducement of presence exist in the context of IVEs?

In Chapter 2 (Section 2.4) we discuss in-depth the notion of an IVE user feeling present in a virtual environment. We take the view that the state of being present in the virtual environment is binary, that is, mutually exclusive where an observer perceives their immediate environment to be either one *or another*. This question asks whether we can find evidence for this view, by demonstrating a visual cues threshold above which IVE presence suddenly commences. This threshold was referred to in the presence research literature as occurring when 'minimal visual cues' were provided (Slater, 2002). We utilise the gaze scanpattern (as well as a pre-established presence measure) in an attempt to show evidence for this. This research question is also investigated in the first experiment (Chapter 4).

Question 3: Could our gaze-scanpattern methodology be useful, given only an approximation of the line-of-sight?

In Experiment I, we utilise both eye-tracker and head-tracker data. The latter is generally available when using standard IVE systems. Eye-tracking devices are, however, notoriously difficult to use (Schnipke and Todd, 2000) and not always available. This question is therefore very much a practical as well as theoretic one, as we would like to obviate the need for eye-tracking. It is addressed in Chapter 5.

Question 4: Does the gaze scanpattern in an IVE correlate with that of the real world, when present?

It may not at first be apparent why we are especially interested in the 'real world' when our specific field of interest is the IVE. However, this is where we have largely moulded our perceptual abilities, and indeed, IVEs are designed specifically to reproduce stimuli as we would sense them in the real world and as we would interact with them in the real world. We also define presence as occurring when someone responds to an IVE as though it were real. Thus, rather than take it for granted, we wish to investigate whether there is further evidence that we actually respond to IVE content in the same way that we do when viewing real-world content. This question is addressed in Chapter 6.

Question 5: If visual cues are provided over and above the minimal visual cues will they affect the gaze scanpattern, and if so would this be indicative of a greater (or perhaps lesser?) presence response?

Much work is done with the aim of enhancing virtual environments by increasing the detail of the content. This is largely carried out from two complementary perspectives, firstly by adding content through the laborious modelling of objects by further refining geometry and textures, and secondly through the development of more complex computational rendering processes that result in increased true-to-life realism (such as pre-computing static shadows or caustics.) This question pertains to the latter issue which sometimes appears to be a short-cut to enhancing a virtual environment. This is because improving rendering enhances all the content displayed, but editing objects is laborious, needing work to be done on an object-by-object basis.

Previous studies have investigated the effects of visual quality on the sense of presence; Zimmons and Panter (2003) and Mania and Robinson (2004) both tested for differences between flat-shaded and radiosity-computed IVEs. Zimmons and Panter (2003) also tested a wireframe IVE and used two texture resolutions. In both papers, no significant difference between conditions was found. Mania and Robinson used a presence questionnaire to detect differences in responses to their experimental conditions. Zimmons and Panter used three types of response measures (a presence questionnaire, skin conductance and heart rate), and employed a 'pit room' that included a sheer drop designed to elicit a stress-response.

According to the findings of Zimmons and Panter (2003) and Mania and Robinson (2004) we might predict little difference between our conditions. However, 'pit room' scenarios as used by Zimmons and Panter lead subjects to focus upon a specific characteristic of the IVE, the stress-inducing depth cues, and this could distract them and prevent other responses that might normally be detectable. In this sense, the results of Mania and Robinson (2004) allow us to generalise better to other IVEs, including ones that are not stress-inducing. The study of Mania and Robinson

(2004) relied upon a presence questionnaire alone though, which is particularly subjective and administered post-hoc. We shall therefore look at this issue once more, using a stress-neutral environment and using the path of the direction of gaze to produce an in-situ, objective response measure.

In a more recent study concerning visual rendering quality (also using the ‘pit room’ scenario) Slater et al. (2009a) did find significant differences using presence questionnaires, skin conductance and heart rate. But their conditions were different in that they were comparing IVEs with and without *dynamic visual effects* - dynamic shadows and reflections. Shadows at least, are known visual depth cues indicating the relationship between objects (Hu et al., 2002; Ware, 2008), and they have also been found to increase the sense of presence seemingly independently of this (Slater et al., 1995). It is therefore not so surprising that the stress-responses measured were greater when dynamic shadows were shown in a ‘pit room’, if only because a dynamic and interactive aspect is introduced, fundamentally changing the experience rather than the quality of the rendered environment only. We will be comparing environments that do not include such dynamic visual effects.

This question is addressed in Chapter 6.

1.4 Research Contributions

The research presented in this thesis makes several contributions to the field of Immersive Virtual Environments, which we list under the headings: Methods and Knowledge.

- Methods

    (i) A method using the gaze scanpattern that can provide an indicator of whether a minimal cues threshold has been exceeded in an IVE, by detecting change in the perception of a presentation.

    (ii) A method for comparing whether an IVE appears to adequately match some reference environment by comparing gaze scanpattern responses.

    Together these methods provide the essential abilities for presence gaze-scanpattern-response testing, which may be used to test the potential of specific IVEs to induce presence. We expect this will be useful to carry out future investigations into the relationships between visual cues in a virtual environment and their impact in the production of presence invoking IVEs. This could lead us further toward a more fundamental understanding of this key issue in IVEs.

- Knowledge

    (i) That the analysis of the gaze scanpattern can provide information regarding the potential of an IVE to induce a sense of presence, under our operational definition of presence (Section 1.2).

(ii) Evidence that the gaze scanpattern behaviour *within IVEs* is not isolated within that context, but that it also correlates with gaze over real-world content, implying that gaze is indeed a valid tool for studying presence under our operational definition.

(iii) Scientific evidence for the existence of IVE minimal visual cues, which have been referred to in research literature but have not until now been demonstrated (Slater, 2002; IJsselsteijn, 2002; Mania et al., 2005).

(iv) Confirmatory evidence that arbitrary visual enhancements to IVEs do not necessarily impact our visual perception of the environments in terms of gaze scanpatterns. This encourages further thought and investigation for the benefit of developing models of the relationships between visual enhancements and their perceptual effects. This would aid our ability to know which enhancements will be effectual when the objective is to produce a presence invoking IVE.

(v) Evidence that non-meaningful IVEs (such as those we create for Experiment I) can consistently result in greater gaze scanpattern entropy than when viewing meaningful IVEs. The generalisation of this finding could suggest numerous avenues for future research where a statistical evaluation of gaze over an image, scene or environment is used.

1.5 Thesis Structure

This thesis is organised to provide a continual narrative to be followed, although it is broken up into digestible and focused chunks. There are 7 chapters that are set out as follows.

After this initial introduction, Chapter 2 provides the necessary background to understand where in the scientific fields this research is seated and relevant. These fields include Virtual Environments, Presence, and Visual Perception.

Chapter 3 is used to develop the approach that is used to probe our research questions, also covering various methods that are either used to carry out the research or that are contextually relevant to the methods used. The methods include physiological measures, gaze scanpattern analysis, and mathematical methods.

Chapter 4 details the first experiment, which is termed ‘Experiment I’ through the remainder of this document. Chapter 5 contains the analysis and evaluation of gaze based on eye and head-tracking, versus head-tracking alone. The results in this chapter have bearing on the way that the following experiment (found in the following chapter) is carried out. Chapter 6 contains Experiment II.

The final Chapter 7 contains discussion of the research work, contextualising it within the domain of this work and within the areas of study that are described in Chapter 2. It is here that we draw the final conclusions from the previous chapters, and provide a summary of the work and an assessment of it as a whole. We also comment on potential future work leading from this thesis.

Chapter 2

Background

2.1 Virtual Environments

In the field of computer science a Virtual Environment (VE) is an artificial, computer generated representation of an environment. The heart of a VE consists of the data describing it and a method of interpreting that data (the latter being made up of computer software and hardware.) Of course, what makes a VE an environment though is the experience of it. To create this experience a system is required to interpret and present the data, and will accept input(s) from a person in order for them to explore the environment. Often, these basic requirements are exceeded to provide additional application specific functions that might be necessary.

Virtual environments come in many forms. Some researchers regard VEs from a particular perspective, but the term ‘virtual environment’ is also often used in a general sense. If we take a broad view of the term, then VEs range from those that are textually based e.g. MUD (Multi-User Dungeon), and MOO (Multi-user Object Oriented) environments (Hand, 1994; Schneider and Godard, 1996) to more graphical ones such as Club Caribe (Morningstar and Farmer, 1991), ActiveWorlds [1], or Second Life [2]. Each of these allows a number of users to interact with the virtual world (and each other) using typed text, the graphical ones rendering representations of the users (avatars) and spatial renderings of their environment. The environments are often segregated into various areas, typically thought of as places (e.g. chat rooms), and hence the inclusion of them *all* under the umbrella term virtual environments. Finally there are those VEs that are presented via an ego-centric three-dimensional projection system, that are commonly thought of as being Virtual Reality (VR) systems. These are rendered using graphical hardware that projects from the three-dimensionally described environment into two image streams that constitute a stereoscopic view.

2.2 Virtual Reality

Jaron Lanier popularised the term ‘virtual reality’ to use in the business that he co-founded in 1984 (VPL Research) [3]. The term has since had numerous interpretations, and its meaning has been investigated by

---

[1] http://www.activeworlds.com , 2010

[2] http://www.secondlife.com , 2010

[3] http://www.jaronlanier.com/general.html , 2010

Steuer (1992). The term ‘Virtual Reality’ has been used in a very general sense (e.g. Hand, 1994) to describe almost any virtual environment. In this document though, we shall with few exceptions constrain our considerations to the types that are spatially three-dimensional and egocentric, that are presented with software and hardware systems that attempt to mimic physical reality. In this definition we also include systems that utilise desktop hardware to present two-dimensional and three-dimensional renderings (known as $2\frac{1}{2}$D and stereoscopic respectively) of three-dimensional environments. These are considered as being particularly limited [4].

The systems used to present and interact with a virtual environment can vary widely. The system hardware often consists of a number of input devices and one or more displays, along with the central processing unit(s). Next, we will briefly consider some of the common elements of a typical VR system. These are of course vital, however the main concern of this work regards the virtual environment’s content and form.

2.2.1 Displays

Perhaps it is for both fiscal and technological reasons that displays tend to be provided for just one of the human senses rather than multiple senses. There is undoubtedly a precedence by which the frequency of each type of display (with respect to the sense it addresses) is to be found. It is probably not surprising that the two most commonly found types of displays are graphical and audio. In fact, devices that address other senses are few and far between, though commercially available tactile and haptic systems are widely available, although limited in scope.

Graphical Displays

Graphical displays come in numerous forms. Although head-mounted displays (HMDs) (Sutherland, 1965) are the most often thought of devices in VR, there are others. The HMD normally consists of two miniature screens mounted within a helmet, so that the wearer’s eyes are presented with a (potentially) stereoscopic image, as generated by computer graphic hardware. Other graphical display systems include the recursively named CAVE (CAVE Automatic Virtual Environment) (Cruz-Neira et al., 1993) systems that consist of several projection screens arranged in a cube shape, allowing a user to stand *within* the display. Apart from the more esoteric graphical display systems such as the BOOM®(Binocular Omni Orientation Monitor — Fakespace Inc.), there is the lowly desktop — which continues to play a part in research, sometimes being used to act as a less immersive point of reference.

Many of these displays are capable of presenting apparent 3D (as opposed to true volumetric 3D) images, by displaying two different images, one to each eye. Technically there are various methods of achieving this, for example by displaying the two images using line-sequential, frame sequential, field sequential (interleaved), side-by-side, top-and-bottom, and through the use of polarized filters, though the hardware is likely to determine which techniques are actually used.

If the image is to be updated, being slaved to head movements (such as with the HMD or CAVE, also see Section 2.2.2 regarding tracking) then for correct perspective a 6 degree of freedom (6DOF) tracker is required, which will measure both direction and position of the head. A three degrees of

---

[4] Volumetric systems are not discussed within this thesis.

freedom (3DOF) tracker could also be used providing a less realistic experience particularly due to the lack of motion parallax.

Given a particular type of display, then perhaps the three most important characteristics are its colour gamut, resolution, field of view, and refresh rate. Graphical hardware providing 24-bit colour is now common, and although the display devices themselves may not always have the ability to display every hue and level of luminosity, most displays provide sufficient colour resolution for general purpose use (May and Badcock, 2002).

In terms of resolution, virtual reality hardware pushes the boundaries of current technology. In a CAVE system, multiple high resolution displays are ideal. For HMDs the issue is not so much the number of displays multiplied by the number of pixels available, but the number of pixels that may be squeezed in to their miniature screens, and the optics necessary to display them correctly. One might say that with a CAVE one is concerned with (logistics and) quantity, and with a HMD, (high-technology and) quality. As regards the number of pixels, CAVEs thus are near the leading edge of what a desktop display and projector can provide multiplied by the number of screens used, and HMDs have a resolution that is similar to a typical desktop monitor[5].

The ‘field of view’ is a term used to describe the angle subtended by the screen, with respect to the viewer. For a CAVE type system, the environment can be fully surrounding (Cruz-Neira et al., 1993), limited only by polarised or shutterglasses that are sometimes used to view stereoscopic scenes. Head-mounted displays however, are mostly limited in this respect, providing a field-of-view from as little as 20 degrees, to around 180 degrees with more expensive hardware. The field-of-view plays an important role in determining how immersive a virtual environment might be (see Section 2.4.3.)

Finally, the refresh rate of a display may become an issue when displaying stereoscopic images. This can occur when a method such as ‘frame sequential’ stereoscopy is used, so that alternate display frames are sent to alternate eyes, creating a sense of depth in the image. In using this method the refresh rate is thus halved, and this can produce a flickery image if the overall refresh rate is not fast enough. Ideally then, distinct channels rather than a temporally multiplexed channel should provide the separate images required for a stereoscopic display.

Auditory Displays

Auditory displays are probably the second most commonly used display. They are usually both directional and spatial, and are implemented with either stereo or surround-sound systems. While CAVE systems will tend to have a number of speakers, a HMD may be used with headphones, making the latter a relatively compact system. Of course, one is not restricted to utilise headphones or speakers in either case. Although it may be thought that the high-technology of a multi-channel, surround sound, speaker system would provide the most immersive experience, Shilling and Shinn-Cunningham (2002) suggest that the use of headphones is ideal because it “reproduces azimuth, elevation, and distance...and offers... the greatest amount of control”. Shilling and Shinn-Cunningham (2002) make an excellent argument for the importance of audio in virtual environments, noting that they are useful in creating an ambiance, or

---

[5] As of July 2010, although both technologies are currently making rapid advances.

providing audio cues to indicate the occurrence of events. A particularly insightful point that they make, is that audio cues can act as a substitute for the tactile sense when collisions occur.

Haptic and Tactile Feedback Devices

Haptic and tactile feedback devices provide a user with a sense of touch within a VE. The word haptic comes from the Greek 'haptesthai', which means *to grasp*; the familiar word 'tactile' is from the Latin 'tactilis', meaning *to touch*. Such devices often provide a complementary combination of input device and output display, detecting the position of and forces from the body (typically the fingers or hand) and producing reactive forces, as the body comes into 'contact' with objects in the virtual environment. The number of degrees of freedom used is variable across devices, and is not necessarily asymmetric with respect to the input and output. Thus, a force could be generated with 3 degrees of freedom (yaw, pitch, magnitude of force) while sensing the position of the point-of-contact, as manipulated by the user, with 6 degrees of freedom using the three-dimensional position and the direction (yaw, pitch, roll). Their main characteristics are therefore degrees of freedom for input and output respectively; the range, resolution and accuracy of the input and outputs; update rates (e.g. rate of tracking the point-of-contact) and latencies; and the physical extents within which the device can operate (i.e. the device's reach). A rather technical description of how haptic effects may be generated by virtual environment software is to be found in Basdogan and Srinivasan (2002).

Although not commonly found, these devices are appearing more frequently in research laboratories. The SensAble PHANToM Desktop is one such haptic device (http://www.sensable.com , 2009), whereas the Cyberglove is a tactile device Cyberglove [6].

2.2.2 Input devices

There are many input devices that may be used with VR, some having been specifically designed for this purpose such as the Cyberglove mentioned above. Simpler more commonly found devices are baseless trackballs, joysticks, and even one-handed keyboards such as the 'Twiddler' [7].

Devices that are held in three-dimensional space will often be spatially tracked with either a 3 or 6 degrees-of-freedom tracking device, (see below). Trackers support effects such as having a representation of the user's hand and/or arm displayed in the virtual environment. This can have a significant impact on the user's experience within the immersive virtual environments (Slater and Usoh, 1993b, 1994), as we shall explain (Section 2.3.2).

Spatial Trackers

Trackers, commonly attached to the head and hand, are used to track the user's body and/or input devices. Having 3 degrees of freedom (or 3DOF in shorthand) affords position in three-dimensional space, or direction (yaw, pitch, roll.) 6DOF however is the ideal, affording both position and direction of the tracked object.

It is usual for the head to be tracked to ensure that correctly rendered images are produced for a user (e.g. on a HMD). These data are also useful in providing the location/direction of the user for other

---

[6] http://www.cyberglovesystems.com , 2009

[7] Unfortunately the Twiddler may no longer be available. See HandyKey Corp, http://www.handykey.com , 2009.

purposes, such as for interaction (e.g. producing appropriate ambient sounds when entering different spaces) or for behavioural analysis.

The hand is often tracked to place some type of cursor device within the environment, but also to allow the display of the user's 'virtual' hand and arm, to aid the sense of feeling present within the environment by reinforcing proprioceptive senses (see Section 2.3.2.)

When the head or limbs are tracked, other physical locations may be roughly estimated from them such as the position of the torso. Alternatively, more complex systems are available that track many points simultaneously (Frey et al., 1996). Although some motion capture systems utilise bodily-worn markers to identify the position of the human limbs, there are now systems that aim to achieve tracking of the human body without the aid of markers (Fua et al., 2002).

The impact of tracking body movements has been investigated by Slater and Usoh (1993b, 1994); Slater et al. (1998), and has been shown to be of great importance in using immersive virtual environments in reinforcing the effect of being present in the virtual world.

Apart from the number of degrees of freedom measured, the other (main) characteristics of spatial trackers are their update rate (how often they produce a sample), latency, physical extents of the volume of space that may be tracked, resolution of the measurements, and their accuracy.

2.3 Immersion

2.3.1 The Meaning of Immersion

A commonly used term in the field of virtual environments is immersion. Immersion may be thought of as the extent to which a person is enveloped by the VE. This is a fairly generic definition, and there are differing views of what immersion is more specifically. As we shall be using the term throughout this thesis, we shall consider its more specific definition within the field of virtual environment research.

In their paper, (Slater and Wilbur, 1997) relate immersion directly to the enabling technology. For them, a more immersive system is one in which the technology provides a more compelling environment:

"Immersion is a description of a technology, and describes the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant." Slater and Wilbur (1997)

For instance, by noticeably improving on the resolution of a display, the level of immersion would be said to have increased, or a system that does not provide force-feedback would not be as immersive as the same system with a force-feedback capability. Technologies meant to provide for such illusions of (virtual) reality naturally have differing levels of capability, for instance the extent of fidelity, and these differences should be quantifiable. This objective view of immersion is of value by virtue of its applicability.

Smith et al. (1998) consider several examples of immersive environments from which we may draw conclusions: physical reality, virtual reality, and dreaming. From these, they propose that a person's senses should be the basis for the definition of immersion. In their description suppression occurs when stimuli from the physical environment are blocked out, such as when a dark environment is used to emphasise the contents of a display screen. Suppression of undesirable (conflicting) stimuli therefore can be used to facilitate the perception of one particular environment model over another. They describe the information derived from the senses as then being filtered according to expectancy and internal models to provide the most parsimonious interpretation of the stimuli. In this way their approach defines immersion as the extent to which senses are catered for, and this is somewhat similar to the argument of Slater and Wilbur although they consider immersion from the opposite side of the technology-sense boundary.

More recently, these two separate views have been unified to a degree in Slater et al. (2009b) where Slater et al. extends Slater's earlier definition to account for the capability of an IVE user to meaningfully interact with the IVE technology. The user's perception of a display, which by its very nature is an interpretation, and the display's gamut are noted as being imperfectly connected. Thus a display may be able to provide different images which are nevertheless indistinguishable to the user. Additionally, an external observer cannot infer what a user's perception of images are by extrapolating from the technical data concerning what is actually shown on a display. This is because the brain's interpretation of stimuli is idiosyncratic and learned through experience rather than through extrapolation alone.

Thus the level of immersion provided by a system is determined by the following "constraints": the ability of the technology, the ability of the senses, and perception — which is made through interpretation (see 2.5.4.) In practice, Slater (2009) explains that immersion may be considered as the extent that a specific technology can support sensorimotor contingencies (SC). As an example of an SC, we can think of the movement of the head in space, which should update the displayed images on a HMD. The SCs that are supported by some particular IVE system, he calls *valid sensorimotor actions*.

Using this further articulated definition, Slater has also developed the concept of immersive equivalence classes, as defined by the relation 'A can simulate B' over the set of IVE systems. This uses the idea that some IVEs could (or can) be used to simulate the use of another IVE system whilst they are used, as was in fact done in an earlier paper (Slater et al., 1994). This development enables the construction of a taxonomy for IVE systems, by which objective comparisons can be made.

In an alternative view to the above, Witmer & Singer defined immersion as:

"a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences". (Witmer and Singer, 1998)

This passage offers an entirely different perspective, because immersion is now framed as a psychological state. Immersion here is thus subjective and cannot be measured directly. As we shall see, a concept termed presence is largely concerned with the human experience of virtual reality systems and thus here the two concepts are confounded (presence is discussed later, in Section 2.4). Unfortunately, it seems unclear what benefit there is in using this definition when the term presence also describes the human experience. This is particularly true as by employing the previous definitions of immersion we can at least distinguish between the experience itself and the technology used to provide it, as well as use Slater's taxonomy.

Table 2.1: Adapted from “Technological Variables Influencing Telepresence” Steuer (1992)

Immersive Virtual Environment				
Vividness		Interactivity		
Breadth	Depth	Speed	Range	Mapping

Slater and Witmer and Singer therefore offer different perspectives on the term immersion, using it to describe aspects of an IVE system and experience that are not mutually exclusive. Part of this conflict is due to differences in terminology[8]. For the sake of clarity, within this thesis we shall use Slater and Wilbur’s (1997) definition of immersion, which is further developed and better delineated. Having made this decision, we shall separately consider the subjective experience that is mentioned in the above quote of Witmer and Singer, which must be defined and measured; as such we describe the phenomenon termed presence in Section 2.4.

2.3.2 Immersive Virtual Environments

Virtual Environments are presented in many forms, largely dependent upon the underlying hardware that is used. Perhaps the most classic example is the HMD (see Section 2.2.1), which is often coupled with a tracked glove used for interaction with the VE. This type and others that project content according to the location of the observer (such as the CAVE) are often described as providing Immersive Virtual Environments (IVEs) due to the high levels of immersion these systems support.

Perhaps it is because we humans can absorb such a vast amount of information using our eyes, and because such a large volume of our brain is devoted to visual processing, that IVE systems tend to emphasise the visual mode. However, the intention of the IVE is to engage the user with the virtual environment by supporting natural movement as input to the system, and responding by providing concomitant feedback to the senses. Slater and Usoh state that:

> “The degree of immersion is increased by adding additional, and consistent modalities, greater degree of body tracking, richer body representations, decreased lag between body movements and resulting changes in sensory data, and so on.” Slater and Usoh (1994)

Ideally then, IVE systems ought to be able to sense all movements (or attempts to move) made in the physical world, and display upon as many of the user’s senses as possible. Therefore a system is said to have a certain ‘breadth’ relating to the modalities that it caters for, and the term ‘depth’ concerns the fidelity provided by each mode of display (Steuer, 1992). Together, Steuer (1992) places breadth and depth under the term ‘vividness’. In addition to vividness, he defines ‘interactivity’, which comprises ‘speed’, ‘range’ and ‘mapping (see Table 2.1). The speed of a system relates to its response time, with faster response times bringing it close to real-time. The range of the system relates to the potential for interacting with the environment, that is, the extent to which this is supported. Mapping is concerned with how the “human [user] actions are connected to actions within a mediated [virtual] environment.”

---

[8] To help clarify, Slater (1999) put forward the qualified terms system immersion and immersive response

When experiencing an IVE, a person's senses and thus perceptual systems are presented with information from a local artificial source (i.e. one or more computers.) The intention is that their perception is at least in part fooled into believing that the person is (1), physically elsewhere, and (2), that there is no mediating device between them and the virtual environment. The latter is described as the illusion of non-mediation (Lombard and Ditton, 1997), and is important in the sense that the user should not be reminded that they are within an artificial environment, but rather only attend the 'virtual' cues suggested by the IVE. If successful, we obtain what has been termed "the (suspension of dis-)belief" (Slater and Usoh, 1993a). This can perhaps be explained as: the disregard of the knowledge that the apparent environment does not honestly depict the immediate physical environment. This then is the immersive virtual environment (IVE.)

2.4 Presence

Having defined immersion as the *enabling technology* of an IVE experience, which may be specified in an objective manner, we must also consider the subjective experience of the participant. This experience has been termed presence and is identified as the key concept of 'virtual reality' (Steuer, 1992); with immersion being a necessary but not sufficient condition to support it (Slater, 2002). Along with IVEs, it was teleoperation (perhaps more so) that suggested the importance of presence as a topic for study (Sheridan, 1992), which have many of the same presence characteristics as immersive virtual environments (Ellis, 1996). Teleoperation concerns the operation of remote devices (such as sub-sea robots), and introduced the concept of telepresence, that is, the sense that one is present at some remote site by artificially reproducing human-sensory stimuli originating from the remote environment (Minsky, 1980).

The importance of presence lies in its potential to enhance many technological innovations including: teleoperation, conferencing and communication, learning and training, therapy, simulation, entertainment, design, and exploration (Lombard and Ditton, 1997). Therefore, in the remainder of this chapter we shall consider the utility of presence, the effects that it leads to, some presence theory, and subsequently, presence determinants and measures.

From the earliest papers regarding presence, its definition seemed to elude researchers and ever since it seems that each field defines presence relative to its own needs. Indeed, Lombard and Ditton (1997) identified six types of presence, each of which may be classified as either physical or social, as pointed out by IJsselsteijn et al. (2001). The field of immersive virtual environments is not an exception, and the concept might well be qualified as spatial presence. In the next section we consider various ways of understanding and defining presence.

2.4.1 More than one type of Presence

At first one might think that the nature of presence is a simple affair, you either feel as though you are standing in a room looking at a computer display, or you feel as if you are in the virtual environment presented by an IVE system. Indeed one common term used to refer to presence is the sense of 'being there' (Heeter, 1992; Steuer, 1992). It is often used as a touchstone phrase to introduce the fundamental concept of presence or simply used as shorthand, but beyond this it appears to be of little use. Defining presence is certainly more complex than invoking this term. In Lombard and Ditton (1997), a total of six conceptualisations of presence (gleaned from their own and others' experiences) are provided, each of which is briefly described below:

Presence as realism This is presence determined by apparent realism of some entity, abstract, or otherwise (e.g. of objects, events, and people.)

Presence as transportation This is sub-typed as "you are there", "it is here", or "we are together": The first of these "you are there" describes the sense that you have been transported to some place.

The second "it is here" describes the presence of objects that have been brought to you.

The final type, "we are together" relates to a sense of presence in a shared space, typically found in conferencing systems.

Presence as immersion This is presence as determined by the extent that one has the real world 'shut out'. It is described in general terms that include both the technological and psychological definitions of immersion as described previously (N.B. This conflicts with the definition of immersion as set out for the purposes of this thesis in Section 2.3.)

Presence as social richness This focuses upon presence as effected by interpersonal communication.

Presence as social actor within a medium Here, one feels present with some actor within a medium.

Presence as medium as a social actor This type of presence is experienced when a user attributes social meaning to a medium itself.

Within this thesis we shall be concerned with how an isolated individual experiences an immersive virtual environment as a 'physical' place. Perhaps the closest definition of these six would be that of 'Presence as transportation: you are there'. That is, as one uses ones senses to determine their immediate environment they 'read' the artificial stimuli from the IVE system, and appear to be somewhere other than in their physical location, and they therefore feel present there. Additionally, this normally occurs even though many of their senses may indicate otherwise. That we have identified the most apt definition of presence from this list (for our purposes) illustrates that there are aspects of presence that our outside the scope of our interest. Other types of presence are just as important and useful, and there is of course overlap between types of presence in any VE system.

The idea that 'you are there' in fact confuses two aspects of presence, 'there' as a space, and 'there' as a place. A space is something that can be probed by our motor-senses. Loomis (1992) describes the perception of space as distal attribution, being able to sense that which is beyond the extents of our body. Ellis (1996) uses a similar notion on which to base his discussion of presence, by abstracting from the 'classically defined virtual image of geometric optics' (to use his terms):

"... a viewer interprets patterned sensory impressions to represent objects in an environment other than that from which the impressions physically originate."

Slater and Usoh (1994) touch on both these ideas at once when they describe an immersive VE as providing: a representation of the immediate surrounds of a participant, consisting of their (virtual) body and an environment as displayed from the unique position and orientation defined by the participant's viewpoint. This description notably includes a virtual body (Slater and Usoh, 1994), and by extension, the notion of proprioception. The inclusion of a virtual body representative of the participant in an IVE has been given great importance, according to Regenbrecht and Schubert (2002) as it determines the 'meshings' between the participant's actual body and the VE. Slater and Usoh (1994) explain that the inclusion of a virtual body is an attempt to reduce the contradiction between sensory data and proprioception. (Proprioception consists of signals regarding the disposition and dynamic behaviour of the human body received by the brain.) In providing the virtual body, proprioception then refers to the virtual rather than real body. The effect of doing this is such that 4 out of 24 of the participants of Slater's experiment began to move their real left arm in sympathy with the virtual body's movements (only the real-right-arm was tracked in space to control the virtual-right-arm, whilst the virtual-left-arm simply mirrored the virtual-right-arm.) The ability of the human mind to adapt to a virtual body was also noted by Slater (2009), where he relates it to the 'rubber hand illusion' as demonstrated by Tsakiris and Haggard (2005), and further, to "whole body displacement" demonstrated by Ehrsson (2007); Petkova and Ehrsson (2008). In studies such as these either a limb, or the entire body of a person is apparently relocated. The person, to some extent, believes that a rubber hand is in fact their own or that their body is elsewhere or is replaced by another person's. For instance, in one experiment (Petkova and Ehrsson, 2008) a knife is (safely) held over the experimenter's arm and this arouses a significantly greater response in the subject than when it is held over their own (actual) arm. Summarising from the findings of such research, Slater goes on to state that "VR can transform not just sense of place, but ... your own body". Due to the evident importance of the virtual body, Slater asserts that this matching of the real and virtual should mean that, for instance, really walking should control walking in an IVE (Slater and Wilbur, 1997). On the other hand, Petkova and Ehrsson (2008) does comment on the surprising malleability and adaptability of the proprioceptive system, and no doubt this would result in some level of tolerance that will be a small relief to IVE designers.

Not only is a virtual body expected to match the movements of a user's real body, but there are other spatial constraints expected. In Slater and Usoh's (1993a) experiment participants were said to have expected realistic constraints such as 'shattering of objects dropped' and 'not walking through walls'. This is reiterated and expanded upon in Slater's later paper (Slater, 2009), under the term 'plausibility'. The link between the virtual environment and the real space should therefore maintain a 'lawful relationship' between efference and afference so that the observer may model the relationship between it (Loomis, 1992). Sheridan (1992) suggests that distortions in the afferent and efferent loop when using an immersive virtual environment should be tested for their effects upon "presence, training efficiency, and performance". Sheridan also states that "strict geometric isomorphism" may not always be the ideal case for modelling IVEs (actually, Sheridan considers both virtual environments and teleoperation systems), due to hardware limitations or constraints of the human body. This is a particularly interesting point as Ellis (1996) later points out scenarios in which performance is improved by distorting or transforming the behaviour of a perceived environment to facilitate human interaction (see 2.4.1). At the extreme, Schubert et al. (1999) comment that 'conflicting stimuli from the real world' must be avoided to ensure presence, and similarly, Held and Durlach (1992) that devices should not generate artificial stimuli that bring attention to their existence.

As the human has multiple spatial senses it is unsurprising that presence is thought to be better supported when more human senses are catered for by an immersive system (Held and Durlach, 1992; Slater et al., 1994; Steuer, 1992; Ellis, 1996; IJsselsteijn et al., 2000). Hecht et al. (2006) found that VE response times were shorter (better) as more sense modalities were supported and argued that this may relate to the sense of presence that their subjects experienced (see Section 2.4.1). They also suggest that the support of one modality may be able to compensate for a lack of support under another as does Zeltzer (1992), and Ellis (1996).

Not only might one expect that different persons may have biases for particular senses (for example, if one has impaired hearing then a phonic bias may be very obvious), but Slater et al. (1994) while considering Neuro-Linguistic-Programming (NLP) suggest that presence may be experienced differently in different modalities.

When we experience a space, we also attribute to it a sense of place. Turner and Turner (2006) suggest that, given an IVE, a sense of place is the conjunction of space and meaning (this definition of place is attributed to Harrison and Dourish, 1996)[9]. Slater and Wilbur (1997) mention that immersion takes place at different levels of cognition, so that we have autonomic responses that relate to, for instance, proprioception and sensorimotor-contingencies, and we have higher-cognitive responses that relate to more considered behaviour. These seem to fit with the separation of space and place concepts, the former relating to low-level cognition and senses and the latter attributing higher-level cognitive meaning to that which is experienced. The sense-of-place and sense-of-space together seem to first have been identified by Slater and Usoh (1993a) as a key indicator of presence after an experimental participant noted that their IVE experience felt more like "somewhere I visited, rather than something I saw (as in a film)". Subsequently the statement "The computer generated world seems to me to be more like (1) something that I saw.... (7) somewhere that I visited" was included in the Slater-Usoh-Steed (SUS) questionnaire (Slater et al., 1994). From a subjective-measures-questionnaire perspective this question is undoubtedly one of the most insightful posed to date, because navigational spatial maps are stored differently to images in the brain, firing specific place cells in the hippocampus (O'Keefe and Nadel, 1978).That the experience is likely to involve the production of a spatial map, that this appears to be reportable by the participant and that Slater realised this, is evidence for the potential use of spatial memory and even neuroscience to aid presence research.

Another reason that presence is difficult to define is because it is a psychological state (Witmer and Singer, 1998). Not only does this mean that it is subjective and thus difficult to measure, but Slater and Wilbur (1997) note that immersion must occur at different levels of the sense, perception, cognition continuum, as there are numerous subjective components involved in its development. Some of these components are autonomic, such as accommodation, and others more volitional

---

[9] Heeter (2003) relates presence to an experience as well as a place.

such as attention. The imaginary worlds constructed when reading books have been said to invoke presence of a sort ('imaginary presence'), as has watching films (Herrera et al., 2006). But although this type of presence may be regarded as being outside the scope of IVEs [10] presence as a psychological state has been described as the 'suspension of disbelief' (Slater and Usoh, 1993a; Hand, 1994). Although it has not been made clear whether this is simply an illustrative description or instead intended as the basis for an operational description of presence, the idea has been criticised by Waterworth and Waterworth (2001) and Lee (2004) who respectively opine and argue that presence is possible without the suspension of disbelief. Waterworth & Waterworth state that it is unnecessary if the presentation is "immediately perceptually engaging", and that the suspension of disbelief is required when the VE presented is 'unrealistic'. Lee on the other hand purports that it is more natural and even an 'easier operation' to accept VE stimulus than to defy it. There has also been evidence against the 'suspension of disbelief' description in the form of work carried out by Herrera et al. (2006). This work investigates the interaction of participants having autism within a 'Virtual Reality'. Herrera et al. remarks that whilst persons with autism generally obtain low scores when asked to act in pretence, persons tested positive for autism have 'accepted' a virtual reality system, interacting naturally with it[11] If true, it seems that virtual environments may have little to do with pretence and belief.

Some researchers have in fact suggested that reality judgements could be used to indicate whether virtual environments induce presence. The idea of a Virtual Turing Test, as put forward by Schloerb (1995), relates the probability that an IVE subject believes that they are in a real environment to a quantifiable presence value. By offering a mediated and potentially degraded experience of both a real world environment and an immersive virtual environment, a Virtual Turing Test could be created whereby experimental subjects are asked to distinguish between the two environments. The degree to which the mediating system degrades the experience is obviously a key issue, but this may be accounted for by the particular hypotheses being tested. Interestingly, that IVEs are mediated has also attracted attention. Lombard and Ditton (1997) have defined presence in terms of mediation, specifically as "the illusion of non-mediation", because the presentation of the virtual environment should make the equipment necessary to view it transparent to the user. By focussing on transparency, this perspective tends emphasise the weaknesses of immersive technology, such as its physical characteristics (such as weight, size, restrictive cables), low-fidelity displays, artefacts, and responsiveness. However, it is interesting to note that presence is attainable despite such problems, as it exists in systems currently available. Finally, in a similar vein Waterworth and Waterworth (2001) interpret presence as: "...such that virtual environments do not require mental modelling (knowledge in the head) to make sense. Rather they contain 'knowledge in the (virtual) world,' which is amenable to direct perceptual processing".

---

[10] Although we are concerned with systems that are by definition immersive, Zeltzer (1992); Steuer (1992) both provide models that relate presence to systems that provide varying levels of technological immersion. One might also note that terms such as 'Virtual Reality' that in some circles are uniquely identified with immersive virtual environment systems elsewhere may refer to any conceptual environment that is not 'real' (Hand, 1994).

[11] Unfortunately details of the particular experiment cited are not publicly available.

This implies that the mediating device is transparent because the stimuli is processed ‘directly’ but avoids mentioning the device(s) directly. If presence is attained through some medium, then there must be some communication, hence Steuer (1992) defines presence (using the term telepresence) as “the experience of presence in an environment by means of a communications medium”. In fact, Freeman et al. (1999) define presence purely in terms of the communication of a message, which is an extremely broad definition. If this is generally accepted, then many of the techniques of communication theory may be employed to directly work with presence. However, if we consider presence in a spatial sense, then we constrain not just the content of the message, but the form in which it is presented.

Another problem due to presence being a psychological state is that such states have the potential to blur. One way in which this is possible is through variance in attention, as illustrated by Bystrom et al. (1999) in their model of “Immersion, Presence and Performance”. Waterworth and Waterworth (2001) provide a model that explicitly describes both attention and presence as continuous in nature, with three axes to represent absence versus presence, virtual versus real world, and conscious versus unconscious (attention to the environment) continua respectively. They are not alone in regarding presence as laying on some continuum, for instance Witmer and Singer (1998) state that “humans experience varying degrees of presence” and include attention as part of their explanation of presence. However, in contrast to this, Biocca (1997) suggests that presence exists in exactly one of three states at any one time: the real environment, the virtual environment, or the imaginal environment. On an intuitive level at least one can identify with this proposal, as we can intentionally bring our attention to each of these quite readily. Schloerb’s (1995) view is that an IVE user may be asked whether they are in environment ‘X’ or not, such that the probability of replying ‘yes’ may be used as a subjective presence measure. Indeed, the idea that presence involves a selection between alternatives has also been proposed by Slater (Slater and Steed, 2000; Slater, 2002). Slater regards an environment as providing stimuli that form a gestalt, with attention selecting from one of these competing signals at any one time, in a binary fashion: you are either here, or you are there. Thus the experiencing of presence depends upon the continuous and immediate interpretation of stimuli at a certain point in time. He uses these ideas to explain that even unexpected stimuli from a non-selected environment may be interpreted as arising from the attended environment. This idea is operationalised through the development of a Markov Model that represents the probability of switching between the real world environment, and the virtual environment. Presence as evoked by a specific IVE is thus seen as the overall integration of periods in which the observer felt present (Slater and Steed, 2000). Although Steuer (1992) had stated that a binary model of presence prevents the comparison of VR systems, it seems that the models provided by Schloerb (1995) and Slater and Steed (2000) both support comparison through the estimation of probability values (that reflect the probability of being present in some IVE.) It should also be noted that IVE systems may be compared according to the immersion-heirarchy concept laid out by Slater (2009) (see Section 2.3).

More recently, Slater’s view of presence has been further developed, dividing it into two main concepts, the Place Illusion (PI), and Plausibility (Psi) (Slater, 2009). PI refers to the spatial aspect, much as it was described by Loomis (1992) as ‘distal attribution’. Loomis provides a concise description that is a little similar to that of Ellis (1996), although defined for the more general case. Slater states Psi to be “concerned with the ‘reality’ of the situation depicted... [including the] credibility of events in comparison with what would be expected in reality in similar circumstances”. Slater states that “PI can be different in different modalities” so that PI could be sustained in the visual sense while the subject “simultaneously [has] a conversation with someone who is outside the virtual environment...clearly in the auditory domain there is no PI”.

While the concepts of PI and Psi are clearly major aspects of the presence phenomenon, we would conjecture that Psi is actually another PI, albeit one thought of as being at perhaps a slightly higher cognitive level of cognition than the spatial PI originally described. Indeed, this would unify the PI and Psi concepts, bringing them more inline with the idea that the brain simply acts (for perception) as a “correlation engine” as Slater states elsewhere (Slater et al., 2009b), and this is also the author’s view. This also supports the general (and intuitive) idea that realism and presence ought to be related. Under this view we would also conjecture that each PI would have its own threshold (Slater, 2002), and have a binary state, such that sufficient IVE cues would be required to exceed the threshold, switching the particular PI from perceiving the real world environment to the IVE. The *number* of PIs connected to the IVE could then be related to the level of presence experienced. This conjecture would link the concepts of PI and presence-as-a-multidimensional-construct as suggested by Ellis (1996) in particular; thus the greater the number of PIs[12] and the greater the size of the effect of each of them, the more likely it would be that the stimulus would induce a sense of presence. Such thresholded multidimensional factors that lead to presence would be analogous to, and could then be directly modelled by, a neural-network. This leads to the idea that this may not just be a method for presence modelling in simulation but may even reflect the modelling of presence as it happens (literally) in the brain. Again though, it should be pointed out that this is conjecture only.

The PI and Psi theory could explain why improvements in visual quality have seemingly not affected presence in some research (Zimmons and Panter, 2003; Mania and Robinson, 2004) but the addition of other types of visual cue (concerning ‘other PIs’) has significantly increased presence (Slater et al., 2009a). Although there is retrospective and fragmentary evidence for Slater’s theory, it has now begun to be tested directly (Slater et al., 2010).

Within the presence literature philosophical considerations are also made (Floridi, 2005), and perhaps one of the most prominent is the Heideggerian and Gibsonian founded work of Zahorik and Jenison (1998). By tying presence together with ontological philosophy they define presence as being “tantamount to successfully supported action in the environment”. Regenbrecht and Schubert (2002) are supporters of this perspective and point out that the mere illusion of the potential to interact should enhance presence.

An affordance (Gibson, 1979)) is the potential to interact or do, so rather than emphasising the perception of the environment itself, we are present within it if we perceive that we may interact with it or objects within it. This emphasis upon interaction in such views is an important point, as it is

---

[12] or perhaps better, the greater the extent of a continuum of PI responses

widely accepted that even low-fidelity environments can induce presence (Sanchez-Vives and Slater, 2005; Harvey and Sanchez-Vives, 2005) and that it is the proprioceptive-display concordance that is most important for presence (Slater and Usoh, 1994; Slater, 2009). Whether Zahorik & Jenison's existence-as-action may be tied to presence is not clear to the author, as one can surely argue that while interaction may increase the probability of presence, one can be present even without any possibility of interaction. It would seem more appropriate to state that unsupported action in an environment can reduce the likelihood of presence. This is particularly the case if one expects to be able to perform an action and cannot, which might initiate the logical question "why can't I do this?" leading to questions regarding limitations in the technology being used. Similarly, Zahorik and Jenison (1998) conjecture "Successfully supported action in the environment is a necessary and sufficient condition for presence." It seems feasible that successfully supported action is a sufficient condition for presence in its most general and abstract form. Consider an example where an IVE is being used, but the user cannot form any idea as to what the surrounding images are (such as random coloured dots and lines). As they move their hand forward and back, they realise that the images changes somehow. At this point they are certainly interacting with the environment. In a most general and abstract way they could be said to be 'present' in the world of randomly coloured dots and lines — perhaps not dissimilar to times when the mind is considering abstract concepts (such as performing mental arithmetic.) However, perhaps this is not so much related to the world of three-dimensionally rendered environments, but rather strongly related to a larger class of presence questions belonging more to the fields of ontology, psychology and philosophy than to presence in immersive virtual environments. Zahoric and Jenison's assertion that successfully supported action in an environment is a necessary condition for presence can also be considered unsatisfactory. As mentioned above, most of the time we as humans can interact immediately with our environment. Sometimes however, this is not the case. "Patient Awareness During Anaesthesia" (Communicore, 1996) rather terrifyingly documents the hazards of using the combination of anaesthetic and neuromuscular blocking agents that are routinely used for surgery. From rather disturbing statements recorded from patients (such as "I was still perfectly lucid when the obstetrician plunged his knife into my abdomen.") and verifiable facts in the document, it appears that a sense of presence may be experienced when interaction is not possible. Floridi (2005) has also criticised Zahorik and Jenison's account of presence for the same reason remarking that it does not account for passive presence, that is, presence without any observable interaction. It seems that Zahorik and Jenison have confused the issues of presence in the context of virtual environments and ontology by assuming that they are one and the same. However, although the phrase 'being there' is often used in the presence research field, it is not the 'being' as opposed to 'not being' that is emphasised, but the 'there' as opposed to 'here'.

Presence and Performance

At the opposite extreme to the philosophical discussion is the mechanistic operationalisation of presence. Ellis (1996) discusses a cybernetic approach for measuring the effectiveness of virtual environments excluding issues of "the 'impression' communicated by the interface", which may refer to the subjective experience of presence as 'being there', and instead focusing on situations involving the communication of some message between the user and system. This type of problem is often tackled using information theory principles, may be framed in terms of task performance, possibly belonging within the field of human-computer-interaction (see Mulder et al., 2004). In his paper, Ellis' point is that task performance while using a virtual environment is not necessarily improved by inducing of a sense of presence. This is backed up with examples in which task performance is increased by reducing the veridicality of displays and hence decreasing presence. While Ellis takes pains to debunk the idea that increased presence should lead to increased task performance with a particular emphasis open teleoperation, Slater et al. (1996) addresses immersive virtual environments more directly. Slater argues that "it is posing the wrong question to consider whether presence per se facilitates task performance" because task performance is related to user interface issues whereas presence is about supporting "natural reactions to a situation (which may or may not have something to do with efficiency of task performance [in a virtual environment])". Hence, he later points out that presence should be measured independently of immersion as immersion might influence presence and task performance in different ways. This view appears to have gained the general support of other researchers in the field (IJsselsteijn et al., 2000). On the other hand, Held and Durlach (1992) suggest that presence and performance have an important relationship that should not be overlooked. Their assertion is that since presence supports natural human interaction and the most general-purpose system available for performing tasks is "us (as operators)", if we regard immersive technologies as providing a platform for performing general tasks there should be a connection between the two concepts. By taking each of these points into consideration it seems that the future of general-purpose immersive systems should support presence as best they can, because transformations that improve human-computer interaction could be added as an extra layer within the application domain rather than being placed within the platform itself. It is because of this that we feel that pursuing presence inducing immersive virtual environments is a valid objective but that task-performance is not directly useful for measuring presence.

2.4.2 The Utility of Presence, and its Effects

The concept of presence is highly esteemed in the world of VR, and this is somewhat demonstrated by the quantity of literature concerning it. In the early 1990s the literature emphasised teleoperation with at least one researcher raising the question whether presence might be of value at all (Sheridan, 1992), and others admitted that its usefulness for this purpose was not clear (Held and Durlach, 1992). In relation to this, it seems that teleoperation was the driving force behind Ellis' (1996) criticism that presence need not necessarily increase performance. Since this period teleoperation seems to have taken a back seat, while presence has been considered in a very general sense for a wide range of applications, and with a particular prominence of virtual environments.

The reason for this change in emphasis must surely be due in part to the ubiquity of virtual environments and the technology to support them. Virtual environments are now so commonly found that it is of little surprise that they are under scrutiny in a variety of scientific fields to determine how they might be leveraged. However, this is much less true of the sub-field of immersive virtual environments that, requiring relatively expensive devices, remains a smaller but quickly growing area. But although they are scarce, IVEs provide a unique way to experience virtual environments, by enabling presence. Whereas the perspective of a virtual environment and the ability to move about it and view it in a natural way (i.e. using the body) is unique, it is not the only benefit to an IVE. Slater et al. (1996) notes that presence is important because it facilitates and evokes behaviour consistent with that which would occur if the virtual environment were real. Examples of the former are architectural walk-throughs, and the development of manufacturing designs. Examples of the latter may be found in more complex situations including training (Vora et al., 2001), and virtual reality therapy[13] (Strickland et al., 1997) such as treatment of acrophobia (Emmelkamp et al., 2002). It almost goes without saying that there are instances of IVEs in which presence is so central that it almost becomes an end in itself, such as for teleconferencing and escapist entertainment, and even relaxation (Freeman et al., 2004).

Although future technologies are likely to reduce the costs of IVE equipment, advances will no doubt improve immersion in terms of its breadth and depth (see Section 2.3.2). But Hand (1994) suggests that waiting for technology advances to obviate the study of presence in IVEs [14] is to attempt to avoid important issues. These issues are due to our ability for high-level cognition, in other words, even if technology sufficiently provides accurate and high resolution displays to our senses, perhaps even including all sense modalities, there is still the consideration (for instance) of what content is displayed and the way in which it is presented. To investigate this problem of content we must look for solutions using presence research.

Slater and Wilbur (1997) states that the concept of presence remains useful if only as a driver for immersive virtual environment research. He explains that if we think of it this way, then we can utilise presence research to help understand what factors distinguish egocentric and exocentric systems. For instance, using Slater's example, we may attempt to use presence to compare IVEs to desktop-displayed virtual environments, in order to find what benefits each has over the other.

Whether there is a link between presence and task performance is highly dependent on the context, but in some cases presence should influence task performance. This was demonstrated by Slater et al. in their experiment using Tri-Dimensional Chess, where they set out to investigate the issue by comparing a desktop system to a highly immersive CAVE (Slater et al., 1996). From their study it was found that task performance was positively associated with an egocentric over an exocentric display, and the egocentric display resulted in significantly increased presence scores being reported over the exocentric display.

A particularly interesting issue regarding the utility of IVEs is found in the simulation of hazardous situations. IVEs appear to lend themselves to such simulations for training purposes, promising to mimic the real world so that responses to such situations can be rehearsed in safety. But Held and Durlach (1992) assert that immersive systems will always have limitations, and that this leads to users developing ways to adapt, usually as their exposure to the systems increases. Whether this adaptation is conscious or not is unimportant, nor is the amount of effort required to optimise it, but rather that the perceptions or responses of users may be aberrant. In the same article they also suggest that familiarisation with an

---

[13] The flexibility of a VE allows full control over the situation, and importantly, the object of the phobia. Phobia exposure therapy can therefore be implemented using finely and well-targeted controls, and can be adapted to an individual's needs.

[14] Although Hand uses different terminology that is more loose

IVE system will enhance the experience of (tele)presence as users habituate — which would appear to compound this problem. However, although IVEs can affect perception of real world situation, such as in virtual reality therapy (Pertaub et al., 2001), there appears to be no research regarding this as a potential problem. In a similar vein there is a second likely problem, namely that within an IVE consequences are also virtual, or may not even exist (that is, they may not be modelled by the IVE system.) [15] This is a mixed blessing. While Zeltzer (1992) explains that some applications are benefited by the fact that dangers are not recreated by immersive virtual environments, this fact could instead have serious detrimental repercussions. To take an extreme example, to train a soldier within a battlefield scenario one can never reproduce the danger that would exist if the scenario were real — that of becoming wounded or even dying. Therefore we cannot expect our soldier to react to the IVE as if it were real, consciously or subconsciously. Therefore, ironically it is likely that the applications that exploit the lack of danger presented by an IVE are those that in fact would benefit by a level of danger to make them truly effective. Even if there are serious consequences introduced, such as keeping the soldiers' IVE performance on their record, one cannot imagine this could compare with the sense of danger of a real situation. One wonders whether this was part of the thinking behind Ivan Sutherland's words when he described the 'ultimate display' and explained that within it a displayed bullet would be lethal. It is fortunate that people respond to IVEs as though they are real even when they are very obviously not, but determining the limits of this effect would be of great benefit, especially for hazardous IVE simulation.

A major thrust of presence research concerns how subjects react to IVE stimuli. Surrogate measures of presence are often defined in terms of these reactions. They reflect the interesting phenomenon that people very often react to immersive virtual environments as though they are in fact real, almost irrespective of the general lack of fidelity and realism that is prevalent at this point in time. A good deal of literature therefore concerns itself with quantifying the perception of virtual stimuli by measuring reaction. Zimmons and Panter (2003) demonstrated this well in an experiment that varied the level of rendering quality across experimental conditions. Measuring the heart rate and skin conductance (both of which are said to reflect stress responses), they had subjects drop objects into a virtual pit (onto a target.) There were five levels of rendering quality: the rendering of black and white lines, to low and high-quality textures without lighting solutions, and low and high-quality with lighting solutions. Using the accuracy of the objects being dropped onto the target, Zimmons and Panter found that task performance did not vary according to rendering condition, although task performance did correlate inversely with heart rate. Having used a training room as well as a separate room that contained the pit, they were able to determine that heart rate increased significantly when subjects entered the pit room. A presence questionnaire was also administered, and it was found that presence responses were not significantly different between rendering conditions. This experiment demonstrates that even low-fidelity IVEs can induce a sense of presence, as evidenced by the increase in heart rate. Therefore experiments such as these do not only tell us something about IVEs, but they also tell us about how we as humans perceive ourselves as being present within environments (e.g. regardless of their apparent detail and accuracy.)

---

[15] Banos et al. (2000) consider the reality of virtual environments, and define reality in terms of consequences that might affect the user.

It seems then that the utility of presence lies in the ability to simulate situations and induce responses that would occur if the situation were real (Slater and Wilbur, 1997). But as we have discussed, there are limits, and although IVEs have been shown to be efficacious for training in some contexts, great care is required in others (and particularly those that involve the simulation of dangerous situations.)

2.4.3 Presence Determinants

Many experimental studies have been carried out to determine presence factors in an exploratory approach to understand the concept better. This is aside from those who have sought to understand the concept by suggesting tentative theoretical models to be built upon (Held and Durlach, 1992; Loomis, 1992; Sheridan, 1992; Steuer, 1992; Zeltzer, 1992; Ellis, 1996; Draper et al., 1998). Sadly there have been few publications such as Hecht et al. (2006)'s that attempt to tie such theories and evidence together either directly or indirectly. The problem is not only that a potentially infinite number of factors will affect a sense of presence, but perhaps that the experimental research has not been sufficiently systematic (Lombard and Ditton, 1997), this being compounded by the fact that, at least for now, there is no independent objective measure for presence.

Determinants Theory

IJsselsteijn et al. (2000) provide an overview of the aforementioned theoretical models and attempt to reduce them to four common factors:

Extent and fidelity of sensory information — essentially Steuer's concepts of 'breadth' and 'depth'.

Match between sensors and the display — "...the mapping between the users' actions and [their effects]" in terms of 'sensory-motor contingencies'.

Content factors — "a broad category including the objects, actors, and events represented by the medium...[and the] ability to interact with the content and to modify it". This also includes 'autonomous factors', that is, autonomous behaviour of agents and objects (such as physical simulation including gravity, fluid dynamics, etc.) as noted by Zeltzer's 1992 'autonomy' in his 'Autonomy, Interaction, Presence' model.

User characteristics — psychological and physical characteristics of the users.

This list provides a good reference for the general determinants of presence, although it does not provide a model as the original authors do. Such models aid the comparison within and between these factors, which is what Ellis (1996) suggests by forming equivalence classes that allow us to estimate likely presence when factors have fixed-values [16] Neither does it help us to compare the relative importance of factors in an ordinal sense, for instance to support or reject Heeter (1992) assertion that responsiveness is more important than fidelity. Zeltzer (1992) argues that as the real world is too detailed to simulate, and because some perceptual cues are not practical to implement, just the necessary sensory cues required for an application could be supported by an IVE. Such a model would aid the testing and possible implementation of this idea as well.

---

[16] Zeltzer (1992) had earlier concluded that sensations could be 'suggested' or evoked by 'substituted' cues.

Specific Determinants

Although presence is thought to be a multidimensional phenomenon (Hendrix and Barfield, 1996), a comprehensive compilation of presence determinants as investigated or theorised has been published by Schuemie et al. (2001). Here we shall consider some of the specific factors that relate to the matters of this thesis as addressed by the research community. Witmer and Singer (1998) state that '...presence measures should assess [the] individual differences as well as characteristics of the VE that may affect presence.' as these are the two groups of presence determinants. Slater and Usoh (1993a) saw that presence factors may be split into user (or subject) exogenous and endogenous factors, and so we shall now use this convention and consider each.

Exogenous Factors Within the experiments of this thesis we utilise a HMD to present the virtual environments. There are several characteristics of such displays that have been associated with presence. A binocular stereoscopic display as opposed to a monocular display resulted in an increase in reported presence (Hendrix, 1994; Hendrix and Barfield, 1996). This result was expected for two reasons. Firstly, Hendrix and Barfield write "This is based on the observations that we use our stereoscopic vision daily to manipulate objects and navigate within a three-dimensional world." Secondly, they make note of the published results of several experiments that indicate how performance had increased when stereoscopic cues are available. Whilst performance and presence may not be related in a bidirectional causal sense, such experiments demonstrate that stereoscopic cues can effect changes in the perception of and interaction with virtual environments. Although their experiment showed a connection between presence and stereoscopy, they did not find that responses to a question on 'realism' changed significantly with the stereoscopic condition. This however was likely to ghosting - a potential artefact when using some types of stereoscopic technologies - as reported by the experimental subjects. In the same article, the effect of a head-tracked display is considered. Hendrix and Barfield explain that head-tracking facilitates the display of a VE in which the viewpoint is dependent upon the dynamic position and orientation of the user's physical head. The results of their experiment provided evidence consistent with the hypothesis that head-tracking would increase reported presence and realism [17]. Head-tracking technology supports optic-flow interaction with the VE, and thus supports many visual cues that Gibson (1979) argues are essential for perceiving our environment. An excellent overview of visual cues that are involved in viewing immersive virtual environments is provided by Rinalducci (1996). Update rate and latency of displays and their content is an important issue for IVEs. These and other factors must be sufficiently supported to minimise what has been called 'cybersickness', and a treatment of motion sickness in the context of virtual environments is provided by Bles and Wertheim (2000). Kennedy et al. (1993) has devised a Simulator Sickness Questionnaire (SSQ) in order to quantify (subjectively) such effects. Meehan et al. (2003) made a connection between presence and display update latency, when investigating whether heart rate increased when viewing a virtual environment that contained a pit, as opposed to a non-threatening training room. Heart rate did increase significantly when in the 'pit room', but it increased by a greater amount (with 'borderline significance', $p = 0.05$) when there was lower

---

[17] The stereoscopic and head-tracking experiments were run separately, and both used within-group designs.

display latency. They compared latencies of 50ms and 90ms. Somewhat relatedly, higher display frame rates have been associated with an increased sense of presence, where rates of 10, 15 and 20Hz were tested (Barfield et al., 1998). Frame rates are dependent upon (and usually constrained by) the system providing the display signal, in this case the computer and associated graphics hardware. This is unlike the display hardware for which the term refresh rate is used to describe the frequency at which the screen is updated (refreshed), and which generally occurs at a minimum of 60Hz. Although subjects could discern between 15 and 20Hz frame rates, presence was found to be increased in both 15 and 20Hz conditions compared to the 10Hz condition, supporting Steuer's earlier (Steuer, 1992) prognosis. Current technology can cope with generating frame rates above 20Hz with ease unless tasked with highly complex scenes. The experiments described within this thesis however contain low to medium-complexity scenes.

The field of view of HMDs is also one of their main characteristics (see Section 2.2.1). Prothero and Hoffman (1995) found that an increased field of view was connected to presence. In their work, the authors compared a (horizontal) 60 degree and 105 degree field of view and found that there was an increase in reported presence in the wide-angle condition.

Pictorial realism has been tested as a potential presence factor. Zimmons and Panter (2003) tested five levels of rendering quality as reported above (see Section 2.4.2). Texture detail and lighting realism were tested, and it was found that reported presence scores were not influenced by either. However, Welch et al. (1996) concluded that pictorial realism did increase reported presence according to their experiment - using a within-groups design whereby subjects could compare alternate conditions repeatedly in order to make their judgements. The experiment had participants driving (or being driven) in a simulator with differing levels of detail in the displayed environment. They also tested interaction (driving) and passive (being driven) effects on presence, whereby interaction was found to increase presence. Slater et al. (1995) found that by displaying dynamic shadows within an immersive virtual environment reported presence scores increased. Slater also used an objective presence measure, the accuracy in pointing toward a virtual radio in contrast to its real world position as betrayed by the emanation of an audible "meaningless tone" (see Section 2.4.4). After grouping subjects according to their visual or auditory dominance [18], it was found that visually dominant subjects had increased presence scores when shadows were displayed. No such association was found for auditory dominant subjects. Cho et al. (2003) provide evidence that stereoscopic, geometry detail, texture detail, object motion, motion detail, and change-of viewpoint factors all significantly increased presence while viewing a 50 inch display using shutterglasses [19]. Subjects viewed an underwater scene containing computer-animated fish (or inanimate fish, depending upon the condition).

Sanchez-Vives and Slater (2005) state that [visual] realism is less important than head-tracking (see Section 2.2.2) or frame rate. Tracking is important in general, because it is tracking that supports the virtual display that corroborates proprioception. Hence, we sense our bodily posture and see this con-

---

[18] Visual and auditory dominance are concepts defined by neuro-liguistic programming (NLP) theory. Subjects were classed through the use of a post-hoc questionnaire.

[19] The distance from which subjects viewed the display is not given, and so the field of view cannot be estimated.

firmed within the virtual environment. Evidence that head-tracking supports presence has been provided by Regenbrecht and Schubert (2002). In their study 'spatial presence' was found to be enhanced when subjects viewed a head-tracked scene as opposed to a non-head-tracked scene. Subjects viewing the non-head-tracked scene saw it instead using pre-recorded movements.

Usoh et al. (1999) empirically tested how walking in an IVE better supports presence when it is achieved by actually walking (in physical space), rather than actually walking-in-place (or 'on the spot'), and how both these methods increase presence significantly in comparison to the use of virtual flying (across the plane of the virtual floor) to navigate the virtual environment. As part of this work they used a wide-area tracking system to allow walking in a real environment, a neural network to determine when a person walked 'on the spot', and a push-button to enable the 'flying' movement across the plane of the virtual floor. Slater et al. (1998) have also investigated natural movement by having experimental subjects look for fruit in an orchard when the trees were of different heights. In one condition subjects had to bend and move around to a greater degree in order to perform the search. The condition that encouraged more bodily movement was found to be related to greater reported presence scores.

Because the body provides the connection between our conscious self and the environment, as we have discussed, at least some body tracking is essential. Slater suggests that the more bodily tracking available, the greater the likelihood that presence is induced (Slater and Usoh, 1994). To reinforce the perception of the IVE it is logical that limbs portrayed in the virtual environment are (ideally) tracked and thus coincident with proprioception. Doing so results in a *virtual body*. Slater and Usoh (1994) tested the importance of a virtual body in an experiment in which one group interacted with an IVE through the use of a three-dimensional arrow (pointer), and a second group were endowed with a virtual body. After exposure to several environments Slater observed that the participants with virtual bodies were more likely to react to virtual dangers, such as a virtual precipice. This result was not reflected by the administered presence questionnaire however. Also, short post-hoc essays made by the participants were analysed according to a neuro-linguistic programming model. Within the group that had the virtual body a positive association was found between the proportion of kinesthetic references and predicates and the reported sense of presence (according to the questionnaire). In the alternate group (having no virtual body) there was also an association, but it was negative rather than positive so that a greater number of kinesthetic terms correlated with a lower reported sense of presence. Given such hindsight, Slater goes insofar as to state that "Immersion...requires a self-representation in the VE - a Virtual Body (VB).".

While the sense of presence is invoked through the use of cues, as Lombard and Ditton (1997) remarked this should result in the illusion of non-mediation, and this implies that the devices employed should avoid the introduction of artefacts. Thus the devices themselves and what they output should be free from signals indicating their existence (Held and Durlach, 1992). But the devices are not the only potential sources of such problems, and Schubert et al. (1999) mentions that there should be suppression of conflicting stimuli from the real world. This is often achieved by depriving the user of real world stimuli, which Witmer and Singer (1998) state leads to more 'immersion'. Slater and Wilbur (1997) also remark that outside events can decrease the sense of 'being there', although this is not always the case as users can interpret and incorporate a stimulus from ‘outside’ the IVE into the experience, perhaps in a similar way in which this can occur in dreams. Methods to reduce the problem of external stimuli being seen or noticed include adding covers to HMDs to block external light, and the use of white noise to cover up external sounds (Renaud et al., 2007).

Endogenous Factors While exogenous factors are often easier to measure, in contrast it can be difficult to probe for relationships between endogenous and presence variables, especially when the former are also subjective, as both may be difficult to define and measure. However, this has not prevented the investigation of endogenous factors, and many of the presence theory publications include conjecture considering potential relationships between such variables. Witmer and Singer (1998) suggest that presence is dependent upon attention as it is directed in a continuous manner between the virtual and real worlds. Their thesis is that without continuous involvement or a continuous sense of the envelopment of the virtual environment, presence is interrupted:

> “presence depends on the ability to focus on one meaningful, coherent VE stimulus set ... presence is based in attention to continuities, connectedness, and coherence of the stimulus flow....[enabling] the focusing of attention...”

This view presents presence as the result of a system of distraction that lures the user away from real world stimuli, and so according to them more involvement with the virtual environment should lead to more presence. Attention and engagement (or involvement) have also been suggested as presence factors. Engagement could be considered the relative of attention; so that when we are engaged in some activity within an IVE, we are normally attending to the IVE content. Barfield and Weghorst (1993) and Nunez (2004) both conjectured that attention to an IVE should manifest itself in the form of mental load, which could be measurable via a secondary task technique, indicating presence. Using a simpler approach, one of Barfield and Weghorst’s (1993) two studies used a post-hoc questionnaire to measure the sense of presence while experiencing a “flythrough” within an IVE. The post-hoc subjective responses to questions regarding “Presence” and the IVE being “engaging” were found to be significantly and positively correlated. However, this relationship would not appear to be generalisable, as Ellis already pointed out that presence and task performance are not necessarily correlated (Ellis, 1996). Also, Spagnolli and Gamberini (2002), who has described how attention may be split between the cues of contending environments, provides ad-hoc evidence that the sense of presence can be robust in terms of attentional interruptions. Spagnolli and Gamberini suggest that IVE presence is not necessarily broken by external (real world, non-IVE) cues or even by technical problems occurring during the use of an IVE system. This could also conflict to some extent with the findings of Slater et al. (2003), where IVE anomalies were artificially introduced into an IVE experience in order to model the ‘breaking’ of the sense of presence.

However, Spagnolli and Gamberini (2002) use their evidence to describe presence differently, in terms of attention and action in a ‘hybrid’ environment made up of not just the IVE and the real world environments, but also of other spaces in which we may (inter-) act (Gamberini and Spagnolli, 2003), such as the mode of internal thought. But this definition of presence pertains to the situation of the IVE user in its entirety (termed a “configuration”), rather than focusing on the more narrow physical, spatial, and proprioceptive view of IVE presence. This very wide view of presence allows for the consideration of an enormous number of types of presence factors, and allows for such intriguing problems as the ‘book problem’ (Nunez, 2004). Thus, the relationship between presence and engagement is then not only understood in terms of the extent to which a user interacts with an IVE (as discussed), but presence is also inferred by the very fact that a user has the ability to (inter-) act in any number of modalities. Therefore perhaps the main research question is no longer “Does the person obtain a sense of presence within the IVE?”, but rather “Where is this person present?”, or more specifically, “How can we define the environment in which the person is present?”.

One potential method for unifying these views is to be found in Riva et al. (2004); Riva (2006). The authors propose that VE presence is maximised when attention and engagement, with respect to the VE, are optimal (maximal.) Their assertion is that maximal presence in an environment is experienced when we (i) are conscious (proto-presence), (ii) perceive the appropriate environment (core presence), and (iii) conceive ourselves as being in that environment (extended presence). In practice, this means that ideally all modes of experience should be focused upon the IVE (see also Waterworth and Waterworth, 2001, and Section 2.4.1). Perhaps the most clearly practical contribution of Riva et al.’s paper is in the underlining of the importance of realistic concepts and content within VEs. Based upon their model, Riva et al. suggest that not only should a VE be perceived in a realistic manner through our senses that interface with the world, but that natural contemplation of a VE should be indistinguishable to that which would occur if the VE were real. In practice, perhaps the most obvious deficiency of this sort is in the lack of realistic content-composition in IVEs, and this is undoubtedly another great avenue for future presence research.

When considering attention in our own experiments we take a simple and pragmatic approach, by giving our participants open-ended tasks to encourage the attending of the IVE for the duration of our experiments (see Section 3.7). However, we shall also see that it appears possible to not only deduce what is being looked at, but what is being perceived in an attentional sense (see Section 3.3). This is important in the context of attention, as it provides evidence that attention is being given over to the IVE, rather than just real world cues. As regards engagement, for our purposes (in our experiments) it is facilitated only by proprioceptive elements, such as turning the head and body, which while not providing much scope for high-level cognitive interaction, does lend itself to visual (and proprioceptive) exploration through the senses.

Apart from attention, Witmer and Singer also hypothesise several other contributing factors of presence originating from the work of Sheridan (1992) and Held and Durlach (1992). They group these into control factors, sensory factors, distraction factors, and realism factors. Witmer and Singer (1998) devised a questionnaire based on these factors, and this was administered to participants of four experiments in order to develop the questionnaire further [20]. Among their findings, simulator sickness as measured using a questionnaire (Kennedy et al., 1993) was found to be negatively correlated with presence.

---

[20] Although the authors believed considerable progress had been made in measuring presence some of the methodology that they employed has been controversial (Section 2.4.4).

Within the same publication they also developed a separate questionnaire to evaluate the propensity of subjects to become involved and immersed in various situations, named the Immersive Tendency Questionnaire (ITQ). Not only does the ITQ promise to help understand immersive tendencies, but it might also be used to find differences in how individuals respond to the questionnaires as a whole. For instance, the variance and skewness of individuals' responses due to factors such as a tendency to express their opinion using extremities of the Likert scale, and it might be used to evaluate the extent to which they are suggestible. Slater and Wilbur (1997) mentioned in passing that susceptibility to hypnosis and to presence could be related. It might be argued that this could provide a measure of one's ability to 'suspend disbelief' (see Section 2.3.2), and could be connected to Regenbrecht and Schubert's (2002) hypothesis that the mere illusion of the ability to interact should enhance presence. The propensity for individuals to respond using particular modes has also been investigated by Slater et al. (1994) using a Neuro-linguistic Programming (NLP) approach by which individuals are rated according to their use of auditory, visual, and kinesthetic terms in communicating. NLP theory asserts that individuals tend to be dominant in one of these modes, and Slater found evidence to support the classification of individuals according to this model and using this as a predictor variable in regression. In addition to the aforementioned endogenous factors are many others, physical and psychological, such as gender, age, health and computer games experience. However, although these are often used as explanatory variables in regressions and are reported upon in a minor way there does not appear to be a comprehensive study of how these variables alone affect presence.

2.4.4 Measuring Presence

After realising the importance of presence and its place in IVEs, we can only learn to manipulate and turn it into a useful tool by understanding it. To understand it, it must be measurable, not just theoretically but in practice.

As Slater and Wilbur (1997) point out, the distinction between immersive and non-immersive virtual environments may be put to use as a discriminating device for identifying important presence factors, and this is an important first step. Subsequent to identifying presence factors, we wish to find how presence may be intentionally induced, and how much effort is required to achieve this (Slater, 2002). Then we would like to know how the sense of presence may be sustained (Slater, 2002; Marsh, 2003). The concept of breaks-in-presence (Slater et al., 2003; Slater and Steed, 2000) could help investigate how to achieve this by inverting the problem, by discovering what obstacles prevent ongoing presence. Interestingly, the literature does not give much consideration as to how presence might best be 'extinguished', perhaps this is because it is so simple to achieve. However, this should not prevent the consideration of this issue, as it is not certain that that such investigation would prove fruitless. In this vein, in one of Slater et al.'s experiments (Slater et al., 1998) a virtual model of the real lab is used to bridge the real and virtual worlds. It was used when the subject both entered and exited the experimental environment, and the experimenter would even communicate with the subjects wearing the donned equipment so long as they were still within the 'virtual lab'. Although the effect of doing this was not studied itself, it is interesting that it was implemented. Another potential issue regarding the termination of a presence experience concerns any after-effects. The after-effects of IVEs are a natural area for study, as they have been relatively severe when compared with simulator sickness, and even sea- and space-sickness (Kennedy and Stanney, 1997). However, after-effects could provide insight into the presence experience according to Welch (1997). Welch suggests that the extent to which an individual adapts to the IVE will be manifest when they exit the IVE, again having to re-adjust (to the real world). It is hypothesized that the re-adjustment will correlate with the experienced level of presence. How much easier and quicker such readjustments to the real world might be relative to the adaptation to the IVE is not discussed, although it is likely to be of some consequence. The consideration of after-effects for both purposes is provided in Stanney et al. (1998).

At each of the aforementioned presence stages, some kind of measure is vital for investigation. It has been pointed out that presence is likely a multidimensional phenomenon (Hendrix and Barfield, 1996), in which case it could (or should) be measured via multiple modalities. Some researchers have therefore suggested that the measure and technique to be used should be dependent on the aim of a particular study (Waterworth and Waterworth, 2001), the experience intended (IJsselsteijn et al., 2001), and the task (Schloerb, 1995). The alternative is to use an aggregate measure, using multiple measures together to estimate the level of presence experienced (IJsselsteijn et al., 2001; Slater and Garau, 2007; Ellis, 1996). An aggregate measure would have to take into consideration how presence is affected by each modality, each of which could differ (Slater et al., 1994) By using several measures, Hecht et al. (2006) provides evidence that when IVEs display data in multiple modalities (using auditory, visual, and haptic modes) overall response time is increased. Although we must be careful not to confuse performance with presence, this does suggest that individuals have the capacity to simultaneously sense from multiple modalities within an IVE [21]. Thus, through diligent and systematic research, multimodal measures could be used to build the equivalence classes that Ellis (1996) suggested. Then not only could we better understand presence, but as he envisaged, IVEs could be designed to support particular levels of presence.

Scientific measures have stringent requirements that must be met. For the purposes of presence measurement, it has been said that such measures must be operational, repeatable, reliable, robust, and useful with respect to IVEs (Sheridan, 1992; Ellis, 1996). Unfortunately, as presence is a subjective phenomenon, measures are not easily defined. So it seems both a blessing and a curse that immersive systems affect an individual at multiple levels, from autonomic to high-cognitive behaviour (Slater and Wilbur, 1997). This is because while alternative approaches to presence measurement are possible (some being more practical than others to implement) it also means that a comprehensive battery of measures may be required to capture the full effects of a presence experience.

Because IVEs fundamentally attempt to produce an environment that is intended to be realistic (from the perspective of presence), it is perhaps obvious that the tests and measures will inevitably use the real world as a reference point, as noted by Lessiter et al. (2001). If we are to learn something of presence from an experiment using questionnaires, then the responses of the subjects and the interpretation of the

---

[21] It may be possible to increase the likelihood of presence this way, or perhaps even provide supporting cues that could aid in maintaining presence when another display modality is weakened.

responses should necessarily be grounded in real world experiences. Indeed, when experimenters use objective measures it is frequently stated that they wish to identify responses that would be expected if the IVE were in fact a real environment. It seems most important then, that presence measures should ‘work’ in the real environments as well as in IVEs. However, there are two important points to be made about this. First, Usoh et al. (2000) have indeed considered the use of (two separate) presence questionnaires in a real environment versus a representative IVE. They found that the two questionnaires could hardly discern between them, although it had been expected that the real environment should result in higher reported presence scores. Mania et al. (2003) confirmed this, again finding that presence questionnaire (Slater et al., 1998) responses after viewing a real environment were not significantly different from the responses after viewing an IVE [22]. The usefulness of presence questionnaires are considered in Section 2.4.4.

Second, Slater et al. (1995) leveraged the relationship between real and virtual environments by presenting conflicting cues to experimental subjects, making for an interesting and effective method that could aid in capturing subtle differences between the two (see Section 2.4.4).

If the real world is our reference point, then an interesting scale against which presence may be marked was invented by Slater et al. (1994). In an experiment they created a ‘Stacking’ of virtual environments where subjects would don IVE equipment within an IVE (thus, in a recursive manner), in an attempt to place more distance between them and the real world. The idea was that this would induce a stronger sense of presence at increasing ‘depths’. It was found that reported presence increased depending upon the ‘depth’ that the subjects experienced. Thus, it seemed possible that the ability of a specific IVE to generate a sense of presence, might be described as being equivalent to some ‘depth’ (perhaps, according to some standard IVE.)

Sheridan (1992) stated his belief that subjective measures are the ‘essential basic measurement’ for presence, because presence itself is a subjective phenomenon. We shall, however, consider both subjective and objective measures of presence. In the next section we shall consider subjective measures of presence (and in particular, questionnaires, which have been the main subjective device for measuring presence.)

Subjective Presence Measures

In this and the following subsection we shall be considering subjective and objective measures respectively. One of the popular subjective techniques for measuring presence is to administer a questionnaire, though much work is being (and has been) done to obviate the need for them in favour of more objective methods. Several questionnaires have been devised such as Slater et al. (1994); Witmer and Singer (1998); Lessiter et al. (2001); Schubert et al. (2001); Hendrix and Barfield (1996); Welch et al. (1996), but no particular one is prominent above all others. This appears to be partly because presence is difficult to define in a generic (inclusive) manner, and thus remains multifaceted. But in addition, researchers themselves differ in their backgrounds, preferred methodologies, and in their objectives.

---

[22] Previously, Mania and Chalmers tested a real environment versus a virtual one. However, the HMD used in this experiment (Mania and Chalmers, 2001) did not appear to be head-tracked, and thus provided little immersion.

As an example of how presence questionnaires differ, we can look at those of Witmer and Singer (1998) and Slater et al. (1994). By studying them it is apparent that Slater et al. tend to ask subjects about their levels of presence using a number of (mostly) direct questions, for instance, asking whether the subject had a 'sense of being there' in the IVE. Witmer and Singer however take the approach of asking questions regarding other factors, ones that they hypothesize contribute to or effect presence. Each of the items of their questionnaire Witmer and Singer relate to one of the following four factor categories: User control, Sensory extents, Distractions, and Realism. By stating that presence is dependent upon attention and involvement, they explain "We expect that Control Factors may affect immersion but not involvement, while Realism Factors should affect involvement but not immersion. We believe Sensory Factors and Distraction Factors should affect both immersion and involvement." (Witmer and Singer, 1998) As they admit themselves, this indirect approach to measuring presence is built upon the assumption that their questions relate to presence determinants, and that these relationships have not yet been established empirically (apriori). In order to provide some credibility for their questionnaire then, Witmer and Singer (1998) carry out an analysis on questionnaire data taken from several experiments. The analysis is intended to both select questionnaire items and validate them as presence correlates in one step [23]. They then proceed to show that the factors gel together as co-determinants, by correlating each questionnaire item with the total. However, this methodology has been criticised because, in general, the sum of a series of observations is expected to correlate with the observations themselves (Slater, 1999).

While the more direct approach of asking subjects about presence avoids assuming presence determinants, it has problems too. Perhaps the most obvious is that because presence is difficult to define, its meaning is difficult to communicate to experimental subjects. Subjects may therefore respond according to their own concept of presence, and this could be entirely unrelated to what was intended to be measured (Slater, 2004; Freeman et al., 1999).

In fact, the use of questionnaires for research purposes is by no means trivial and there are numerous potential pitfalls to be avoided and careful considerations to be made. For instance, language must be clear and should be concise, as the questions are otherwise less likely to provoke a considered response; questions can be logically grouped or (conversely) the distribution of related questions can be increased or randomised to obtain more independence in their respective responses; the ordering of the questions can affect the answers given (Bradburn et al., 2004). Of course there are many such considerations, and because much research has been done to investigate these, there are standard texts available detailing the scientific approach for the design of research questionnaires (Bradburn et al., 2004; Munn and Drever, 2004).

When researching a novel field wherein a particular response variable can only be measured indirectly, questionnaires may in some cases be developed using a statistical technique called factor analysis. Factor analysis can be used to analyse common-but-unspecified factors underlying a battery of related questions. This is achieved by computing the covariance of question-responses to determine a (new) set

---

[23] Generally these two steps should be separated and independent, for instance by first selecting the items as they have done, and then using a cross-validation method to show that the resulting measure generalises.

of orthogonal variables that have ‘loadings’ onto each question. The loadings, which reflect the relationship between generated factors and the original questions, can then be used to investigate the relationship between the questions themselves. An excellent in-depth explanation of factor analysis can be found in Chatfield and Collins (1980). In the following paragraphs, we mention several research investigations that have used factor analysis to grapple with the difficult-to-measure concept of presence.

Lessiter et al. (2001) developed a generic questionnaire for measuring presence across media types. The rationale behind their objective (to create a cross-media questionnaire) was that previous research had demonstrated that reported presence could not only describe experiences within IVEs, but also that it had discriminated, for instance, between mono and stereoscopic presentations using 20 inch, colour CRT monitors (Freeman et al., 2000) — which is not an especially immersive technology, if it can be called immersive at all. The Lessiter et al. (2001) questionnaire leveraged the work and hypotheses of many others, including those of Witmer and Singer (1998) and Slater et al. (1994) mentioned, in order to generate a preliminary set of questions. After its creation, a factor analysis was carried out on data arising from the questionnaire, and this resulted in a compelling factor structure. Also, the analysis was validated using cross-validation, and the resulting questionnaire was shown to be internally reliable. The four major factors as identified by Lessiter, Davidoff, Keogh, & Freeman were ‘Sense of Physical Space’ (sense of being in the virtual environment, ‘Engagement’ (involvement with the content), ‘Ecological Validity’ (believability, and realness), and ‘Negative Effects’ (mostly physiological reactions, for instance, nausea or eye strain). Some may argue that presence is a phenomenon that requires immersion, where immersion is defined as having an element of three-dimensional spatial interaction. Although this constraint is not upheld by the research just described, one can argue that there are elements of the displays that did visually appear as in the real world. For instance, the very basic cues consisting of colour, edges, textures and so on. This debate though, must remain outside the scope of this thesis, but one hopes that presence researchers will address this in detail in the near future. The questionnaire is available for free use for non-commercial research (see Lessiter et al., 2001 for details).

Schubert et al. (2001) developed a presence questionnaire in a similar way, again leveraging work from various sources that include the commonly found works of Witmer and Singer (1998) and Slater et al. (1994). After conducting an exploratory factor analysis on data captured using an extended version of their resultant questionnaire, they found eight major factors, explaining 50% of the total response variance. The first of these factors (having the greatest eigenvalue from the analysis) explained over 20% of the variance and was termed ‘Spatial Presence’ based upon the questionnaire items that loaded greatly upon the factor. The item that loaded most on the factor was the question from the Slater et al. (1994) questionnaire, ‘In the virtual environment, I had a sense of being there...’. Of the remaining factors and the items that most loaded upon them, attribution was given (mostly) to Witmer and Singer (1998) work. A second level factor analysis suggested a set of three factors relating specifically to presence: ‘Spatial Presence’, ‘Involvement’, and ‘Realness’. Remaining factors (i.e. not belonging to the presence category) were categorised under the headings ‘Immersion’ and ‘Interaction’. Schubert, Friedman & Regenbrecht also include details of a second study, the data of which is used to perform a confirmatory factor analysis, though this was to consider presence factors only. The results of this supported the first analysis, particularly ‘Spatial Presence’ and ‘Involvement’, which were very stable. The authors themselves point out that their conclusions are supported by the work of Slater et al. (1994)(SUS), Witmer and Singer (1998), and Lessiter et al. (2001) (the analysis in the latter research resulted in similar conclusions.) The questionnaire is free for use, and is provided in several languages [24]. It contains 14 questions, but if presence and not interaction is being measured, then the original SUS questionnaire may be more efficacious, having just three or so [25] questions.

Several problems with presence questionnaires have been pointed out by researchers. As mentioned above, the questionnaires are prone to become fairly long. Factor analyses can help with this by reducing the number of items that appear, and thus preventing persons from becoming weary and inattentive. A benefit of the SUS questionnaire is its brevity. Despite its few items it achieves its purpose well, as others’ analyses have shown (see above).

Another concern regarding questionnaires is that of memory. It is known that simply by questioning an individual, one can induce or implicate thoughts, ideas, and apparently even memories. Work by Loftus (2003) has been carried out to show that this is the case, and that memories are often distorted when being reported.

The communication of questionnaire items to respondents is a potential problem so that researchers must take pains to provide unambiguous and comprehensible questions. Given that no precise presence definition has been agreed amongst the research community, it must be that respondents interpret questionnaire items differently. A critical assessment of this situation has been published by Slater (2004), wherein he demonstrates that a questionnaire regarding an arbitrary fictional construct can produce reliable data. His conclusion then, is that the use of presence questionnaires might best be regarded as a means to generate hypotheses, rather than conclusions. In a paper subtitled ‘Why Questionnaires cannot Assess Presence in Virtual Environments’, Slater (2004) gives a rationale for his more recent position regarding (his opposition to) the use of questionnaires for presence. He asserts that one difficulty in asking subjects about presence is that respondents have no experience in making such judgements.

In order to test the effectiveness of presence questionnaires, Usoh et al. (2000) administered both the Slater et al. (1994)(SUS) and Witmer and Singer (1998) questionnaires to one group of subjects who viewed a real-world environment, and to another group that viewed an IVE rendition of the same environment. The hypothesis was that in order to be effective, such questionnaires ought to be able to discriminate between the two conditions, with presence responses being higher when viewing a real environment. Both questionnaires failed to do this, although the SUS showed an elevated response for the real environment for two (out of six) of the questions [26].

One explanation for this was that individuals may reinterpret the questionnaires differently depending upon the real and virtual contexts, so that the presence question regarding ‘being there’ is understood

---

[24] http://www.igroup.com/pq/ipq/

[25] In later studies, further questions are added to the SUS questionnaire. For example, see Usoh et al. (2000).

[26] The results of Slater et al. were also confirmed by a separate study by Mania et al. (2003).

to regard a sense of involvement, as it may appear illogical to ask whether they felt present there when they were indeed in that space.

The conclusion of Usoh et al was that the use of questionnaires is "doubtful for the comparison of experiences across environments, such as immersive virtual compared to real, or desktop compared to immersive virtual" (Usoh et al., 2000).

In contrast to this conclusion, Slater later provides an alternative explanation for this discrepancy Slater (2002), that the cues presented by the IVE may have been sufficient to provide an experience that was just as real (in terms of presence) as the actual, physical, environment.

There have also been two distinct problematic issues raised, regarding the methods used in analysing questionnaire data in presence research, although these criticisms are not limited to the presence per se. The first regards the use of Likert scales, which have been used widely. Gardner and Martin (2007) point out that Likert data cannot be interpreted as being on a continuous scale, and is 'lumpy' (to use their term), and too coarse for techniques such as regression and analysis of variance. They also state their concern over the averaging of Likert data across questions, to produce a single, overall score. Again, this is because the data does not lay on a continuous scale, and in fact it may not even be possible to align ordinal scales to create such aggregate scores at all. Slater and Garau (2007) explain the rationale behind some of techniques used in processing Likert data, and in particular refer to a method long used by Slater, namely, the use of logistic regression. By counting 'high' responses to Likert data, Slater produces a binary measure to each Likert question, and these are instead used for a logistic regression. Although this method loses some of the information it is a conservative approach to handling such data, and generating inferences, especially if results are used for exploratory rather than confirmatory research as Slater (2004) suggests.

The second criticism is regarding the use of exploratory factor analysis. Waller and Bachmann (2006) explain the history and uses of this technique, and go on to suggest that confirmatory factor analyses would be better used in furthering presence research. In their opinion, too much exploratory work has been published, and not enough confirmatory results. We have described the development of two of the main questionnaires in which exploratory factor analysis has been employed, Schubert et al. (2001), and Lessiter et al. (2001). The former in fact used both exploratory and confirmatory analyses. However, it seems that Waller and Bachmann are suggesting that a wide scale comparison (across research groups' work) of presence factors is performed to come to an ultimate conclusion, ideally leading to an optimal presence questionnaire given current knowledge.

Of course, questionnaires are not the only method of eliciting subjective presence responses from subjects. Several alternative methods, have been devised. For instance, IJsselsteijn et al. (1998); Freeman et al. (1999) had experimental subjects use a physical slider to indicate a presence response throughout their experience. This allowed the provision of a continuous value for their sense of presence that could be sampled throughout the trial. Another method is magnitude estimation. This is a psychophysical technique explained by IJsselsteijn et al. (2000) in which subjects respond to stimuli by suggesting a numeric value (such as between 1 and 100) that reflects the 'strength' of the characteristic under investigation (Freeman et al., 2000). IJsselsteijn et al. (2000) also describe a technique termed Cross-Modality Matching, whereby multiple stimuli are presented, and subjects may manipulate some parameter(s) in order to equalise the strength of some phenomenon across the stimuli.

A particular problem with within-group subjective measures is that they sometimes request a judgement between stimuli. Doing this can often introduce bias into their reports, as they then feel obligated to distinguish between them. Such a situation could easily lead to a subject interpreting or reinterpreting a question accordingly, as has been suggested (Slater, 2004; Freeman et al., 1999). Freeman et al. (1999) showed that subjects bias their reports dependent upon prior experience, for instance prior experimental conditions, although the potential for such effects is well known (Mohsin, 1984). Thus it seems that when subjective measures are used, they are best used with a between-groups experimental design, or when the subjective response can be made immediately and without ambiguity, or reflexively (that is, without consideration, such as when relying upon an autonomic response). However, we shall consider the latter to be an *objective* measure.

Techniques that support the implementation of subjective measures include the dual task, where an observer is presented with cues from the real and virtual worlds in an ambiguous manner. Their judgements are then used as an indicator of their sense of presence. An example of this may be seen in Slater et al. (1995), where a real and virtual radio are provided but in two different positions. When the real radio is momentarily switched on (producing a “meaningless tone”), the experimental subjects are asked to point at ‘the’ radio’s position, and the angle, between themselves and the real radio and them and the direction in which they point in, is then used to construct a measure. This measure was found to correlate (significantly) with subjective presence scores according to their presence questionnaire.

Schloerb’s (1995) idea of using a virtual Turing Test presents a subject with either a real or virtual environment (randomly chosen), and defines a measure of presence as the probability that the subject will decide that they are physically present in a displayed environment (the alternative being that they decide they are using an IVE.) By displaying the real environment that has somehow been degraded should allow the possibility that the two conditions are indiscernible. Indeed, the level to which the two conditions must be degraded is potentially an indicator of the ability of an IVE to mimic a real-world environment. Also, Slater and Steed (2000) has had subjects indicate when their sense of presence is ‘broken’, to record these events in time — events known as ‘Breaks-in-Presence’ (or BIPs). Estimates of the ability of an IVE to induce a sense of presence are then computed, according to a Markov model built upon this data.

Questionnaires, however, remain the most popularly used technique for measuring a presence response. The greatest weakness of all such techniques, is the *level* of subjectivity involved. In the next section we shall consider more objective measures.

Objective Presence Measures

Given the disadvantages of subjective presence measures, as discussed in the previous section, many look toward objective measures to avoid them. Objective measures are generally well-defined, and subsequently there may be less ambiguity in the communication of researchers and when inter-relating their work. That objective measures are well-defined could also aid the repeatability of experiments, as they may be less reliant on the opinions of subjects.

If an individual attains a sense of presence in an IVE, then we should expect them to behave as though the virtual environment were real. This idea has been suggested by numerous authors including Sheridan (1992); Held and Durlach (1992); Slater and Usoh (1993a); Hand (1994); Ellis (1996); Freeman et al. (2000); Meehan et al. (2002); Sanchez-Vives and Slater (2005). This should not only be true of outward behaviour (such as gesturing), but also of (the normally hidden) physiological behaviour, both of which we shall consider next.

Behavioural Responses

The identification of natural, behavioural responses as an indicator of presence was suggested early on by both Sheridan (1992) and Held and Durlach (1992). Sheridan (1992) put forward ideas such as measuring presence by responses to (virtual) threatening stimuli, or by eliciting socially conditioned responses from subjects when experiencing an IVE. Held and Durlach (1992) similarly suggested the use of both psychological and physiological devices, such as a 'startle' response when an object suddenly looms [27]. Although a human judge could record when such behaviours occur, it is more reliable and less susceptible to bias to use a strict methodology through the use of some technology, such as the tracking device used by Freeman et al. (2000). Slater has pointed out that subjective reports can omit data that a behavioural measure could pick up, giving an example of a firefighter who stepped back on perceiving an precipice in front of him, but who upon questioning "said that he felt nothing when seeing the pit" (Slater, 2002). Thus a behavioural measure can pick up on events that are subconscious... or perhaps also, events too embarrassing for an individual to report.

An example of behavioural presence is found in the experiment of Freeman et al. (2000), where the reported sense of presence (estimated by magnitude estimation) was found to correlate with the behaviour of the subjects when viewing a visual stimulus. Subjects were shown either a still image of a rally track, or a video from within a rally car as it travelled around a track. The stimuli were presented in different but counterbalanced order, with subjects rating their sense of presence after experiencing each condition. As well as providing a presence rating, subjects' positions were tracked throughout the experiment using a 6-DOF Fastrak ('Flock of Birds') system. The tracking device showed that as the rally car travelled around the track, subjects would move into the 'corners' as they passed by. As expected, this was not found to occur when subjects viewed the still image, and so in this condition less movement occurred. Not only was presence deemed to be stronger when viewing the video (from within the car), but the reported level of presence was greater when the stimulus was displayed stereoscopically as opposed to monoscopically.

Physiology

Perhaps one of the most objective methods for measuring presence could be to utilise some kind of physiological measuring device (or devices.) Physiological responses are well documented, and often

---

[27] (The looming effect occurs when a subject ducks to avoid an object travelling toward them at the level of their upper body or head.)

easier to measure than behavioural gestures. Unfortunately, physiological data is often confounded with phenomenon that is unrelated to the response under investigation: for instance the physical movement of subjects can create transients in such data.

(i) Electrodermal Activity (EDA)

Electrodermal Activity measures have been investigated as to their potential use as a (surrogate) presence indicator (Dillon et al., 2000; Wiederhold et al., 2001; Meehan, 2001; Meehan et al., 2002; Slater et al., 2003; Brogni et al., 2003). EDA may be measured either actively (by applying a small electrical current) or passively (without external current). In active systems it is a Skin Conductance Response (SCR) or Skin Resistance Response (SRR) that is used to determine when an event occurs, with respect to the underlying tonic level: either the Skin Conductance Level (SCL), or Skin Resistance Level (SRL) respectively. An EDA device can indicate stress in a subject (Fenz and Epstein, 1967; Dawson et al., 1990; Andreassi, 2000) and is useful for detecting short transient responses as well as arousal occurring over greater spans of time. Responses to a particular event or stimulus is termed an 'Event-Related' response, whereas those that occur without being attributable to a specific event are termed 'Non-Specific' responses (Dawson et al., 1990). Thus, if a subject is placed in an anxiety provoking situation, one might expect an event-related response as they perceive the situation, followed by further non-specific responses if they continue to feel anxious. In order to induce such arousal some stimulus is required and that will usually require the introduction of an anxiety provoking or novel stimulus. This constraint severely limits the practical use of EDA responses as presence indicators (e.g. to situations in which anxiety is induced, or some unexpected item is presented.) In addition, the movement of a subject and the ambient temperature (for instance, inducing sweating, which could increase the conductivity of the skin) can confound EDA measurements, and so care must be taken to minimise such factors. A further concern when measuring EDA is that individuals are characterised as being either electrodermally labile or stabile. Labiles have a tendency to exhibit a greater number of non-specific skin conductance responses, relative to stabiles or delayed reaction to event-related skin conductance responses (Dawson et al., 1990). This factor must therefore be taken into account when designing an experiment.

Two EDA (skin conductance) analysis methodologies are evident within presence research: one considers the overall SCL (such as Meehan 2001, and the other is based on individual event-related SCRs (momentary events) (used, for example, in Slater et al. 2003.) An (hypothetical) example of an SCR is shown in Figure 3.3. (The figure is adapted from Dawson et al. 2007.) After an evoking stimulus is presented, there is a delay of about 1-3 seconds until the skin conductance increases. The increase then lasts for around 1-3 seconds, after which the skin conductance begins to decreases. The original skin conductance level will not be reached for some time though, as the rate of decrease has in an almost logarithmic character, becoming slower with time. The amplitude of the SCR is stated to be about 0.1-1.0 microSiemens. Given these few rough values, it is not too difficult to filter skin conductance data and subsequently (automatically) identify SCR events. Depending upon the research being carried out, either the overall SCL or SCRs will be analysed, and Dawson et al. (1990) do not favour one method over the other. If one is to only consider the overall skin conductance level (SCL) one should take care that it is not affected by factors such as temperature, whereby a subject in a very warm environment may begin to perspire, increasing the skin conductivity as time progresses. When one uses an event-related skin conductance response this is likely to be less of a problem, as a slow increase in SCL is unlikely to be identified as an SCR. However, movements of the hand or arm that an EDA device is connected to can introduce acute (high-frequency) artefacts, and these could appear as events to software that is designed to automatically detect SCRs. Therefore, skin conductance data should be filtered appropriately: when working with the SCL, a high-frequency (i.e. low-pass) filter should be employed, and when using SCRs a band-pass filter should be used (or at least a low-pass filter).

(ii) Heart Rate

Heart rate (HR) is normally measured using an Electrocardiogram (ECG/EKG). The change in heart rate may be calculated to find how heart rate is affected according to some experimentally controlled variable. As heart rate will differ between individuals, the *change* in heart rate may be computed to somewhat obviate this variability.

Although heart rate may be affected by many phenomena (Andreassi, 2000) it has been employed to measure anxiety where an increase (in heart rate) has been expected to occur. This is exemplified by Meehan's experiments (Meehan, 2001; Meehan et al., 2002) in which subjects were placed in precarious situations that could induce vertigo. In the experiments of Meehan's thesis, heart rate was found to be a more powerful indicator of anxiety than EDA (Meehan, 2001). Heart rate has thus been used in presence research to find an increase in beats-per-minute over a span of time. Numerous other measures of heart rate variability (HRV) have also been investigated in this fashion (that is *before* versus *after* some event within an IVE) — see Guger et al. (2004). However, the heart's inter-beat interval appears to be useful for identifying instantaneous events. This has been demonstrated by Guger et al. (2004) who showed that HR decreased when experimental subjects experienced a sudden break in presence.

(iii) Electrodermal Activity and Heart Rate measures in IVEs

In order to investigate the use of physiological measures in the context of immersive virtual environments, experiments have been conducted by Meehan et al. Meehan (2001) used a virtual environment consisting of a training room, and a room containing a deep pit, to investigate how physiology may be affected by the virtual stimuli. A stereoscopic HMD was tracked to produce an immersive virtual environment, and a physiological device recorded skin conductance, heart rate and skin temperature. In their experiment, subjects learned how to pick up virtual books and move them about within the virtual training room. They were then instructed to take a book into the second room (the 'pit room') and place it on a chair. The chair was positioned at the far end of the room, and could only be reached by navigating a ledge around the room to avoid the pit, or by directly walking over the pit. The pit could be walked across without any consequence, seeming as though it were covered by glass. Meehan found that both skin conductance and heart rate increased significantly as subjects entered the 'pit room', demonstrating that physiology may be affected by virtual (immersive) experiences. Skin temperature did not change significantly. The SUS presence questionnaire (Slater et al., 1994) was administered to subjects, and it was found that reported presence correlated with physiological responses, although in this case the correlation was found to be stronger and more consistent with the heart rate measure than the skin conductance measure.

That physiology may be affected by an immersive virtual environment experience is also confirmed by the results published in Wiederhold et al. (2001). Skin resistance and heart rate were measured while individuals experienced a six-minute virtual aeroplane flight. The IVE was displayed using a tracked HMD, and the system also provided vibratory and audio feedback. Presence and realism were subjectively reported using magnitude estimation scales (that ranged from 0 to 100, the actual questions are not described). It was found that reported presence and reported realism were significantly correlated (negatively) with skin resistance and heart rate, so that greater reported realism and presence were related to greater arousal. Although heart rate was not expected to negatively correlate with increased presence (and realism), Wiederhold et al. cite evidence that the same correlation between heart rate and anxiety has previously been demonstrated.

(iv) Eye Movements

Measurement of eye movements, and in particular, fixations (foveations) and saccades (an eye movement taking the eye from one fixation to another) have been considered for providing correlates of presence, as they ought to reflect visual attention to elements of the virtual environment, and possibly even characterise the overall perception of a virtual environment.

According to Stark (1995) immersive virtual environments can be described as the perception of the virtual world as though real, and we utilise this view within this thesis heavily. To illustrate the relationship between eye movements and visual perception, we may look at the work of Yarbus (1967), who made numerous studies of subjects' saccades and foveations, recording them to produce eye scanpaths. He found that the scanpath was somewhat predictable, by showing the repeated return of the eye to various salient points of presented images. This was further investigated by Ellis and Stark (1978, 1979) whereupon it was found that fixations on 'regions of interest' (ROIs) and the saccades between them are affected by the *perception* of the stimulus.

As regards eye movements and *presence*, Laarni et al. (2003), has suggested the simultaneous use of eye-tracking and heart rate measures to determine where attention is directed and the extent of that attention respectively. More concretely, Renaud et al. (2007) recorded eye movements when using an IVE to investigate whether the self-similarity of eye gaze movements would correlate with subjectively reported presence. He had previously found that the recording of eye gaze can produce self-similar data (Renaud et al., 2003). Renaud et al. argues that the sense of presence is likely to be mediated by oculomotor behaviour, because such behaviours are related to visual perception (see above.) His main measure comprised the determination of the angle between two vectors: the vector from the subject's eye to the point where the subject's gaze fell, and the vector from the subject's eye to a fixed reference point in the scene (which was set directly ahead of the subject's starting position). The virtual environment that was presented consisted of a room containing numerous items of furniture, and subjects were instructed to search for differences between the environment and a previously presented environment (the details of which are not provided.) Subjective presence was measured using Witmer and Singer (1998) questionnaire. The main result of the experiment was that reported presence was negatively correlated with the measure of self-similarity. However, the angle that was computed did not (in itself) correlate with reported presence, but only its self-similarity. Renaud concluded that the complexity of eye-movements, from the perspective of fractal dynamics, is likely related to the level of presence so that as subjects engaged with the environment their eye-movements were more difficult to explain using the self-similarity measure. It could also be argued that when there is low presence, subjects produce repetitive behaviours that are characterised by increased self-similarity, while they tend to ignore the IVE presentation.

Apart from such publications there is a distinct lack of presence research that leverages eye-tracking, although this avenue has been suggested several times (Barfield and Weghorst, 1993; Danforth et al., 2000; Laarni et al., 2003).

2.5 Vision

"The eye is the window of the soul... The eye is the window of the human body through which it feels its way and enjoys the beauty of the world." (Da Vinci, 1452-1519)

Although these words were penned many years ago, we still appreciate their meaning. The eye is not just a passive device, but interacts with the environment through both autonomous and volitional mechanisms. In many ways the eye may be viewed as being part of a visual feedback loop, and part of this loop can be observed externally as the eyes (and their components) move. Whilst it is not known when the eye was first thought to give away information about the observer, there is now a large body of scientific work given to this subject that subsumes many fields of research.

Our interest is of course in the sense of presence experienced whilst viewing an Immersive Virtual Environment. Given this perspective, we could reason that monitoring the eye might allow inferences to be made, regarding when presence is induced. But can the eye provide this information? This question is fundamental to this work.

Consider how the information arriving at ones senses combines (both immediately and over time) to provide a 'complete-picture' of your environment in every day reality. This continuous 'picture' is what most people have lived with since birth. It has been used to build their perception of what it is to be present in the normal, every day, real world. The concepts and presence and perception appear to be inextricable, in that presence is the assumption of a particular perception over another. Before considering perception however, we shall first look at the underlying visual system.

2.5.1 Anatomy of the Eye

To understand how we use our eyes, it is helpful to have basic knowledge of its components and operation. Figure 2.1 shows how light enters the eye (for instance, along the visual axis) through the cornea (a film across the front of the eye), aqueous humour (a clear watery liquid), and the lens situated behind it. It then passes through the vitreous, which is the liquid filling the main chamber of the eye. At the back of the eye is the retina, the central part of which is named the macula, and the most central part of the macula is the fovea. The fovea is the area at which we have sharpest vision. When looking directly at an object, the light from it is projected onto the fovea. Although light is admitted through the pupil it is

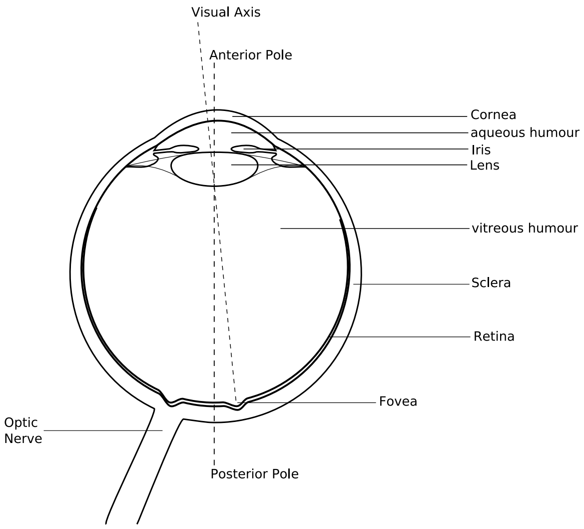


Figure 2.1: Simplified Anatomy of the Eye. (Constructed by the author, from details in Bron et al., 1997; Youngson, 1994.)

attenuated by the iris, which controls the level of light falling on the retina. The lens of the eye changes its shape to focus the light as is passes through towards the retina. The outer (white of) the eye is the called the sclera.

Light falling upon the retina is converted into electro-chemical signals through photoreceptors. There are two types of photoreceptors; rods, which are achromatic; and cones, which are sensitive to specific (bands of) wavelengths of light. The rods, of which there are on average 92 million, are more sensitive to light than the cones (average of 4.6 million), and are predominantly found outside the central fovea (Bron et al., 1997). Measures of acuity and colour perception vary dependent upon light intensity, wavelength and the area of the retina upon which stimuli falls (Buser and Imbert, 1987).

The entire eye is rotated by six muscles, enabling it to move with great flexibility. These muscles are the Medial Rectus (inner), Lateral Rectus (outer), Superior Rectus (upper), Inferior Rectus (lower), Superior Oblique (upper, running obliquely), and Inferior Oblique (lower, running obliquely.) See Figure 2.2.

Together, the two eyes provide stereoscopic vision, allowing depth to be perceived. To achieve this there must be binocular overlap, meaning that both eyes must be able to view the same area of space, and this is true for an approximately +/-50 degree solid angle directly ahead of a person (Bron et al., 1997). Outside this area (to the left of it, and to the right of it) we have only monocular vision.

Each of these components function together to provide sensory information to the brain, to the areas that process the visual signals.

2.5.2 Vision

When the light reaches the retina, it is converted by the photoreceptors into electro-chemical signals, and these are routed to the rear of the brain. Although the brain is not composed of modular parts, general areas have been identified as being sensitive to particular types of stimuli (Atkinson and Braddick, 2003).

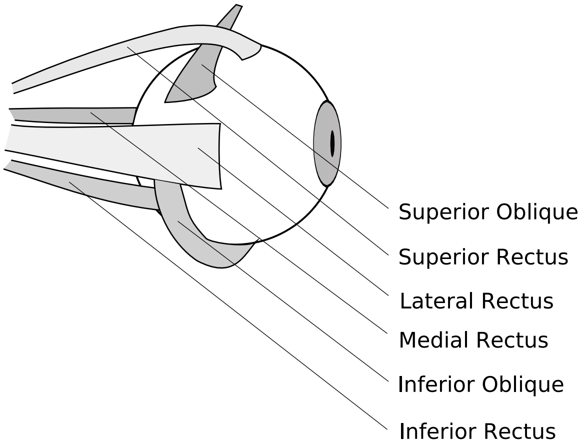


Figure 2.2: Ocular Muscles. Drawn by the author, with reference to Bron et al., 1997.

The main visual part of the brain, the primary visual cortex (V1) is situated at the rear and it is here that the signals from the retina are first processed at a low-level. In V1 some simple aspects of the stimulus are discriminated such as edges, lines, and colours (Ullman, 1996). Along the route to V1, it has been noted that there are more connections in the direction of the eyes, than toward the brain (Gregory, 1998). He notes that this interesting fact could be indicative of the extent to which what we see is influenced by the brain. Moving on from the V1 area, two major paths have been discovered, termed the ventral and dorsal streams. These have been identified as having ‘what’ and ‘where’ aspects of that which is seen (Atkinson and Braddick, 2003). Atkinson and Braddick state the ventral stream is concerned with “aspects of objects, such as form, colour and face recognition”. The dorsal stream on the other hand, is “intimately linked to the eye movements of selective attention”, which as we shall see are central to this thesis. We shall be focussing more on where we look rather than what we are looking at in particular, so with this emphasis, we consider eye related movements next.

2.5.3 Eye Movements and Control

The eye as described has numerous parts that must be controlled. As discussed previously (Section 2.4) attention and interaction are important parts of the IVE experience, without which there is little (if any) presence. Both attention and interaction should be measurable to some degree, using eye and head tracking systems. For any kind of investigation a measure is required, and so we shall first consider types of measurable phenomenon and some of the characteristics that they indicate. We can immediately identify several major types of eye movement, of which we shall describe a subset. This subset, shown below in bold, represents the phenomena most relevant to this thesis:

- Pupil Dilation
- Accommodation
- **Vergence**
- Blinking

- **Saccades**
- **Microsaccades**
- **Foveations**
- Smooth Pursuit
- Opto-kinetic Response
- **Vestibular-Ocular Reflex**

Vergence

When fixating a point in three-dimensional space (or, ‘3-space’), each eye rotates so that the optical axis is aligned with the target point. The ability of the eyes to move together in order to locate such a point is termed vergence. Vergence of the eyes is disconjugate when the eyes move in the opposite direction and conjugate otherwise.

If the line of sight for both eyes is known, we may use our understanding of vergence to determine the approximate location of fixation (in the 3-space.) It is possible that the place of fixation may be estimated from the line-of-sight of just one eye, as it alone will (in an enclosed environment) intersect a fixated object. But the use of vergence overcomes problems such as when there is a partially transparent object in the line-of-sight, and it may then not be known if it is this object, or a more distant object that is being fixated. Additionally, if a subject is ‘staring into space’, be it either looking toward infinity or any other place apart from the intersection of the line-of-sight with an object, then we would likely not know about it without exploiting vergence.

Fixations

Fixations (or Foveations) occur when the eyes focus upon a particular point of interest. This allows the region of interest to be projected onto the fovea. Fixations typically last between 150ms and 600ms (Duchowski, 2003).

Fixations and attention clearly are related. But fixations do not necessarily imply that attention is given to some stimulus, simply because they fall upon it. A second problem for making inferences from fixations is that even when attention is directed to a stimulus being fixated, the percept is not known. That is, we do not know what is being perceived. Both problems may be addressed to a degree by composing fixations and saccades to form eye scanpaths.

Saccades

Saccades are movements from one point of fixation to another. They may be characterised and measured by their direction, speed, duration, latency and accuracy. These movements are planned (having a latency of around 150ms Carpenter, 1977), and ballistic. Hence from initial conditions, measured at the onset of a saccade, their final destination can be determined. Saccades are extremely fast, potentially moving at speeds of over 700 degrees/second (Carpenter, 1977), and they typically last between 10ms and 100ms (Duchowski, 2003). Saccades may under or over-shoot, which provokes subsequent smaller corrective saccades. Whilst a saccade is executed, vision is practically lost (Findlay and Gilchrist, 2003), a phenomenon termed ‘saccadic suppression’, implying that most of the time we see a number of ‘exposures’ between saccades, rather than a continuous flow of visual information.

Microsaccades

Microsaccades are small eye movements that occur even when the eye appears to be in a stable position. These movements are independent of the drift and tremor that also occur when ‘stable’, and are generally smaller in rotation than 1 degree (Engbert and Kliegl, 2004). Microsaccades are important to retain the image on retina. This was demonstrated by Yarbus (1967) when a stable image was presented to the eye of an experimental participant by means of a rubber suction device that attached to the eyeball. This ensured that the image was perfectly still relative to the eyeball and thus retina. Yarbus found that after just 1-3 seconds the image disappeared, and did not reappear given time. This appears to be due to a habituation effect of the retinal receptors. Despite such findings, research into microsaccades is still in its infancy, and they could provide information regarding the cognitive state of a subject pertaining to perception. For instance, it is thought that micro-saccades provide signals that correlate with attention shifts (Hafed and Clark, 2002). To measure microsaccades high-resolution eye movement recordings must be made, which requires specialist eye tracking equipment for this purpose.

Vestibular Ocular Reflex (VOR)

This reflex occurs as the head moves (turns) and the eyes turn in the opposite direction to compensate. A situation in which this reflex is found occurs when examining a static object by moving around it. This appears to be a form of smooth pursuit, in that the eyes move slowly and *smoothly*, tracking a position in space.

2.5.4 High-Level Vision

Having considered the more fundamental parts and workings of the eye, we progress to consider a great body of work regarding vision that is set in a more abstracted plane. Much of this work comes under the heading of ‘High Level Vision’. The main reason for this heading is that the brain is a complex interconnection of neurons, so it is not possible to produce a modular functional map of its workings to help comprehend its mechanisms. Therefore much of the research occurring is abstracted and relies upon making inferences, theories, and models in a more general sense before testing their specifics empirically. This black-box approach means that for the various theories that have been put forward, there is often both evidence and counter-evidence whilst the concepts are refined. Although this is an expected problem when using such an approach, it could also suggest that there is a multitude of mechanisms performing similar tasks simultaneously, either co-operatively, or in competition, or a mixture of the two. Vision may be divided into passive vision (emphasising what is before the eyes), and active vision (emphasising how and where the eyes move).

Passive Vision (‘What’)

As we will be mostly concerned with how and where the eyes move in an environment, we shall only consider vision from an active perspective in any detail. For the interested, passive vision has been theorised over greatly by those such as Marr (1982) (known in particular for his Primal Sketch, and the $2\frac{1}{2}$-D sketch), Gibson (1979) (“The Ecological Approach to Visual Perception”), Ullman (1996) (Visual Routines), and Biederman (1987) (the Geon theory).

Active Vision (‘Where’)

The above theories have a major focus on what is seen, but as mentioned the location (the ‘where’) of that which is seen is also important. For instance, Gibson describes perception as affordances for action, rather than the interpretation of images alone, and this necessarily implies an enviro-spatial aspect of visual comprehension. He maintains the importance of position and situational awareness as part of the visual experience.

Although it might be naively thought that in an instant we see the immediate environment before us as a whole, in fact we require eye movements to re-orient the eye, so that points of interest are aligned with our fovea to view details. To think otherwise would be to suggest that we obtain a complete internal image of our environments. To demonstrate that this is not the case, experiments have been carried out that show a phenomenon called change blindness (Grimes, 1996). Change blindness refers to an undetected change in an image, which has been shown to occur when the change is made while executing a saccade. In an experiment, Grimes showed experimental participants pictures on a computer screen, while using an eye-tracker to detect when sufficiently large saccades were made. Saccades had to be of at least a certain size to be sure of detecting them (the system used could detect such a saccade within just 4 msec.) While a saccade was being executed, the image on the screen would change slightly, and the participant was subsequently asked whether they had seen any change in the image. To the experimenter’s surprise, very few changes were detected, and this held true even when testing with large differences in the images. The images used included a city skyline (wherein a prominent building becomes 25% larger), two men (whose hats are exchanged), and a scene including 30 puffins (of which 33% are removed.) From such experiments it has been concluded that little image detail is retained between saccades. But it was the insight of Von Helmholtz that long before led him to say “We let our eyes roam continually over the visual field, because that is the only way we can see as distinctly as possible all the individual parts of the field in turn.” (Helmholtz, 1925), quoted by Duchowski (2003).

There is an important distinction that must be made between eye movements and attention. It is obvious that a person is not necessarily attentive to whatever is before their eyes, for instance rather than being attentive they may be in some mode of contemplation. It is also possible to attend to peripheral visual areas without moving the eyes, and this type of covert attention has been investigated by Posner et al. (1980). They suggested that attention can be described metaphorically as an ‘internal spotlight’ after finding that it is possible to attend to greater and lesser degrees round about a particular area of the visual field. But although one cannot be certain what is attended at any particular point in time solely from eye movements, it is clear that data from eye movements can be indicative of what is seen, attended and even perceived (as we shall see). Findlay and Gilchrist (2003) point out that some passive vision theorists appear to dismiss eye movements as simply a nuisance factor. Such a perspective can only limit the understanding of vision as they point out, because the fovea only subtends an approximately 2 degree solid angle, and so it must be moved in natural circumstances to see one's environment. Conversely, active vision researchers must bear in mind that while they deal with the overt, there are the passive covert processes supporting vision; of course, both passive and active vision must be understood as working together.

We shall now consider eye movements when initially viewing a scene. When first observing the image of a scene, eye movements are employed to direct attention toward the salient regions within it. We shall consider two methods by which this could be facilitated. The first method consists of obtaining a general but fast perception of the scene's gist. Oliva and Torralba (2006) have demonstrated that at least in some cases, the gist of a scene can quickly be determined from a low spatial-frequency statistical analysis that extracts general and global features as basic forms. The idea of extracting these underlying forms has parallels with Biederman's Geon theory, whereby such forms provide the basis for the brain's modelling of objects. Using this idea as a framework, they successfully demonstrated the potential of the theory by automatically grouping images according to spatial semantic categories. Hence, this type of analysis could be used to glean information from the first glance of a scene for subsequently guiding visual attention. Extending this framework could potentially model vision at multiple levels of detail rather than being only used to derive the overall scene gist. This method could therefore be used to derive an alternative passive vision theory or contribute to those discussed above.

However, a more specific theory has already been proposed by Itti and Koch (2000), that low-level visual processes could build saliency maps to direct attention to salient regions of an image, and this theory appears popular at the moment. As mentioned previously, the process of generating saliency maps would have to be executed quickly, so that eye movements could subsequently use them to scan the image for features. Through experimental investigation, visual search over images has been used to provide evidence that low-level features can be identified in parallel (Treisman and Gelade, 1980). Treisman and Gelade's Feature Integration Theory (FIT) purports that an item having certain uniquely identifiable features may be searched for in parallel across an image, but if the search depends on finding an item having a conjunction of unique features, and the image includes 'distractors' (items with some but not all of the features being searched for) then a serial search is carried out.

A parallel search, or the possibility of an array of search functions, as described by the FIT would complement a visual attention model such as that proposed by Itti and Koch(Itti and Koch, 2000; Koch and Ullman, 1985). Their model provides a description of how an input image could be processed to direct attention to areas that could be described as regions-of-interest. The process operates as follows:

1. Filtering is performed (for colour, intensity and orientation) at different spatial levels.
2. 'Centre-surround' differencing is then used (finding differences between values at different spatial levels) to produce feature maps (respectively for each type of filter that occurred in step (1))
3. The multiple feature maps are then combined into a final conspicuity map (again, respectively for each type of filter that occurred in step (1)).
4. A linear combination of the three conspicuity maps is used to produce the final saliency map.

5. Finally, a ‘winner takes all’ procedure is used to direct attention, though this is intentionally biased by an inhibition-of-return scheme.

Itti and Koch (2000) tested their model to find whether feature-searches and conjunctive-searches could be carried out using synthetic stimuli similar to that used by Treisman. According to Itti and Koch, whilst their conjunctive search failed, the plain feature search appeared successful, finding features almost immediately. The model was then tested using ‘complex natural scenes’ as stimuli. The process was found to provide ‘robust performance’ in finding salient locations within the images. What their research demonstrates is the potential for the directing of movements of attention, including eye-movements, based solely on low-level information. Itti and Koch’s model (Itti et al., 1998) has been further evaluated by Marmitt and Duchowski (2004). They found that when compared to human scanpaths recorded from walkthroughs of virtual environments, the Itti and Koch model performed less well than expected. [28] Marmitt and Duchowski conjecture that this unexpected poor performance is likely due to the dynamic nature of the image, in contrast to the static images used in Itti and Koch’s own analyses. They also suggest that this problem may be surmountable if (memory of) previously viewed images are taken into account, and empirically based data is also included to reflect the finding that gaze appears to be biased toward the centre of images. This criticism is therefore likely to affect others’ similar methods that rely upon static images, such as those of Privitera and Stark (2000) and Parkhurst and Niebur (2005) (see also Oliva et al., 2003; Kootstra et al., 2006; Hwang et al., 2009)

Oliva et al. (2003) has criticised this general approach due to the fact that it tends to be purely visual-feature based and do not take into account the semantic value of the scene and its sub-features. Therefore, Oliva et al. begin to address this problem by not just finding localised salient visual features, but also using spatial-frequency based information taken from across the entire scene to provide context. This spatial information is compared to training data in an attempt to deduce an appropriate context. Comparing their method of attention-prediction with human subjects across sets of images, they found that their approach significantly outperforms a basic saliency approach in predicting fixation points, including the approach of Itti et al. (1998).

When wishing to predict gaze locations Cater et al. (2003); Yang and Chalmers (2005) have taken the approach of providing set-tasks to constrict the pattern of gaze itself. Their ultimate intention is to exploit inattentional blindness (Mack and Rock, 1982) to render only parts of a VE in high-quality. In both papers, a virtual fly-through is presented to subjects that are tasked with counting objects in the scene (either teapots or pencils). When given the task (as opposed to free-viewing), subjects are less likely to detect a difference between scenes that contain lower-quality rendered regions (or objects) versus scenes that are rendered entirely in high-quality. Because of this different approach to gaze prediction, these two experiments seem to largely obviate the aforementioned concerns of Marmitt and Duchowski (they both use moving images.) However, it must be remembered that the issues

---

[28] The comparison was performed using the method found in Privitera and Stark (2000), which is based on string-editing measures (see Section 3.4.3).

described by Marmitt and Duchowski (2004) concern IVEs rather than desktop VEs, and an interactive and proprioceptively based system with a dynamic field-of-view could pose further problems.

Although models for predicting gaze have been somewhat successful, care must be taken as it has also been reported that predictive methods can weaken over time, as gaze begins to wander from the predicted regions to attend “less ‘attractive’ scene regions” (Haber et al., 2001).

Returning to the model of Itti and Koch, which is the undoubtedly the most widely used model of visual attention, perhaps the part of their model that is most consequential with respect to this thesis, is the ‘inhibition of return’ (IOR). This generally refers to a tendency of attention to avoid returning to locations already visited. The phenomenon is defined by Posner and Cohen (1984) after their finding that saccadic reaction times were longer when a saccade was about to be made to a previously visited location, than when the location had not been visited before. They suggested that this longer reaction time could ensure that the next saccade is less likely to arrive at a former location. This tendency has therefore been proposed as aiding visual search by reducing the occurrence of saccades wasted in visiting such locations (being dubbed a ‘foraging facilitator’). IOR is thus a potentially useful function for visual attention. However, the evidence that IOR facilitates visual search is in fact mixed: whilst the idea has been championed by Klein, it has also been in part challenged by him and others (Klein, 1988; Klein and McInnes, 1999; Hooge et al., 2005). Modelling an experiment on Treisman’s work, Klein found that the latency in a saccade that returns the fixation to an immediately prior location was greater than that when saccading to a new fixation location (Klein, 1988). This result is similar to that of Posner and Cohen, and the response variable is also reaction time – as used by Posner and Cohen. But in an experiment where participants viewed various pictures and scenes, Hooge et al. (2005) explicitly tested for return saccades, that is, saccades that were of the same amplitude but in the opposite direction of a previous saccade. Their results indicate that return saccades are not prevented due to IOR. In their own words, this is because “the number of return fixations is higher than may be expected on the basis of chance.” Secondly, while there is evidence that IOR exists in terms of increased foveation time before return-saccades are made, they did not find that the number of return saccades increased when IOR was stronger. Therefore one should be cautious regarding the part of Itti and Koch’s model which could be biased toward type I error when tested; an IOR implementation that inhibits returning saccades in a spatial rather than temporal sense could artificially aid their technique for finding salient regions in images.

In 1954, Attneave proposed that important features of an image were (amongst others) contours such as object edges, and homogenous areas such as solid colours or even textures. His ideas are founded on information theory principles, describing ways that images may be efficiently encoded by the brain, and his line of reasoning leads to a support of and alignment with Gestalt Psychology perspectives on perception. This proposal was tested to an extent by Mackworth and Morandi (1967) by showing 20 experimental participants two pictures and recording their gaze using an ‘eye-camera’. The pictures were divided up into 64 squares. They then had a separate group of 20 participants label the squares according to informativeness, although the term ‘recognizability’ was used for the participants. Thus the participants decided which squares would be easy to recognize again on a scale of 0-9. Their results indicated that the number of fixations upon a particular square was highly correlated to the ascribed informativeness of it. From this it appears that there is truth in Attneave's proposal that informativeness directs eye movements (as measured through fixations), and this could support the Itti and Koch model, as both models rely upon what are described as features. However, whereas the features of Itti and Koch's model have no semantic value, it could be argued that the participants of the Mackworth and Morandi experiment could attribute meaning to the group of features within a square. This would be possible if there was sufficient detail within a square to recognise its contents. More recently, experiments have been carried out that appear to overcome this potential criticism of Mackworth and Morandi's investigation. Both Krieger et al. (2000) and Reinagel and Zador (1999) have analysed statistics of viewed scenes, and found that areas containing high levels of information, in an information theoretic sense, are foveated most. Krieger et al. produced a model from which "corners, junctions, curved lines and edges" were found to be most attractive to gaze, and then to a lesser extent "straight lines and edges" using information-theoretic 'intrinsic dimensionality'. These areas contain such features as "curved lines and edges, occlusions, isolated spots" (Krieger et al., 2000). Reinagel and Zador (1999) found that for the foveated regions of images presented on a computer display, the contrast was greater and the spatial correlation between pixel intensities was lower than for other areas of the images, indicating a greater level of information (entropy) present in those regions.

Another important element of eye movement analysis regards how eye movements and their characteristics change over time. Very early on, Buswell (1935) suggested that fixations when viewing pictures had two modes, firstly facilitating an overview of the image, and then a more in-depth course of observations. Since then, Antes (1974) executed an investigation into this very issue. In an experiment based on that of Mackworth and Morandi (1967, described above), Antes found that over time the number of fixations when viewing an (achromatic) image would reduce, as the time spent fixating grew longer. The pattern of fixation-duration over time, followed a logarithmic curve that appears to fit his experiment duration well. Also, the saccade length grew shorter with time (on average). This reflected the way in which early saccades first moved around to 'scan' an image noting salient regions, and afterward began inspecting local region details before making a saccade to another salient region. His experiment lasted 20 seconds, and approximately 14 seconds after it began, the fixation-duration appears to level out (to about 0.3 seconds per fixation.)

Top-down Perception

The low-level features of the retinal image undoubtedly drive vision, and the mechanisms involved in achieving this are termed bottom-up processes. But they are not alone in stimulating visual perception, as information based upon prior experience is also employed in what are known as top-down processes. Part of the perceptual process involves eye movements, which are necessary to allow the details of an image or scene to be directed onto the fovea. While Itti and Koch (2000) have promoted their model of attention that help demonstrate how eye movements could be directed in a bottom-up manner, it has long been known that top-down information affects eye movements.

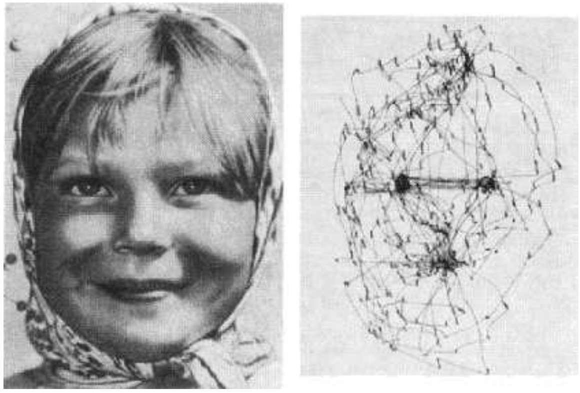


Figure 2.3: Eye scanpath over the picture “Girl from the Volga” as used by Yarbus (1967) (photograph by S. Fridlyand.) Eye Movements and Vision, 1967, originally published by Plenum Press with kind permission of Springer Science and Business Media.

One seminal work that demonstrates the top-down nature of eye movements is that of Buswell (1935). In his work, experimental participants viewed over fifty pictures whilst their eye movements were recorded. To analyse the results, each image was then overlaid by the eye scanpaths for inspection. Rather than randomly or systematically searching the pictures that were presented, Buswell documented how the participants’ eyes are drawn almost immediately and directly to the salient features with the pictures. He found that when viewing the pictures, eye scanpaths are idiosyncratic and repetitive with respect to the observer, and the image itself. His results demonstrated that top-down processes control eye movements to a large extent in order to have them visit meaningful loci within images.

A second seminal work, this time by the Russian scientist Yarbus (1967), demonstrated how priming an experimental participant with different tasks affects eye movements appropriately. By visual inspection, Yarbus confirmed that eye movements were idiosyncratic with respect to the viewer as well as the stimulus using his own set of images, one of the most prominent being shown in Figure 2.3. But more importantly, he also demonstrated that the eye scanpath is greatly affected by differences in the tasks that subjects were set. To do this he had subjects observe a picture (“An Unexpected Visitor” by I. E. Repin, see Figure 2.4) under various tasks, namely (1) Free examination of the picture (2) Estimate the material circumstances of the family in the picture (3) Give the ages of the people (4) Surmise what the family had been doing before the arrival of the unexpected visitor (5) Remember the clothes worn by the people (6) Remember the position of the people and objects in the room (7) Estimate how long the unexpected visitor had been away from the family. Each of these tasks led to apparent differences in the eye scanpaths over the picture.

From the evidence as a whole, it is widely believed that both bottom-up and top-down work together in order to fully perceive an image (Itti and Koch, 2000; Ware, 2008).

Hypothesis Testing

As we have described, eye scanpaths could be used to provide us with information that allows us to investigate how images are perceived and recognised. In perceiving the content of an image, it has

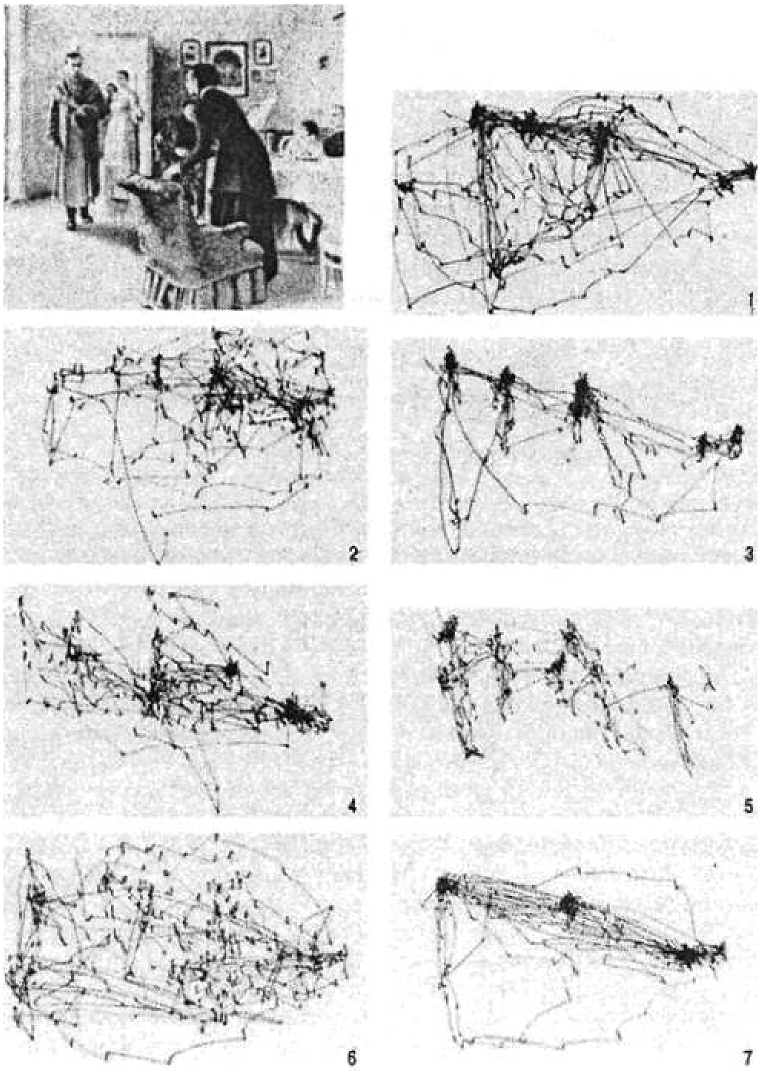


Figure 2.4: Eye scanpath over the painting “An unexpected Visitor” as used by Yarbus (1967) under seven different tasks (Painting by I. E. Repin.) Eye Movements and Vision, 1967, originally published by Plenum Press with kind permission of Springer Science and Business Media.

been suggested that a process termed Hypothesis Testing occurs (Gregory, 1998). Here we shall only consider the visual sense, though this general theory could potentially be applicable to other senses such as touch. Hypothesis Testing is a cycle of Bottom Up (BU) and Top Down (TD) (Stark et al., 2001) information gathering that seeks to determine the nature of a sensed image(s). In this, the brain takes sensory information originating from some stimulus and produces various hypotheses to match higher-level conceptual models. The Bottom Up information is that taken directly from the physiological senses. The Top Down information is taken from the cognitive models, knowledge, and memory (which admittedly are probably not separate entities.) Thus, using both sources of information hypotheses are generated, and the best match is accepted as the perception (or best explanation) of the sensed image(s). As further evidence appears from the stimulus, hypotheses are re-tested and are re-generated, so that this is a continual process. This theory of Hypothesis Testing could explain the repetitive nature of the eye scanpath as shown by Yarbus (for instance, see Figure 2.3.)

Hypothesis Testing has been suggested to be the reason that IVEs are perceived as actual environments even though they do not provide the rich content that we see in the real world Stark (1995) (see Section 2.5.5 and Chapter 3, Sections 3.2 to 3.3). This is because IVEs would then provide at least sufficient cues to trigger (or select) the hypothesis that we are actually in the place depicted by the virtual environment. In our first experiment (Chapter 4) we test this idea by introducing visual cues to an IVE slowly, whilst watching for this ‘trigger’ to occur (when the IVE is perceptually selected) by monitoring eye-movements.

Hypothesis Testing theory may be useful in investigating how the brain classifies images (Stark et al., 2001), and could shed light on which parts of an image are deemed most important for perception. Indeed, Noton and Stark proposed a perceptual theory that was based on eye scanpaths and hinged on Hypothesis Testing (Noton and Stark, 1971; Stark et al., 2001). Their theory set out that internal cognitive models drove the eye movements that produced the scanpath, and that these in turn drove perception. Their idea was that as eye movements traced out the scanpath, this would trigger hypotheses. This has implications regarding the relationship between eye scanpaths and perception that are of practical importance, because it then appears that they are tightly bound together (also see Section 2.5.4). Their theory continued in this vein such that if no perception occurred when viewing a stimulus, then a new cognitive model would be constructed (i.e. learned) (Noton and Stark, 1971). One experiment showed how eye movements followed the expected scanpath when simply remembering an image, and this seemed to provide evidence supporting their theory (Choi et al., 1995). Although some experiments of Stark’s have had success, the bold overall theory of Noton and Stark has been controversial and not widely accepted although it is often mentioned in the comprehensive texts that describe perceptual theories. A major opponent of the theory was Viviani (1990). He strongly argued that although eye movements and cognition are mutually influenced (to use his terms), the internal mechanisms of perception could never be inferred from eye movement experimentation. Perhaps his greatest objection to scanpath theory is that images may be perceived in extremely brief periods (e.g. in around 125ms, Potter, 1976), which is too short to allow sufficient eye movements to occur, although perhaps covert attention might be argued to execute more quickly. Additionally, in Noton and Stark's (1971) experiment, the scanpath evidence only appeared within data from 65% of their participants, and this figure was only 50% of participants in a study carried out by Locher and Nodine (1974). These are low figures when the theory should play a central role in such a fundamental and ubiquitous process as visual perception.

In this thesis we do not utilise Noton and Stark's Scanpath Theory itself, partly because it is felt that the theory does not have sufficient consensual support from the research community, but also because it is not necessary to show how the 'perceptual engine' works for our purposes. Using eye-movements in order to discover how the 'perceptual engine' works was also Viviani's prime criticism of the Scanpath Theory. However, we shall use as a premise the finding that specific perceptions are reflected by eye movements, even when the visual stimulus is the same. This premise is evidenced by research (Ellis and Stark, 1978, 1979) that may be viewed as either part-of or weakly related-to the Scanpath Theory, but it is certainly not sufficient to support the Scanpath Theory. Whereas the Scanpath Theory prescribes a specific control mechanism between eye-movements, perception, and cognitive models, we find sufficiently useful the finding that eye-movements and perception appear to be tightly bound (see Section 3.3). Viviani's second criticism of Scanpath Theory is that recognition of images has been shown to occur too quickly for the theory to be valid, because an insufficient number of eye movements could have taken place in this short time, as explained above. Again, in this thesis we shall not be assuming that eye-movements lead to perception per se, but rather that they become more deterministic once a perception is settled upon because they appear at least then directed by a cognitive model (see Chapter 3). Therefore, it is of no consequence that the Scanpath Theory may or may not be valid in itself, only that a specific perception and eye-movements are linked as per Ellis and Stark's follow up work. In fact, in our first experiment we slowly introduce a visual environment such that we expect subjects to suddenly perceive it as a meaningful environment at some moment in time. It would seem that the faster the perceptual switch occurs, between not-perceiving and then perceiving the environment, the less support there appears for the Scanpath Theory[29]. However, as shall be seen, our methods do not have sufficient temporal resolution to completely rule out the claim of Scanpath Theory.

We therefore build on the findings of Ellis and Stark (1978, 1979), that eye movements and perception are related. This is because as an extension of their finding, eye scanpaths could potentially be used to indicate the perception of not just images, but even environments. In this case of environments the idea may be justifiable by considering them as providing a single surrounding image. In fact, rather than a scanpath per se, Choi et al. (1995) coined the term 'searchpath' to refer to a path that is idiosyncratic with respect to a spatial map (as opposed to a cognitive model related to stimulus perception). The term scanpath was used to describe the eye-gaze path over some stimulus such as an image or object which is (for instance) segmented from a scene. However, the term searchpath was used to describe the eye-gaze path over an entire scene itself, generally between the objects or regions within it. Also, the

---

[29] Hypothesis Testing as an explanation for perceptual selection seems viable with or without acceptance of the Scanpath Theory. The Hypothesis Testing theory does not dictate the exact relationship between eye-movements and perception, nor does it dictate an upper bound on the time within which perception must occur.

searchpath was, at least originally, used in the context of the observer performing a search task (whereas the ‘scanpath’ only concerned the recognition or perception of images or objects, mainly within the Scanpath Theory literature described above). This was used when they investigated the searchpath of a stereoscopic image on a screen, which can be thought of as a window from their (real) environment into another (the virtual one). In this body of work though, we shall consider searchpaths in an all encompassing, three-dimensional and stereoscopic environment. In fact, we prefer (and use) the term scanpattern, which is more generic and does not imply the use of an explicit search (Henderson, 2003). We shall also be constraining our scope by considering static scenes only, although eye scanpaths in dynamic scenes have been investigated to a degree (Blackmon et al., 1999).

2.5.5 Perception in an Immersive Virtual Environment

Upon entering an IVE one or more of your most compelling senses are presented with an alternative view of your physical-local environment whilst your real-world environment is shut out (as much as it is possible.) It seems that because the perceptual system is trained over many years to recognise our ‘everyday reality’, it thus has no experience to distinguish it from a ‘virtual reality’. So perhaps it is not too surprising that the mind in attempting to subjectively classify the IVE experience identifies it as being our actual, physical reality. It could be said that this is what presence research is really concerned with, that is, at what point does the mind believe in the presented stimuli as being the reality.

It was Lawrence Stark who said “Why does virtual reality work? Because reality is Virtual!” (Stark and Choi, 1996), and in “How Virtual Reality Works” (Stark, 1995) he theorises that we do not require great detail from virtual reality systems because even in reality, most of what we ‘see’ around us exists as a model, in our minds. The perception of our immediate environment is therefore based upon little actual sensory information, and is for the most part illusory; an IVE simply provides cues that are a sufficient match for our inner conceptual models of what it is to be in an environment. Ironically, we rely upon our perceptive system so much that it enables us to be fooled. For instance, stage magicians rely upon this fact by providing basic cues that purposely misinform our perceptual system and leave us wondering how we have apparently jumped from one world-state to another. Just as when we see an illusion and are able to accept the perhaps ‘odd’ perspective that is implied, when we view an IVE we can accept the virtual world perspective implied over the real world.

When visual cues are strong enough, we can find it extremely difficult to resist seeing an illusion, and we suggest that this can be true of a presented IVE. Strangely, such perceptions may occur even though the detail that normally exists in a real world environment is missing, and the IVE is imperfectly presented. Gregory’s thesis (1977; 1998) provides a perspective on this issue, with his statement that perception occurs when low-level information is subject to hypothesis testing based on previous experiences. This again implies that IVEs ‘work’ because they present a closer match to an immersive environment than other plausible explanations for the presented information.

This inconsistency, between the actual stimulus and that which is perceived, can also be framed in the concepts of Gestalt Psychology as conceived in 1911 by Max Wertheimer. Gestalt Psychology sought to shake the foundations of the existing theories of perception (structuralism) by posing problems that could not be explained using the existing frameworks (Goldstein, 2002). It presented several principles that are often cleverly demonstrated pictorially, and that are known as the Gestalt Laws of Perceptual Organisation. One of the most important Gestalt principles, is that the whole is not equal to the sum of the individual parts, meaning that a composition of elements has meaning in itself and as such it is not possible to break this down into components without changing the character of the whole. This view was adopted by Slater and Steed (2000) who proposed that presence is a binary state and so an observer at any moment is either present in the IVE, or in the real world. This important Gestalt principle suggests that perception is very much an all-or-nothing process, in that either one perceives a stimulus, or they do not. In fact, such a perception (that is perceived as a whole, or not perceived at all) is termed simply a Gestalt. Interestingly, Koffka (1935) described a specific Gestalt principle that he called maximum-minimum simplicity. He proposed that a reduced amount of available energy in the brain would lead to the perception of less detail, and more uniformity. If correct, then this point of view might be used to explain how the development of percepts could converge to those previously experienced because such a process would require a minimum amount of energy. Chertoff et al. (2008) argue that “memories, processes and behaviours are guided by schema”, and so again rather than there being a piecemeal construction of the perception of (or sense of presence within) an IVE, this would be described as a schema matching problem. This understanding of how IVEs work according to Gestalt concepts has, potentially, further support in the form of schema matching. In the work of Mania et al. (2005) subjects that had previously viewed an IVE, reported seeing objects that were consistent with the scene presented but that did not in fact exist in the IVE model. Objects that were inconsistent with the scene appeared more difficult for subjects to remember, suggesting that the schema used to perceive or remember the IVE made recall either impossible or more difficult respectively. Rendering quality had little effect on recall, possibly being evidence of its less important role in comparison with overall schema matching.

IVEs exploit this perceptual resilience, and an explicit example of this exists in the use of level of detail (LOD). In order to reduce rendering time, complex objects in a virtual environment are often represented by separate models each at a different level of detail. This allows a simplified version of an object to be rendered when greater detail is not required, such as when the object is at a distance and the details would not be discernible to an observer (Reddy, 1997).

As described in Section 2.4.2, rather than taking a geometric perspective, Zimmons and Panter (2003) and Mania and Robinson (2004) investigated the quality of rendering to see whether improvements would affect the sense of presence experienced in an immersive virtual environment. Neither Zimmons and Panter nor Mania and Robinson found a significant effect, suggesting that a perceptual Gestalt could be behind the sense of presence because once this state is achieved improving rendering quality is no longer of import.

Although the perceptual processes are very robust, to instil a sense of presence, one’s perception of stimuli presented by a virtual reality system should map to one’s normal perception of reality. Incoherencies such as loss, lack, or imbalance of information could - if noticed - reduce or even nullify the sense of presence. Given certain types of misinformation the perceptual system may find itself simply unable to comprehend its inputs, as they are just too inconsistent. As an example, this can happen when viewing a three-dimensional picture using a stereoscopic device. If the interocular distance (often a parameter of stereographical systems) becomes too great, then instead of visualising a three-dimensional image, an incoherent double image is viewed. This phenomenon is called diplopia, and is described in the context of virtual reality in Ware and Balakrishnan (1994).

2.5.6 Gaze in Three-Dimensional Scenes

Although we have looked at how eye movements are influenced by static images and then provided an explanation for how we perceive IVEs, we now combine these topics to consider eye movements as studied in more complex scenarios. Specifically, we proceed to look at areas of research concerning eye movements in three-dimensional scenes (known more concisely as 3-space).

When moving from two to three-dimensional studies we must consider head position and direction (Danforth et al., 2000), which accounts for an extra six degrees of freedom. This can either be achieved by directly measuring the head position and direction in-situ (Duchowski, 2003)[30] , or it may be implicitly achieved by simply recording the view before the observer at each instant (with a video camera), and then synchronising and superimposing the video with the eye-tracking data to find the areas fixated (Pelz et al., 2000). One of the great advantages of using the former method is that we may investigate experiences and tasks as they occur (more or less) naturally in the real world or in (immersive) virtual worlds, rather than only examining responses to flat images and pictures. An overview of how gaze is used in natural environments may be found in Hayhoe and Ballard (2005), including work carried out in both virtual environments and the real, physical, world. Examples of the experimental tasks include moving toy bricks (Pelz et al., 2001b), pedestrian collision avoidance (Jovancevic et al., 2006), both in virtual environments, and sandwich making (Land and Hayhoe, 2001; Hayhoe et al., 2003) and hand washing (Pelz et al., 2001a) both in the real world. In these examples, just as with two-dimensional stimuli it is found that gaze is largely governed by task, but it is also directed by learned visual routines[31] and reward-based neural circuitry which may be related to the finding that eye movements over 2D stimuli follow idiosyncratic paths dependent upon the image stimulus. We therefore suggest that when using 3-space and given a task, there will continue to be some deterministic process that drives the eye scan-pattern. The structure in these paths could contain information relevant for developing response-based measures for presence research.

Although the progression into 3-space investigations broadens the application domain for eye-movement studies, there are also advantages due to the larger space over which the stimulus is presented. This allows a wider spatial spread of the individual elements of the stimulus (i.e. scene), which in turn means that for some applications a lower resolution of gaze measurement is sufficient to record the regions in space where gaze falls. In fact, Land (2004) has found that eye movements account for smaller shifts in gaze (generally, < 10 degrees) whereas larger shifts in gaze are generated by the head and trunk of the body. It therefore seems that head-tracking equipment can be used to analyse gaze without re-

---

[30] which in the case of a real environment will require registration of the head-tracking system with the real environment.

[31] not to be confused with the differently defined ‘visual routines’ of Ullman (1996).

course to eye-tracking devices in some cases, though this is largely dependent on the accuracy required for each specific application. However, although gaze and head movements are coupled to some extent this association should not be taken for granted in general. Biguer et al. (1984) conducted an experiment in which participants pointed to targets while their eye, head and hand movements were recorded. They found that although head movement followed gaze, it only followed for approximately $\frac{1}{2}$ to $\frac{2}{3}$ of the rotation, with the eyes accounting for the remaining rotation of the gaze. More specifically, for targets at 10, 20, 30, and 40 degrees, the mean head rotation was only 5, 12, 19 and 25 degrees. Having said this, the participants were sat down for the task, and were under pressure of time — being told to point at the targets as quickly as possible, which were displayed for 5 seconds with an inter-stimulus interval of 5 +/- 1 second. These differences may explain the discrepancy between this study's findings and that of Land's. Specific work on the dissociation of eye from head movements has been shown by Ron and Berthoz (1991) although their work appears to frame this phenomenon as an exception to the rule rather than being part of the 'normal' visual routines. Examples of head tracking being successfully substituted for a composite eye and head tracked system are provided by Ohshima et al. (1996); Watson et al. (1995). Zimmons (2004) used head-tracking alone to determine the length of time that experimental subjects spent looking at various areas of a virtual art gallery. Although subjects experienced varying lighting conditions, subjective presence scores (Slater et al., 1994) did not reveal any effect of lighting on presence. In measuring attention however, Zimmons found that there was a significant difference between areas of the gallery attended dependent upon the "contrast ratio of the lighting configuration". Additionally, he was able to determine that there was a significant difference between the times that objects (two paintings and two vases) in the gallery were attended, again, dependent upon lighting condition.

2.6 Summary

In this chapter we have described immersion, presence, and perception as interrelated concepts of key importance to the IVE experience. Some of the general technologies that are used with IVEs have been briefly described, and the relationship between types of VEs (in a broad sense) has been mentioned. A number of techniques that are used for the measurement of presence have been discussed, and we have described the areas in which they fall short.

The definitions of presence most appropriate for this work are: presence as a Response As If in the Real world, known as RAIR (Sections 1 and 2.4.4) and 'presence as transportation' (Section 2.4.1). The latter is useful in an informal sense, because it relays the concept of presence in simple terms. However, the former, definition that frames presence in terms of a response is operational, and is the definition we must relate to in any formal discussion, such as the remainder of this thesis.

We have seen that eye movements may be analysed in both two and three-dimensional scenarios, and that they shed light on how and what an observer sees. We have also briefly described numerous perceptual theories that could aid us as we go on to develop the methods and strategy for analysing responses to IVEs.

Chapter 3

Methods and Strategy

3.1 Introduction

In this chapter we develop the general strategy and methods that will be used in the investigations presented in the subsequent chapters of the thesis. Section 3.2 gives an overview of the ideas and research that underlie this thesis and its approach. In the following Section 3.3 we consider research to be leveraged for our investigation purposes, specifically, research that brings insight into the relation between scanpaths, searchpaths, the scanpattern and perception. Section 3.4 looks at two common ways in which scanpaths and searchpaths are modelled; measures of the selected model are examined in Section 3.5. Section 3.6 to Section 3.8 describe remaining concepts and methods that we shall need to carry out our investigations. The chapter is closed with Section 3.9, which provides a summary.

3.2 Overview

This thesis investigates visual IVEs because visual systems are the most widespread type of IVE. As is evidenced by the equipment itself, the visual system of an IVE is a closed loop. The displays provide the input to the human visual sense, and yet head movements change the content of the displays. Similarly, the image in front of the eyes is projected onto the retina, but because detail can only be obtained by projection onto the fovea, muscles move the eye, and this may also be viewed as a feedback mechanism. While we cannot monitor the thoughts and cognition to ascertain what is seen, we may monitor these feedback control mechanisms, i.e. head and eye movements. From these we can make inferences and take measures.

As has been described in the previous chapter, eye and head movements are not random (Section 2.5.4). The eye scanpaths and searchpaths have been shown to be repetitive and idiosyncratic, being dependent upon the stimulus and task of the observer. In this thesis we exploit this knowledge, but also go further by taking into consideration research that relates eye movements to perception, to be discussed in the following section (Section 3.3). Although eye and head movements are not truly random, they do appear to be probabilistic. Because of this we may model movements of gaze as random processes, and these can be characterised using information theory, and specifically entropy. Entropy may be thought of as describing how deterministic such processes are, or alternatively how much ‘information’ we may

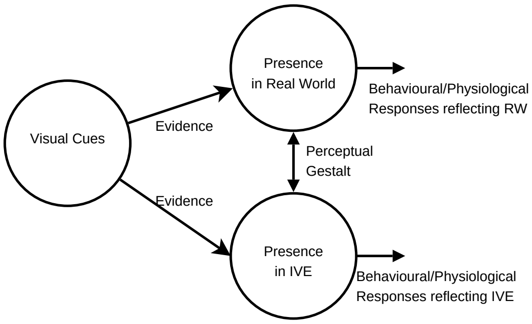


Figure 3.1: Visual cues, leading to a state of presence.

obtain from them. Such analyses will hopefully provide insights into how an IVE is being perceived. The basic concepts of information theory and entropy shall be explained in the later sections of this chapter.

In the presence research literature it has been suggested that there is a point at which a minimal number of visual cues exist that are only just sufficient to induce an intended perception and sense of presence in an IVE (Slater, 2002; IJsselsteijn, 2002; Mania et al., 2005). An IVE presents an alternative perception of a person's environment, and it is a phenomenological acceptance of this alternative that may be thought of as presence that leads to RAIR. Because it is a state of mind, it is difficult to define the point at which this perception arises.

Within this thesis, we hypothesise that when the sufficient visual cues (i.e. the minimal cues) are presented, just as in a common illusion, a subject may experience a sudden Gestalt (see Section 2.5.5). This Gestalt represents a switch to perceive the presented illusion of the IVE as reality (Stark, 1995; Slater, 2002). This is illustrated in Figure 3.1.

However, this introduces another problem, which is determining when such a Gestalt might have occurred. In an attempt to obtain evidence to indicate when this might occur, we shall monitor behavioural and physiological responses including gaze scan patterns. We shall expect the gaze scan patterns to reflect the findings of a number of researchers including Buswell (1935); Yarbus (1967); Ellis and Stark (1978, 1979) (see Section 2.5.4). Using the findings of Ellis and Stark; Ellis and Stark in particular, we hope to be able to discern differences in what is being perceived over a specific period of time. If a switch in perception is found to occur at the same time as presence in the IVE is induced, then we would have evidence that such a point of Gestalt does indeed exist. This would help link the concept of 'a sense of presence' to a discontinuity in perception, and this link is fundamental to the concept of minimal visual cues. In order to detect this switch, we shall employ a measure of gaze entropy.

To investigate people's responses to IVEs using gaze entropy, part of our strategy is to identify the presence state of a subject while varying the number of cues presented (Experiment I). A means for the introduction and removal of visual cues will be required, and although numerous methods could be used or invented to do this, we shall introduce cues randomly. This is so that any subsequent findings are not biased due to a systematic method of cue introduction, such as by the introduction of cues using a level-of-detail algorithm. Methods for introducing visual cues are certainly of interest as regards presence research, but to include an investigation of their effects is not feasible within the scope of this thesis.

Not only can we monitor responses in order to detect changes in a subject's perceptual state, but we may also compare responses to those elicited by real world environments. If these are similar, then we have evidence that the IVEs are being perceived as though they were real world environments. To investigate this possibility we may be able to compare perceptions as they are reflected by a person's gaze, as explained by the Hypothesis Testing theory (Section 2.5.4). This has been implemented using an information theory measure that reflects the rate of information arising from observers' gaze when presented with both a real environment and a virtual depiction of that environment.

Since changes in the state of presence are described herein as distinct switches, or perceptual Gestalts, there should be a threshold in the sense of a minimal set of visual cues for presence in an IVE. But in addition, once a state of presence in an IVE has been achieved then improving or adding to the realism of the environment should not 'increase' a presence response. This is because an increase no longer has meaning under this model of presence. This shall be investigated in Experiment II.

3.3 Eye Scanpaths, Searchpaths, and Perception

The eye scanpath provides a rich data source to which numerous methods of analysis may be applied, and is superior to fixation-only analysis. In an IVE we know what is displayed before a subject's eyes at any point in time, but also, with an eye tracker we know the area of the display being looked at. One can determine the most frequently visited points and the Regions of Interest (ROIs) by computing the line-of-sight and relating it to the IVE to determine whether the main features are viewed, and even that artefacts are not foveated for any significant amount of time. This bears a resemblance to an analysis of attention based only on fixations.

A major problem for presence research is that a subject may be attending stimuli other than that being presented or, more importantly, perceiving stimuli in an unintended manner. Perhaps the ultimate concern is that a subject appears to interact convincingly (in the eyes of the experimenter), but in fact did not perceive the IVE as intended [1]. There are many ways in which this situation can occur. Perhaps the most obvious is when a subject notices rendering artefacts, which is a concern for IVE designers. One might think that these situations could be detected by monitoring where the gaze falls within the IVE. For instance, if the gaze falls upon an area of the IVE display that does not correspond to a salient region within the IVE, then it is less likely that the subject is engaging with the IVE. It is possible that they are in fact observing something as part of the real world environment. But if we consider the converse situation, then when such a point of gaze is salient with respect to the IVE, then they could either be perceiving the real world environment (and thus a point which is relevant to the IVE only coincidentally), or as hoped, by perceiving the IVE. Because of this problem, our objective is to find some way in which we could statistically determine the likelihood that the subject was present in the IVE. As we shall now see, this problem is addressed to an extent by considering the sequence of (or the dependence between) fixations

---

[1] A common occurrence of such a problem has been exemplified when utilising a type of 3D shutter-glasses that turn on when opening up the arms. It has frequently been known for the uninitiated to view stimuli for extensive periods, without realising that the glasses are not turned on (thus seeing 2D double images) because these persons have not experienced such displays before, and they do not know what viewing experience to expect.

rather than the location of fixations only, which includes information regarding the recurrent foveations and recurrent saccades to the same location (Yarbus, 1967; Duchowski, 2003).

Once identified, ROIs can be labelled and the order of transitions be analysed to produce the gaze scanpattern. Perhaps surprisingly, Ellis & Stark have provided evidence that from an eye scanpath one can determine how a stimulus is being perceived as well as simply what is seen. This is obviously of major significance as this is a very similar problem, if not the same problem, to that which all presence research has pursued to date, that is, ‘how is the subject perceiving the IVE?’. Ellis & Stark showed how the perceptual state manifests itself through the use of ambiguous figures, by pointing out clear distinctions in the transitions and ROIs of alternative perceptions, given a single stable stimulus (Ellis and Stark, 1978, 1979). In their experiments they had a subject look at a particular stimulus and recorded the foveations and saccades that occurred over ROIs, stringing them together to create an eye scanpath over the image stimulus. Therefore as the perception switched, so did the scanpath. It was already known and documented much earlier (Yarbus, 1967) that after being given a task, a subject would view a visual stimulus in a stereotypical way. However, in addition they themselves had discovered that the eye scanpath was also determined by the subject’s perception of the image stimulus. One of the ambiguous images that was used is shown in Figure 3.2. In the paper itself it is overlaid with different scanpaths according to the perception held by the observer. The scanpath is thus shown to reflect how the subject perceives the stimulus at a high-level at some point in time. These switches in perception can occur over a time period in the order of 2-3 seconds as shown by Pelton and Solley (1968) when they conducted an experiment that had participants switch perception as fast as possible. This provides an upper limit on the speed that perceptual switching can occur.

3.3.1 Searchpaths

Later in this thesis we study the hypothesis that it is possible to differentiate between perceptions of a subject solely by studying the scanpatterns that a person’s gaze follows, and so our context and constraints differ somewhat from the original research that inspired our use of gaze scanpatterns. In Choi et al. (1995), a searchpath was defined as a path that is idiosyncratic with respect to a spatial map (as opposed to a cognitive map, as is the scanpath.) Thus, the form of a searchpath is much the same as the scanpath, but covers a wider spatial area and so is likely to include head-movements, whereas one would be hard-pressed to find a scanpath that included head-movements whilst retaining its integrity. Thus our use of gaze scanpatterns is akin to the searchpath (but without the search task implied by this name.)

In the first experiment, we hypothesise that there is a ‘way’ of looking around an environment that differs depending upon whether it invokes a high-level perception. Hence, we are interested in environments rather than images (be they mono- or stereoscopic images), and we are also linking the onset of a high-level perception to a switch in the presence state.

In the second experiment we compare the scanpattern response between a real world environment and a representative IVE. This is important to show that an IVE is viewed similarly to a real world environment. Also, in the same experiment we consider whether IVE scanpatterns become more like

Figure 3.2: Triply Ambiguous Figure, ‘Mother, Father, Daughter’ by Fisher G.H. (1968) Utilised by Stark & Ellis (1981). From THE AMERICAN JOURNAL OF PSYCHOLOGY. Copyright 1968 by the Board of Trustees of the University of Illinois. Used with permission of the University of Illinois Press.

those found when viewing a real world environment, when the IVE has more realistic rendering (using radiosity) applied.

3.4 Scanpattern Modelling

Searchpaths (and scanpaths) are not only made up of a collection or set of regions, but also the sequence of these, as its name suggests. In order to analyse such sequences we must select ways in which to model them. This selection is dependent upon our goal(s) and thus our hypotheses. Although there are many models available that could be used, we shall limit ourselves to discuss only those utilised in later chapters (the experiments), whilst briefly mentioning alternative models that were potential candidates but were discarded for some reason. This discussion is presented in the remaining sub-sections of 3.5.

In instances such as ours the analysis of scanpatterns first requires the capture, transformation and composition of eye-tracked and head-tracked data in order to obtain the Point Of Gaze (POG). Secondly the regionalisation of the image space into ROIs is needed. These two requirements are discussed in the following subsections 3.4.1 and 3.4.2 respectively.

3.4.1 Point of Gaze from Eye and Head Tracking

When eye-tracking is over a two-dimensional stimulus display, we are often able to linearly map measurements from eye tracker coordinates to stimulus coordinates. This is because non-linear and linear transformations are routinely dealt with by eye-tracking hardware/software, given the necessary calibration data. When using an IVE system with an eye-tracking device, we typically utilise a six degrees-of-freedom head tracker, and we therefore have a few additional considerations. We must simultaneously measure both the head and eye positions and directions. Composing these we may compute the Point of Gaze (POG), which is the three-dimensional point in virtual world model coordinates upon which the gaze falls.

Because an IVE utilises a three-dimensional model to determine what is displayed before a subject, we require head-tracking data that is transformed to provide the ‘camera’ coordinate system. This coordinate system is used to project and subsequently render the model onto the display(s). By recording this head-tracking data when a subject is viewing an IVE, we may thus determine what is or was rendered onto the display at any time.

Once the camera’s details are known, we must also know the characteristics of the rendering frustum, used in the projection transformation. From this information we can calculate the position of the centre of projection (COP). The two-dimensional coordinates from the eye-tracker are used to calculate the deflection of a ray, which initially passes from the COP into the centre of the display. By doing this we have generated a ray representing the subject’s optical axis in virtual world coordinates (i.e. model coordinates). By intersecting this ray with objects in the virtual environment, the point of gaze and the object being viewed may be found. Similar methods and the details of such transformations are detailed in Duchowski (2003).

It is important to also consider a method to distinguish foveations from momentary samples of saccades. Without some kind of filtering many false foveations would be recorded. Fortunately this can be achieved simply by setting a temporal threshold after which a foveation will have been deemed to occur if the POG has remained within some specified region. This method is known as a ‘position-variance’ scheme, but there are alternative algorithms that may be implemented. For the interested reader Duchowski (2003); Salvucci and Goldberg (2000) provide details of these methods, which are of particular use when real-time saccade tracking is required. For instance, velocity-based saccade detection can not just detect the onset of a saccade, but can allow the prediction of where the gaze will land when the saccade terminates.

3.4.2 Determining Regions of Interest

A scanpattern is generated based on the point of gaze as it visits a number of ROIs. In many studies the ROIs are pre-determined by the experimenter, specifically selected to contain the salient elements of the stimulus. In others they are often defined post-hoc, that is, after it is known where the gaze dwells upon the stimulus. Although the latter approach has been performed manually, it could be done automatically and objectively using algorithmic clustering techniques (Everitt et al., 2001; Privitera and Stark, 2000; Itti and Koch, 2000).

An alternative approach to defining ROIs post-hoc, or subjectively according to salient points of the stimulus, is to subdivide the environment into equal areas/volumes. We shall take this approach because we wish to objectively measure characteristics of the scanpattern without the need for human intervention, and because various models formed may then be easily compared. We therefore divide a three-dimensional environment in a pre-determined but unbiased and regular fashion. The scene is thus segmented into a number of congruent (i.e. equal shape and size) elements that together will capture the entire environment.

Searchpath ROIs normally contain the main components of each object in the scene, which may be said to exist mainly at an inter-object level. For identifying *scanpaths*, saccades and fixations at the intra-object level for an IVE scene would prove difficult if not impossible. This is firstly because the distances between the observer and the fixated objects can mean that fixations over objects subtend angles too small to discern using a standard eye-tracker — even if such small images do continue to elicit stereotyped scanpaths. Secondly, the resolution of the displays can limit what is perceptible, as objects appear smaller with distance. However, as we shall be presenting immersive environments and looking at *scanpatterns*, our stimuli will extend around the observer. This means that the ROIs may be relatively large with respect to the display size, because the contents of the display change as the observer's head moves around to view the scene.

3.4.3 Searchpath and Scanpath Modelling

Once some consistent system has been selected to define the regions of interest, it is typically quite simple to allocate successive foveations to them in order to record the eye and head scanpatterns as a series of transitions between the regions. Upon the detection of each new fixation, its occurrence is used to update and hence incrementally construct a model.

Selection of a model in the first place is fully dependent upon the type of analysis that is subsequently to be carried out. The two main models that appear in the searchpath/scanpath literature are the 'string model', and the Markov Model. A string model supports the use of measures such as the string Edit Distance.

The string Edit Distance

In order to find similar eye scanpaths, Stark and Choi (1996) generated a sequenced string of letters, each letter denoting one of the regions of interest within the image. Given two such strings, a string editing algorithm may be used to find the difference between them by determining the number of atomic manipulations required to change one string into another. One such algorithm is the Edit Distance (Navarro, 2001), which calculates the minimum number of insertions, deletions and substitutions to turn one string into another. Using string models, such algorithms provide practical ways to measure the difference between scanpaths as well as searchpaths (Choi et al., 1995). Due to the view that paths are idiosyncratic while retaining a degree of unpredictability, strings used with string editing should be as short as possible, and should be recorded in response to some specific stimulus. Doing so optimises the probability of obtaining shorter distances and more correlated strings in an investigation. Thus the string editing distance is most suitable for short strings, and hence short recordings of searchpaths and scanpaths.

Searchpaths and scanpaths are likely to contain many short strings, each relating to the current sub-task (or ‘visual routines’) being carried out (Ullman, 1996; Pelz et al., 1999). However, automatically splicing strings into sub-tasks or routines is infeasible given the current level of knowledge and understanding of eye movements. Doing this manually may theoretically be possible, but even so it would be extremely tedious.

Markov Models

An alternative method for modelling searchpaths and scanpaths (or, scanpatterns in general) is to use a transition matrix that represents a Markov Process. A Markov Process of the first order is defined by a set of states (known as an alphabet), transitions between them, and some weight attached to each transition. In other words, it may be defined as a directed, weighted graph. Each weight lies in the interval [0,1], as each is a probability. Transitions weights need not be symmetric, so that the weight placed upon the transition between states A and B, $A \rightarrow B$, does not necessarily have to equal that of $B \rightarrow A$. Such a Markov Process must have the Markov Property, which is defined as:

$$\begin{aligned} P(X(t+1) = i_{t+1} | X(t) = i_t) \\ = P(X(t+1) = i_{t+1} | X(t) = i_t, X(t-1) = i_{t-1}), \ldots, X(0) = i_0) \end{aligned}$$

Where $X(t) = i_t$ asserts that the random variable $X(t)$ takes on the value i_t at time t, and $\forall t : 0 \leq t$. We shall only consider the case in which transitions occur in discrete steps (i.e. t takes on discrete values only.)

A Markov Model of the first order may be implemented as a simple probability transition matrix. Searchpaths that are modelled as matrices are more amenable to common numerical analysis than are strings. Not only can one easily calculate the probability of any particular path, but there are also numerous other numeric measures that are well known, documented, and understood. With such transition matrices, each individual region of interest is denoted by both a row and column, and so the probability of a transition from one ROI to another is stored in the appropriate matrix element. Once built, such matrices have also been used to generate scanpaths for simulation purposes (Stark, 1993) by executing a random walk through the graph.

3.4.4 Choosing a model

In this thesis only Markov Models shall be used to model scanpatterns in the form of probability transition matrices. This is because they are more amenable to mathematical analysis. The greatest advantage of using strings to model scanpatterns is that we may use the string Edit Distance algorithm to obtain a ‘distance’ between them. However, with transition matrices there are numerous methods that may be used to more straightforwardly find the distance between two Markov Models (see Section 3.5.3), thus making the string Edit Distance method redundant for our purposes. Further, Markov Models have specialised algorithms associated with them that may be useful. For instance, algorithms to find converging sets of states, thereby identifying the eventual subset of states that will be repeatedly visited. This could be useful for determining the stability of a stimulus.

3.5 Potential Model-Measures

Naturally, a model must be put to use. Because the nature of this work is one of investigation, we shall be looking to obtain useful measures that will give insight into the properties of IVEs. In the following sub-sections the measures used in this thesis will be discussed, and then further supporting methods in Section 3.6.

3.5.1 Information Theory

Information theory may be used to attribute quantifiable values to events, and processes based on the information they provide us with. Due to the apparent unpredictability of a person's gaze, in this thesis we model it as a stochastic process, and may thus use information theory as a tool for investigation.

Information theory has a number of facets, having different uses and definitions depending on the perspective of the user of it. It provides methods to quantify the level of uncertainty of events, given their probability, in terms of the amount of 'information' that the occurrence of such events would provide.

By way of example, let us consider a statistically biased television weather broadcast, which every day correctly tells us that the probability of it raining tomorrow is high; however, the probability of it being sunny is extremely small. If we assume that the broadcast probabilities are correct, but somehow we personally could predict (perfectly) the weather tomorrow, then in which case would our ability be most useful? Of course, the answer is when it is to be sunny. This is because most of the time persons following the advice of the broadcaster would find that they are correct: that it was likely to be raining, and that it did indeed rain. On such occasions our predictive ability would appear of little value to any person who had had access to the broadcast. On the other hand, before a sunny day our advice is at great odds with the broadcaster, which makes our advice (or information) extremely valuable.

In this example one soon recognises that the value of being able to perfectly predict the weather (the information) is very much tied to the probability of unlikely events. Information theory attributes a larger quantity of information as being communicated when an event occurs that was less likely. The quantity of information is measured in bits, and given the probability $p(e)$ of an independent event e, the information $H(e)$ is calculated as:

$$H(e) = log_2(\frac{1}{p}) = -log_2(p) \tag{3.1}$$

Let us consider a more commonly found example, in this case adapted from Attneave (1959) (and found elsewhere.) A fair coin that is tossed into the air has the probability of landing with the 'head' side up of $\frac{1}{2}$. Thus the information obtained when this occurs is:

$$H(`heads') = log_2(\frac{1}{\frac{1}{2}}) = log_2(2) = 1\ bit \tag{3.2}$$

Of course, for 'tails' side up, we have the same value occurring. It is interesting to note also, that to communicate the fact that a 'head' or 'tail' has occurred over a communications system, we require 1 bit of information to be transmitted.

If we change our second example slightly and consider a bent coin, perhaps a coin that has a probability of landing with the 'head' side up of $\frac{3}{4}$ then this modifies the situation as follows:

$$H(`heads') = log_2(\frac{1}{(\frac{3}{4})}) = log_2(4/3) = 0.415\,bits \tag{3.3}$$

Interestingly, the information has reduced. This may be ‘explained’ because as ‘heads’ is more likely, we are less surprised by its occurrence, and so we learn less from this event. The information that we obtain from the occurrence of ‘tails’ is:

$$H(`tails') = log_2(\frac{1}{(\frac{1}{4})}) = log_2(4) = 2\,bits \tag{3.4}$$

In the case of ‘tails’ we have an increase in information, because we are more surprised that we obtain a ‘tails’ than in the case where ‘heads’ and ‘tails’ are equally likely.

The above illustration is useful in understanding the concept of information as relayed by a single event. However, we are also interested in evaluating models of processes. A common (perhaps one could say ‘standard’) model used to work with information theory is the communications channel. Information theory tells us that the number of bits required to communicate the occurrence of ‘heads’ (H) or ‘tails’ (T) over a channel is calculated as follows:

$$H(`coin\ tossing\ process') = p(T)log_2(\frac{1}{p(T)}) + p(H)log_2(\frac{1}{p(H)}) \tag{3.5}$$

Note that this expression may be viewed as the average information that we obtain from our process per event.

In the fair coin example, we obtain:

$$H(`coin\ tossing\ process') = \frac{1}{2}log_2(2) + \frac{1}{2}log_2(2) = 0.5 \times 1 + 0.5 \times 1 = 1\,bit \tag{3.6}$$

For our biased coin problem, we find:

$$\begin{aligned} H(`coin\ tossing\ process') &= \frac{1}{4}log_2(4) + \frac{3}{4}log_2(4/3) \\ &= 0.25 \times 2 + 0.75 \times 0.415 \\ &= 0.5 + 0.311 \\ &= 0.811\,bits \end{aligned} \tag{3.7}$$

This is lower than our previous (fair coin) example, the explanation being that the unfair coin is more predictable, whereas a fair coin is *entirely* unpredictable. In fact, it may be of interest to know that this value reflects the number of bits required in an optimal encoding scheme (Attneave, 1959).

In general we can use the following equation to compute the average information per event arising from some process, where X represents an ensemble of potential events X_i (see MacKay, 2003, p32):

$$H(X) = \sum_{i=1}^{n} p(X_i)log_2(\frac{1}{p(X_i)}) \tag{3.8}$$

The information attributed to an event or process is therefore related to the uncertainty of the event. This uncertainty is termed entropy, and we may discuss this as the entropy of an event, a string of events, or a process.

In the next section we see how the information theoretic method may be used to characterise scan-pattern models.

3.5.2 Conditional Entropy

Ellis and Stark (1986) used a combination of statistics and information theory to analyse transition matrices, created from scanpattern data, to determine the statistical dependency in the transition probabilities. It is important to note that this dependency is in the transitions of a process (second order information), and is not simply a measure over the states visited (information of the first order) in that process. For a Markov Model, a notion of statistical dependency reflects the overall bias away from a probability transition matrix with maximal entropy wherein the next state in the process is random and is fully independent of the current state, toward a deterministic one whereby once in a particular state the next state is entirely pre-determined. Such a concept could prove to be useful for our purposes because, although explained in more detail later, we would expect little statistical dependency (or at least, a different level of statistical dependency) when presented with a nonsensical IVE; and perhaps also a low sense of presence in such an IVE. Therefore a measure of entropy might make for some type of a surrogate presence measure.

The statistical dependency over an entire matrix that describes a Markov process may be measured using information theory, and such a measure has been termed ‘total conditional information’ by Ellis and Stark (1986), or the ’Conditional Entropy Rate’. The Conditional Entropy Rate in a transition probability matrix is determined using the following formula noted by Ellis and Stark, although it was originally obtained from Brillouin (1962):

$$H_c = -\sum_{i=1}^{n} p(i) \left[\sum_{j=1, j\neq i}^{n} p(i,j) log_2 p(i,j) \right] \tag{3.9}$$

The term $p(i,j)$ represents the conditional probability of transitioning from ROI i to ROI j given that ROI i is currently foveated. $p(i)$ is the independent probability of ROI i being foveated. Thus the inner sum represents the entropy of transitions $p(i,j)$ (for all j where $i \neq j$) and each is weighted by $p(i)$.

3.5.3 Markov Model distance measures

If we assume that we have two Markov Models and can determine their similarity or distance from each other, then we could directly compare between scanpatterns modelled in this way. With this ability we may begin to consider experiments that could evaluate an IVE experience against a corresponding real world environment, and possibly find correlations that might frame a presence measure. In this section we consider two potential distance measures.

Euclidean Distance

The most obvious measure to use is a Euclidean distance. The Euclidean distance between two matrices M and N may be computed simply as:

$$d(M,N) = \sqrt{\sum (M_{ij} - N_{ij})^2} \tag{3.10}$$

However, because this function treats the matrix as though it were actually one long vector of dimension $|M|^2$ (or equivalently $|N|^2$), it ignores the fact that the ‘dimensions’ of the vector are not independent. However, if this weakness is understood, then this function could be a useful estimator, and is certainly easy to calculate and understand.

Path likelihood

A second method for determining distances between Markov Models is by comparing the likelihoods of paths through the states in the models.

Given a sequence of states (or path) s_i where $i = 1 \ldots n$, it is trivial to compute the likelihood that the path arose from a particular model. This is achieved by multiplying out the probabilities of each path transition according to that model. Rather than multiplying though, which could result in extremely small numbers that can become inaccurate due to rounding error, we may take the log of such probabilities, adding as we traverse the state sequence.

To compare multiple Markov Models then, we need an independent source of paths (S) for which we can compute the likelihoods of them arising from each model (M_k). This provides us with the 'distance' of each of the models from the paths, as probabilities $p(S, M_k)$. Each time we compute such a value, it forms a point in a distribution, and these distributions may be tested in a frequentist sense. However, when comparing just two models with respect to some set of paths (which acts as a reference point), we may alternatively take a Bayesian approach, and compute the log-likelihood ratio:

$$logL(S, M_a, M_b) = log\left[p(S, M_a)/p(S, M_b)\right] = log(p(S, M_a)) - log(p(S, M_b)) \tag{3.11}$$

Whereas the likelihood ratio test would be given as (Cover and Thomas, 2006):

$$\begin{aligned} L(S, M_a, M_b) &> 1 \\ \Rightarrow \frac{p(S, M_a)}{p(S, M_b)} &> 1 \end{aligned} \tag{3.12}$$

(assuming each hypothesis had an equal probability of being in error)

the log-likelihood ratio test would thus be:

$$\begin{aligned} log(L(S, M_a, M_b)) &> 0 \\ log\frac{p(S, M_a)}{p(S, M_b)} &> 0 \\ log(p(S, M_a)) - log(p(S, M_b)) &> 0 \\ log(p(S, M_a)) &> log(p(S, M_b)) \end{aligned} \tag{3.13}$$

3.6 Supporting Methods

3.6.1 Bayesian Inference

When building a statistical model of eye and/or head movements using scanpatterns, we employ Bayesian theory. All analysis takes place using movements of the eye and head that have been classified as belonging to some ROI, with the transitions between these regions being our main focus. One of the difficulties in modelling such transitions is that the model space is large. For n ROIs, we have n^2 possible transitions, and if we obtain transition data experimentally then obtaining a large enough data set to create a model can be challenging if it is possible at all. The problem with an insufficiently defined model is not simply that it contains relatively little data, but rather that such a model in the form of a transition frequency matrix contains zeros. When using such a matrix to compute a subsequent transition probability matrix, the resulting matrix will also contain zeros. This is not only an inaccurate reflection of reality (in which it is possible for any transition to occur, and a model should reflect this), but impossible transitions also present a mathematical problem. When multiplying out the probabilities of a path that includes such a zero element, the probability of the path suddenly becomes zero.

To illustrate this problem of zero matrix elements, we shall consider a concrete example. Let us conservatively assume that we have 20 ROIs (and hence 400 possible transitions), and that we obtain data samples (transitions) at a rate of 1 per two seconds [2]. To obtain approximately 10 samples per matrix element, we require $10 \times 400 \times 2 = 8000$ seconds of data, which turns out to be 133 minutes and 20 seconds, or 2 hours, 13 minutes and 20 seconds. Assuming that we have a static scene, the time for which we record data regarding a subject should be relatively short, in order to obtain their most autonomic and/or natural response to it. In keeping with this conservative example, let us assume that 2 minutes are allocated to each subject. It should now be obvious that we require over 60 subjects (not including any experimental failures or outlier data.) Given the nature of eye-movement experiments which can be difficult to carry out (Schnipke and Todd, 2000), this is an exceptionally large experimental undertaking, even having assumed such conservatively estimated parameters. Unfortunately this example also contains an implicit assumption that further exacerbates the problem; that captured transitions are randomly distributed over the ROIs in a uniform manner. Of course in reality this is not true, and we would expect that we are in fact dealing with a phenomenon better modelled as an exponential distribution. This means that to obtain sample values for all elements of the matrix would take an exponentially long time.

Help from Bayes'

Fortunately, by using Bayesian inference we can 'kick-start' a model, which provides us with a meaningful although somewhat synthetic model. This technique may be thought of as initialising a model and then modifying it with empirical data, or equally, merging both a synthetic and empirical model and allowing bias toward one or the other.

Dirichlet models

In our case, we have the probability transition matrix that models the transitions between ROIs. Each ROI (i) then has some probability of transitioning to another ROI (j), which we shall denote p_{ij}, and so for each value of i we have a Dirichlet probability distribution across j, $D_i(j)$. Because our model is a first order Markov process, each of the distributions D_i are considered independent.

Data collection

When collecting data it is processed and appears in the form of transition frequencies, which are the count of the number of times we observed a transition from ROI i to ROI j. These values are thus in the

---

[2] This is a very rough estimate based on the work presented later in this thesis.

form of a number of multinomial distributions M_i each having a random variable $M_i(j)$ that may take a non-negative integer value for each j ($j = 1 \ldots n$, where n represents the total number of ROIs).

Applying Bayes' Theorem

Using a Bayesian probability model, we are able to compose the collected experimental data with a 'synthetic' model[3]. This synthetic model is set up according to our prior beliefs, which are stated as a probability distribution (our own prior parameter estimates for transition probabilities). The composition of these produces a *posterior distribution*.

Bayes' Theorem is often stated as:

$$Posterior \propto Likelihood \times Prior$$

Or, using mathematical notation:

$$\pi'(\theta) = L(\theta).\pi(\theta) \tag{3.14}$$

where θ is an unknown free variable. As stated above, we are dealing with the distribution of probabilities, which are those of transitioning from one Markov process state to another, the posterior distribution is of the class Dirichlet. It should be noted that the conjugate prior distribution of the multinomial distribution is also a Dirichlet distribution, and this is of import as it simplifies the mathematics involved (Kern, 2006). Knowledge of this relationship allows us to take advantage of Bayes' Theorem by utilising a Dirichlet prior distribution along with our empirical data to produce posterior probabilities, which will then act as the final transition probabilities in our model.

Using the multinomial distribution, we can state that for a set of empirical data (d_i where $d = \sum_i d_i$), the probability of obtaining that data under the distribution θ is:

$$p(d_1 \ldots d_i|\theta) = \frac{d!}{\prod_i d_i!}.\prod_i \theta_i^{d_i} \tag{3.15}$$

Here the values d_i are viewed as fixed (the empirical data), and θ is a free variable. We can therefore state that the likelihood of θ is:

$$L(\theta) \propto \prod_i \theta_i^{d_i} \tag{3.16}$$

Our prior (Dirichlet) distribution provides the probability of θ given our prior beliefs as described through the hyper-parameter α (and where $A = \sum_i \alpha_i$):

$$p(\theta|\alpha_1 \ldots \alpha_i) = \Gamma(A).\prod_i (\frac{\theta_i^{\alpha_i - 1}}{\Gamma(\alpha_i)}) \tag{3.17}$$

In our case, the hyper-parameter α is set up in a most simple fashion by giving each α_i the same value for all i (doing so emulates a uniform prior distribution.) The constant that we select is $\alpha_i = \frac{1}{|ROI|}$ where $|ROI|$ represents the number of ROIs. Choosing this particular constant is analogous to providing for a single transition away from each state (ROI), evenly (uniformly) split between the possible destination states.

---

[3] The explanation here is attributed in the main to the excellent elaboration of Kern (2006).

Substituting equations 3.16 and 3.17 into 3.14, we obtain the following posterior distribution:

$$p(\theta) \propto \prod_i (\theta_i^{d_i+\alpha_i-1}) \tag{3.18}$$

This clearly is also a Dirichlet distribution, but with the parameter $b_i = (d_i + \alpha_i)$. The marginal expectation of each θ_i in a Dirichlet distribution is well known to be:

$$\theta_i = \frac{d_i + \alpha_i}{d + \alpha} \tag{3.19}$$

This simple and elegant expression provides a direct equation with which we may obtain the marginal posterior transition probabilities θ_i. It is a clear and intuitive trade off between prior distribution and the empirical data, where the α_i values appear as prior observations of transitions. This may be very simply implemented by initialising every element of the transition frequency matrix (generated before the transition probability matrix is computed) to α_i before adding the observed transitions from the scanpattern data.

3.6.2 Cluster Analysis

Cluster Analysis takes the approach of using measures in normally more than one dimension, to hierarchically split (or conversely, join) items of data to build clusters of related data. Clustering is useful in many situations, one of which is to reveal previously unknown taxonomies, as well as for post-hoc classification. We shall see that the latter problem arises later within this thesis, when we would like to determine the existence of a discontinuity in our data. We may also use such a method to create regions of interest (ROIs) from fixational data, by finding clusters.

There are many ways in which clustering can be performed, and for the most part these differences relate to the way in which distances between data, and thus clusters, are computed. For the purposes of this work these nuances of clustering are of secondary importance, so we shall only briefly describe the method used herein and its context.

In Everitt et al. (2001) we find an enumeration of what are known as the standard hierarchical agglomerative methods (see Table 3.1). From his overview and reference to empirical studies, one of the better algorithms in terms of cluster-interpretability and robustness is the Unweighted Pair-group with Arithmetic Mean (UPGMA.) A seemingly close second in this respect, is UPGMC (Unweighted Pair-group Mean Centroid), which is computationally less expensive as the cluster centre and weight may exploit a computer's memory (in the form of a cache) in a straightforward manner. Such caching effectively reduces the cluster computation time from $O(n^3)$ (UPGMA) to $O(n^2)$ (UPGMC).

Given a function d which determines the distance between two points, and two clusters P and Q having m and n points respectively, the distances between the clusters is calculated as follows:

(UPGMC)

$$D_{p,q} = d\left(\frac{1}{m}\sum_i^m p_i, \frac{1}{n}\sum_j^n q_j\right) \tag{3.20}$$

Method	Alternative Name*	Usually used with:	Distance between clusters defined as:
Single linkage Sneath (1957)	Nearest neighbour	Similarity or distance	Minimum distance between pair of objects, one in one cluster, one in the other.
Complete linkage Sorensen (1948)	Furthest neighbour	Similarity or distance	Maximum distance between pair of objects, one in one, and one in the other.
(Group) Average Linkage Sokal and Michener (1958)	UPGMA	Similarity or distance	Average distance between pair of objects, one in one cluster, one in the other.
Centroid Linkage Sokal and Michener (1958)	UPGMC	Distance (requires raw data)	Squared Euclidean distance between mean vectors (centroids).
Median linkage Gower (1967b)	WPGMC	Distance (requires raw data)	Squared Euclidean distance between weighted centroids.
Ward's method Ward (1963)	Minimum sum of squares	Distance (requires raw data)	Increase in sum of squares within clusters, after fusion, summed over all variables.

*U, unweighted; W, weighted; PG, pair group; A, average; C, centroid

Table 3.1: Standard agglomerative hierarchical clustering methods. Reproduced and adapted with permission of B.S. Everitt, see Everitt et al. (2001) for further details and remarks.

(UPGMA)

$$D_{p,q} = \frac{1}{mn} \sum_{i}^{m} \sum_{j}^{n} d(p_i, q_j) \tag{3.21}$$

An important feature of clustering algorithms is the general requirement (there are exceptions) for the predetermination or pre-selection of the number of clusters, before clustering actually commences.

One method that may be used to estimate the number of clusters is ‘v-fold cross validation’. This method firstly removes a random subset of the data, called the hold-out data. The remainder of the data is then clustered, and then the hold-out data is compared to the clusters to determine an error value. The error value is computed to reflect how well the cluster model fits the hold-out data. By repeating this procedure with different sets of hold-out data, a confidence interval can be constructed for the error value. Finally, in doing this for an increasing number of clusters, it is possible to estimate the number of clusters that best fits the data as a whole by analysing the error values. Although described here as a method in the context of clustering, v-fold cross validation is used with other types of analyses.

3.7 Experimental Tasks

As with most experiments, great consideration must be given to the task set for subjects. Within each experiment set out in this thesis, the task is the same for all experimental conditions.

When considering gaze, the task given to the subjects is known to explicitly affect a scanpath in a manner consistent with the task, although the scanpath is and continues to be idiosyncratic. This is exemplified by Yarbus, where subjects are given distinct tasks, and for each the scanpath is recorded (Yarbus, 1967). The recorded scanpaths are clearly related to the task even though there is a great deal of what might be seen as ‘noise’. A brief summary of the effects of task on gaze is provided by Henderson (2003), and the importance of task and further insights as to how it is and may be manipulated is commented on by Hayhoe and Ballard (2005).

To find a suitable task for an experiment is not particularly easy. On the one hand it must be designed to constrain the scanpattern as much as possible, so that random effects are reduced, aiding subsequent analyses. At the same time, the task should allow the subject enough freedom to experience the environment (and so attain a sense of presence in the IVE.) The subjects must also be kept attentive (with respect to the IVE) for as much of the time as is possible, however in spite of this we utilise only static environments to keep the stimuli as stable as we can, and thus reduce confounding factors.

In order to achieve these goals we utilise an open-ended task in the first experiment of this thesis. The experiment has four minutes per exposure, and each subject is asked to search for a ‘small flower’ and told that it will not be easy to find. In fact, the subject is kept searching throughout the experience because no such ‘small flower’ exists in any of the environments presented. Thus a task is set to prime the subject, and provide them with an objective that should drive the participant throughout the experiment. The second experiment has a much shorter duration, lasting only 1 minute per exposure. That it is shorter should help reduce variation in the scanpattern, not only because there should be fewer effects of ‘visual routines’ (see Section 2.5.6), but because these initial gaze movements should be largely made up of an orienting response, whereby the observer makes an initial assessment of their immediate environment. In this experiment, each subject is asked to determine what uses the environments might be put. Thereby, the subject is also ‘primed’ with a task.

3.8 A Physiological Measure

Motivation

Part of the methodology involved in this work is inspired by an experiment investigating the concept of presence and Electrodermal Activity (EDA) in which the author was involved. The results of the EDA analysis of the experiment were not included in the eventual publications (Freeman et al., 2003, 2005a,b). One of the secondary aims of the experiments was to test whether a physiological device could be used to detect an anxiety response from a virtual stimulus, thus indicating a likely sense of presence. The experiment placed subjects within an immersive virtual environment that was designed to look like a library, and had several animated figures (known as avatars) within it. Although the study was originally designed with the aim of testing the effectiveness of avatars, it also appeared that the physiological measuring device was sensitive to the subjects' emotions and perceptions as affected by the virtual stimuli. Interestingly, it was later noted (though not published) that the physiology data was correlated with the proximity of the subject to the avatars[4]. The study played an important part in the development of this thesis, in that the same physiological expedient (EDA) was used to support the our first experiment. Secondly, this finding acted as a springboard into the use of physiological devices in general — a field that includes the eye tracking methodologies that appear herein.

Physiological measures have recently made an impact on presence research (Pugnetti et al., 2001; Meehan, 2001; Meehan et al., 2002, 2003; Slater et al., 2003; Guger et al., 2004; Slater et al., 2006b), as described in Section 2.4.4. Their exploitation has been achieved in particular by provoking mild anxiety or startle responses, from which physiological responses are detectable. As previously mentioned, some kind of objective physiological measure will later be required to validate the use of scanpatterns for detecting perceptual changes. The main sources of physiological data for consideration are EDA and EKG because research has already been carried out supporting these methods for this purpose (Meehan, 2001). Exactly which of these two measures is most useful is obviously dependent upon the type of experiment being considered at the time, although there is certainly overlap in their utility.

Method

Meehan (2001) found that EKG and EDA data was amenable to analysis by averaging over the full duration of an experimental treatment, and comparing differences in these averages with respect to treatments. This method of analysis is simple to execute once the data is obtained and processed.

Whilst EKG data requires manual intervention in order to reliably locate (temporally) the beats of the heart (for instance, by confirming identification of the 'QRS' signal, Papillo and Shapiro, 1990), EDA data on the other hand may be processed in an entirely automated manner. This was one reason that in this work we favour EDA over EKG. The other reason that we focus on EDA is due to our prior experience in initially investigating it as a potential measure, and subsequently using it successfully in studies such as Slater et al. (2003), and later Slater et al. (2006a).

The method used in this work is based on Skin Conductance Responses (SCRs). This method relies upon the detection of response events, rather than changes in the overall level of skin conductance (which is highly variable.) The only disadvantage of this method is that it is a little more complex than mere averaging skin conductance data. The primary aim of the technique is to determine the occurrence of Skin Conductance Responses (SCRs), which are illustrated in Figure 3.3.

The process firstly applies a low-pass filter to the EDA signal in order to smooth it. After this, turning points of the signal are located by differentiation of the first and second orders so as to find the points in time at which a response (Skin Conductance Response, or SCR) begins and ends. Constraints are also applied in order to provide a threshold for the size of the response (below which no response is deemed to have occurred), and for the maximal duration of the response. Once obtained, the number of SCR events within a 'time-window' may be counted, obtaining an SCR rate. The total number of events within a time-window may be compared to determine significant trends in the data.

---

[4] Personal communication with Mel Slater

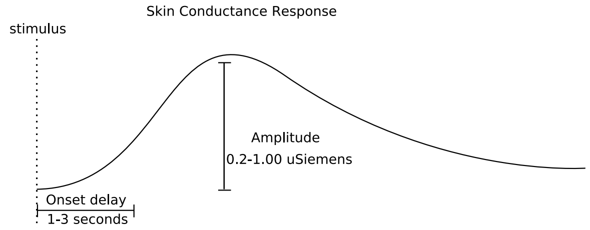


Figure 3.3: Approximate SCR Characteristics (values from Dawson et al., 1990)

3.9 Summary

In this chapter we have discussed the inspirations and general approach that will be used for the investigations to follow in this thesis. To summarise briefly, the approach consists of using gaze scanpatterns, which according to existing scientific literature enable us to potentially detect states of perception. This involves the processing of eye tracking and head tracking data to compute the line of sight, and/or ‘point of gaze’, with respect to the virtual environment being viewed. We utilise some method for the determination of regions of interest, providing for the classification of eye movements. We then record the sequence of visited regions, and thus transitions between these regions. Models are then built from the data using Bayesian inference, and may be analysed using the mathematical methods mentioned, such as information theory or Euclidean geometry. As a supporting tool, we utilise physiological measures and ‘presence questionnaires’ that have been established in the presence research literature for detecting presence.

There are two general approaches that we shall use to analyse scanpattern data. Firstly, we have the statistical dependency in ROI transitions that can be used to characterise scanpatterns. This is loosely based on the statistical dependency scanpath work carried out by Ellis and Stark (1986). Secondly, we have the measures of Markov Model distance, as calculated by the Euclidean and path likelihood functions. Each of these methods may be useful to detect the changes in the scanpatterns predicted by Hypothesis Testing and scanpattern research. We expect that the scanpattern will change with respect to the environment when a Gestalt switch is made, that is, as a person perceives their environment/situation in a different way. We also expect that real world scanpatterns will be reflected by scanpatterns arising from immersive virtual environments that are designed to represent them. Finally, after achieving a state of presence in an IVE, we do not expect that adding further realism will lead to scanpatterns that are significantly closer to those that arise when viewing the real world environment that the IVE represents.

Chapter 4

Experiment I: Investigating Minimal Cues for Presence using Gaze Tracking

4.1 Introduction

Having explored much of the background to our research questions as well as related analysis methods, we are now able to move to the next stage of this thesis: the investigation.

To recapitulate, Research Question 1 is "**Can the gaze scanpattern be used to detect RAIR in IVE users?**" To investigate this, we designed an experiment that tested for presence by constructing a measure of gaze entropy which, according to perceptual theories (see below), should indicate when an observer perceives a meaningful environment. This type of perception (i.e. a meaningful perception) would be expected when viewing most everyday scenes. In parallel, we employed an already established presence measure. By varying the displayed environment, we show that the two measures are correlated.

According to Gregory (1977), perception works through a continual process of hypothesis generation. Perceptual hypotheses are generated as potential descriptions of visual information, from which is selected a particular perception. The selection is continually re-evaluated, but also, as any new information is acquired further hypotheses may be generated. The re-evaluation of these hypotheses is thought to be a major reason that the idiosyncratic scanpaths found by Yarbus (1967) appear. Regions of a stimulus are re-visited over and over again in order to re-evaluate the hypothesis(es). Gregory (1977) has also implied that the inability to produce a favoured hypothesis from stimuli leads to a lack of perception. Thus, it seems that without suitable hypotheses, or even with quickly changing hypotheses, the visual stimuli will not be re-evaluated but rather (perhaps) new visual information will be sought to help generate acceptable hypotheses.

We use this theory of Gregory's to investigate whether a meaningful perception is achieved by an IVE subject. In everyday life we generally perceive our environment as a visually meaningful one, and thus when our IVE is also perceptible as meaningful we expect a positive presence-response. To be sure of this, we also require falsification which we provide through the use of an alternative condition, in which a non-meaningful IVE is developed. This should result in the lack of a positive presence-response for 'meaningfulness'.

If we obtain a positive result from this indicator, it cannot be regarded as a sufficient condition, and may not even be a necessary one for presence. In the general case though we should expect meaningful perception when viewing an environment.

One interesting difference between the work of Yarbus and Stark versus our own, is that we are using a three-dimensional environment stimulus rather than a two-dimensional image. Here we wish to investigate the repetition of a *scanpattern* more akin to the searchpath study carried out by Choi et al. (1995). In their study, a three-dimensional image was used without head-tracking. Whereas a scanpath is defined to be a “repetitive, idiosyncratic sequence of saccades alternating with fixations based upon a cognitive model” (say, over an object), a searchpath is defined as a “repetitive, idiosyncratic sequence of saccades alternative with fixations based upon a spatial model” (Choi et al., 1995).

We also address Research Question 2 in this experiment, “**Does a visual cues threshold for the inducement of presence exist in the context of IVEs?**” To investigate this, we test for presence-responses whilst the visual cues of our virtual environment are being progressively introduced. We then analyse the measured data in the hope that we might find evidence of a discontinuity — indicating the change from having no sense of presence in the IVE, to becoming present in the IVE. This expectation is due mainly to the school of Gestalt Psychology, which maintains that perceptions appear in an all-or-nothing manner (see Section 2.5.5). Therefore, once sufficient visual cues are present in the IVE, an environment is suddenly perceived. The results of this experiment have been published (Jordan and Slater, 2009).

It should be remembered that throughout this thesis we consider a person to have attained a sense of presence if they respond to the IVE in a way that they would if the IVE were in fact real. Thus, we are interested in just a subset of the responses that we would expect when present in the IVE (as though it were real.)

This chapter is divided into 10 sections. The following Section 2 details our hypothesis and the rationale behind it. Section 3 contains Assumptions made, and Section 4 the Experimental Design. In Sections 5 and 6 respectively, the Procedure and Materials are described. Sections 7 defines the Response Variables and Section 8 the Manipulated and Explanatory Variables. Finally, Section 9 provides the experiment’s results, and in Section 10 we draw our final conclusions.

4.2 Operational Hypotheses

Our method involves displaying an immersive virtual environment whilst developing it over time, by slowly adding an increased number of polygons to the scene in order to defer the perception of the scene whilst continuing to present an observable stimulus.

There are two operational hypotheses to be tested:

1. *Once the minimal visual cues are presented, a perception of a meaningful virtual environment is achieved, and a stabilised perception should commence. We hypothesise that this will be reflected by a significant drop in the entropy of eye-in-head movements.*

   This is because once a meaningful perception is found, there should subsequently be increased structure in the scanpattern as fixations are made to the meaningful regions. This should occur according to Gregory’s and Stark’s theses (see Section 4.1): without a stabilised perception there is expected to be little structure within the scanpattern as the gaze will shift in an un-structured fashion to find new evidence to create and validate further ‘perceptual hypotheses’.

To provide a control for this effect we use an additional virtual environment that is non-meaningful even when fully developed. The non-meaningful environment consisting of random polygons should not provide a source of minimal visual cues at any time throughout the experiment.

Thus, the method we construct is designed to detect a significant difference between scanpatterns generated as a response to a meaningful environment, versus those generated as a response to a non-meaningful environment.

2. *Once minimal visual cues are presented, a perception of a meaningful virtual environment is achieved, and there should be a physiological stress response given the premise that the environment is designed and is able to provoke such a physiological response.*

   EDA is an excellent indicator of stress, and we exploit this by using a stress inducing environment and a stress neutral (control) environment to determine whether or not participants eventually perceive the developing environment presented. This indicator was designed to corroborate the anticipated results of our gaze analysis.

Evidence for our first hypothesis is corroborated by that of the second. That is, skin conductance as measured by an EDA device was used to detect a stress response that should occur when a meaningful perception was achieved. This response acts as a presence indicator, to corroborate the anticipated change in scanpatterns expected when the environment is perceived as meaningful.

The Null Hypotheses may be stated as follows:

1. Eye movements continue to follow the same unstructured paths before and after the intended perception is attained, and thus any change in the conditional entropy of eye and head movements is statistically indiscernible.

2. A stress response is not detectable (as a significant increase in a response measure) contingent upon the presentation of a stressful virtual environment.

4.3 Assumptions

Visual cues displayed around a person in a virtual environment collectively determine the environment that they perceive.

The scanpattern (or more specifically, scanpath and searchpath) characteristics hold true although the subjects view an immersive virtual (and hence egocentric) environment rather than an allocentric two-dimensional image (Yarbus, 1967) or three-dimensional image (Choi et al., 1995).

The virtual environment was designed with intent to induce vertigo, or at least a state of heightened awareness. This state may thus be measured via the physiological devices (see Section 2.4.4). This state will only be elicited if a state of presence is induced (by way of the perception of the virtual environment as we intended.)

4.4 Experimental Design

4.4.1 Participant Population

Subsequent to approval of our experiment by the UCL Ethics Committee, advertisements were placed in and around the University College London campus. Participants were asked to take part in a paid ‘Virtual Reality study’ titled ‘Investigating Environments’, lasting approximately 45 minutes in total. They would be paid a total of £5 if not employees of the university. The advertisement stated that participants could only take part if they had good uncorrected eyesight, precluding contact lens and glasses wearers. The intention of this restriction was to reduce major variances due to differences in eyesight, as well as to facilitate the wearing of the head-mounted-display and eye-tracking equipment that could be intrusive, dependent upon physical facial and head structures. Willing participants were instructed to email the experimenter in order to book an appointment. Forty-two ($n = 42$) participants qualified though only twenty-eight ($n = 28$) provided eye-tracking data. In this chapter, we shall only consider data from participants who provided eye-tracking data. The number of male and female participants within each condition (to be described) were equal.

4.4.2 Experimental Conditions

The experiment has a within-groups design having one (1) factor; the environment presented, which has three (3) levels:

- Stress inducing Environment (SE), participants: $n = 8$,
- (Stress) Neutral Environment (NE), participants: $n = 6$,
- Random Environment (RE), participants: $n = 14$.

Participants viewing SE or NE also saw a second environment, NE or SE respectively, such that the order of presentation was reversed for half the participants. Participants that experienced environment RE did not view a second environment. However, analysis of the second exposures was not carried out; analysis of just the first exposures makes for a simpler design.

Environment NE was used as a control, in which EDA was expected to be relatively steady over time. However, environment SE was expected to show an increased number of SCRs (Skin Conductance Responses) after the point that minimal visual cues are presented. Environment RE was to act as a control for the investigation of eye movements, so that environments SE and NE could be considered in contrast to RE.

Environment SE (Stress inducing Environment) was a simple room environment with the exception that it was designed specifically to induce vertigo, because it also contained a tall column upon which our participants would virtually be standing.

The virtual environment consists of a room measuring 3 meters square, and 6 meters high. In the room were several items of furniture (three chairs and two sofas), a door, and two empty picture frames on the wall. Also, in the centre of the room was the column of approximate width and length of 40cm, and a height of 3 meters, upon which was stood the virtual body of the participant. The environment was presented such that participants would view it from a standing position on top of the column. The model contained 3,500 polygons.

Environment NE was created identical to SE with the exception that the column's height was reduced to approximately 1cm (appearing as a simple square mat beneath the participant's feet) to provide a 'stress neutral' condition. The virtual body and viewpoint were also displaced accordingly, so that the participant viewed the room from the top of the virtual body that is practically standing upon the 'floor' of the room.

Environment RE also contained the same polygons from environment SE (and thus NE), but they were rotated randomly about the centre of the model to create a meaningless environment.

In all environments SE, NE, and RE, the virtual body was supplied in the form of a headless avatar, that extended from the floor of the lab up to just below the centre of projection (where the participant's head would be.) This is shown in Figure 4.5. Therefore, when a subject looked down they would see a virtual torso, arms and legs approximately registered where their real body would be.

See Figure 4.1 for examples of environment SE at various levels of emergence, with Figures 4.2, 4.3, and 4.4 showing environments SE, NE, and RE in their final states (full detail.)

4.5 Procedure

The procedure used with each participant as they arrived is described as follows. Only one participant could take part at any one time.

Before the experiment was carried out, each participant completed a general demographic questionnaire and a Simulator Sickness Questionnaire (SSQ) Kennedy et al. (1993). They then proceeded to put on the equipment. Next, the participant would follow a standard (Applied Science Labs 501) procedure to calibrate the eye-tracking equipment. Once completed, they would be directed to stand over a floor marker set at a place corresponding to that at which the 'virtual body' (see Section 2.4.1 and Section 2.4.3) would stand in the virtual environment.

Participants were now shown a virtual training environment, which consisted of everyday items of furniture along with the virtual body. This training period, which lasted one minute, was used to test the equipment, allow the participants to get used to wearing the equipment, and also to record baseline data for each type of measure we used.

Next, the task was explained to the participant: to look for a small flower. This was set in an attempt to stabilise the scanpattern as much as possible in terms of having a fixed task for all participants (priming each participant consistently), whilst encouraging the participants to visually explore the scene. No flower was ever presented. The participants were required to keep their feet planted in the same spot throughout the experiment, but were allowed to move their body and limbs apart from this limitation. Participants were instructed to report their findings at the very end of the trial on a questionnaire.

The experiment was then started. In every condition, the experience lasted 4 minutes. At the start, the environment displayed was an empty black void. After subsequent periods of 4 seconds a number of visible polygons were added to the immediate environment being displayed. The number of polygons added after each period increased exponentially. The polygons to be added were randomly selected, and

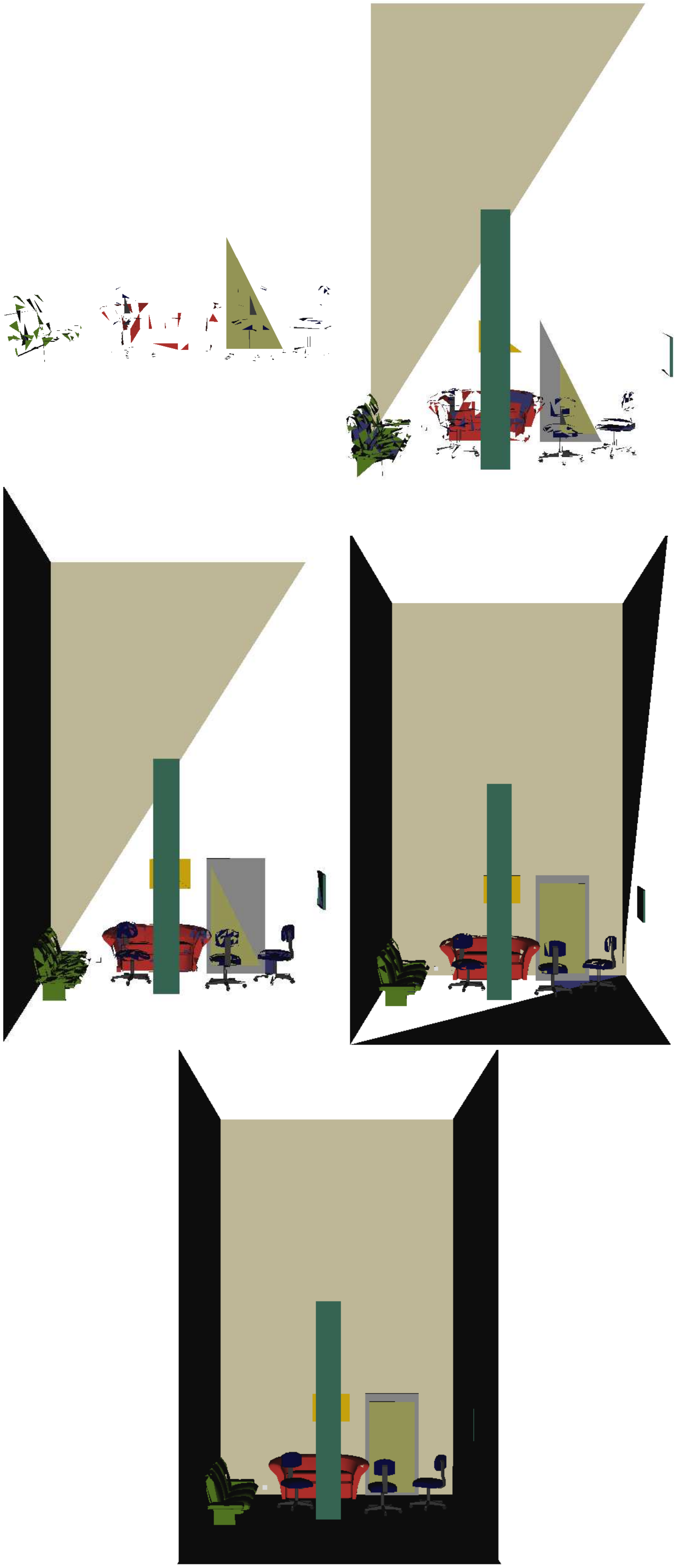


Figure 4.1: Examples of environment SE with an increasing number of polygons.

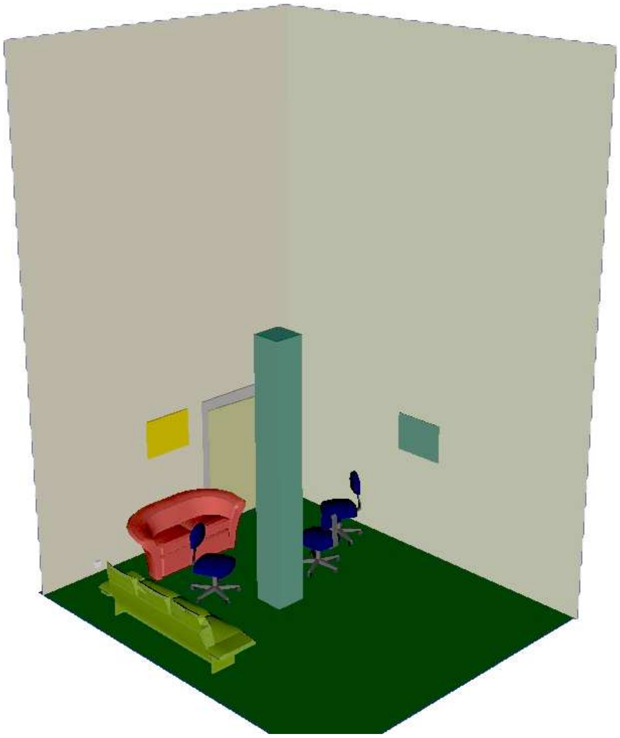


Figure 4.2: The complete SE model.

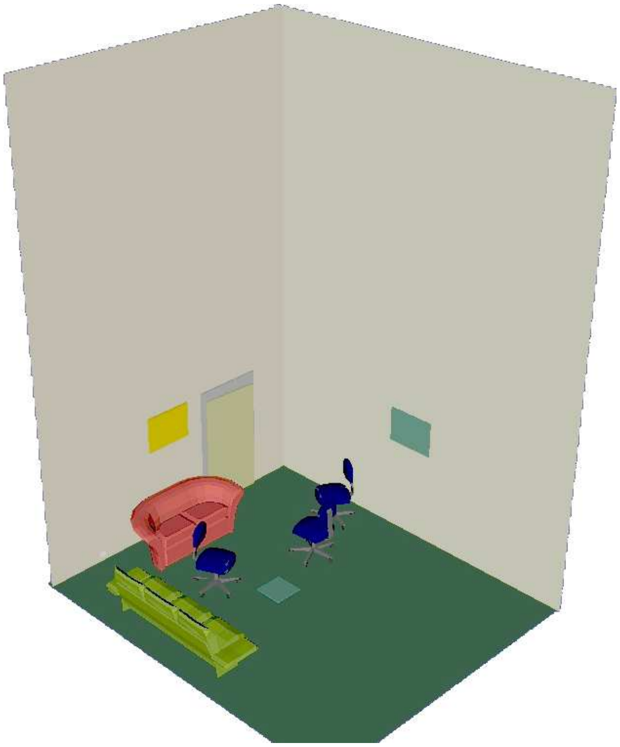


Figure 4.3: The complete NE model.

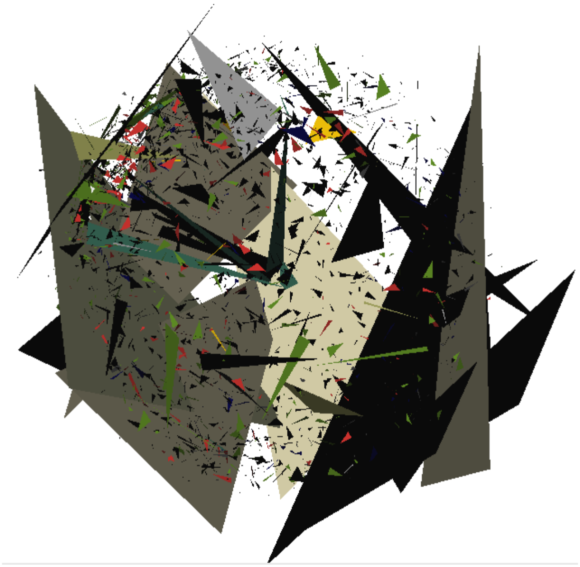


Figure 4.4: The complete RE model.

Figure 4.5: The virtual body.

(a) Side view. (b) Inside, showing the optics and eye-tracker.

Figure 4.6: Head-Mounted Display with affixed eye-tracker (Applied Science Labs 501.)

appeared in the same order for all participants. This continued until all polygons were displayed after a total of 120 seconds. A further 120 seconds viewing time was allowed after this point, with the then static environment.

After the experiment the eye-tracker calibration was again executed, to allow the detection of cases in which the eye-tracker had slipped so as to render the data unusable. The participants were then able to remove the equipment, and were asked to complete the complimentary part of the Simulator Sickness Questionnaire, and our presence questionnaire based on that of the SUS (Slater et al., 1994) (see Section 4.7.3.

4.6 Materials

A 1.8GHz PC drives the main application, which provides the graphics (using an AGP PNY GeForce FX5500 graphics card.)

The VE was displayed using a Virtual Research VR8 head mounted display (HMD), coupled to a Polhemus Fastrak head-tracker. The Virtual Research VR8 head mounted display (HMD) was used to display colour stereo images, having a refresh rate of 60Hz, and a resolution of 640 by 400 pixels per screen. The viewing angle was 60° across the diagonal. The 6DOF Polhemus Fastrak was used to track the head position and orientation at a rate of 120Hz.

Attached to the HMD was a single camera based eye-tracker (ASL 501) for the left eye, that updates at a frequency of 50Hz (constrained by the camera's refresh rate.) The accuracy of the computed line-of-sight, using the eye-tracker and head-tracker together, was tested by having a spectator view the corners of the 'picture' on the wall in the NE condition. Accuracy was found to be +/-1.5 degrees, with the maximum sampled error being just under 3 degrees. The HMD and eye-tracker are shown in Figure 4.6.

A ProComp+ physiological instrument was used to measure skin conductance, EKG (Electrocardiograph), and respiration, recorded at a rate of 32Hz.

All data from the above devices were recorded, as well as the times of each event — an event being defined as the appearance of a set of polygons in the environment.

A separate SGI O2 machine was used to record all data via VRPN (Taylor et al., 2001) software.

4.7 Response Variables

The two main response variables of interest were derived from Electro-dermal Activity (EDA) and the composite eye-head movements. In addition there were the responses to a presence questionnaire.

4.7.1 Physiological Response Variable

The EDA data was recorded as Skin Conductance (SC) (in μSiemens) at a rate of 32 Hz. The measure used was the number of Skin Conductance Responses (SCR), computed as follows. First the signal was smoothed, which was achieved through the use of a wavelet decomposition function that effectively acts as a low-pass filter. Decomposition at 6 levels was performed, and then a reconstruction of the wavelet coefficients at the greatest level that provides us with an error of less than 0.05 μSiemens is selected as the new series. The 'error' in this case is defined as the maximum spot difference between the original data series and the reconstructed (smooth) series. The second order derivative of this smoothed series indicates the points in time at which the signal accelerates and decelerates (i.e. maximal turning points), these being identified as potential SCRs. An SCR was defined using this method as a local maximum that has an amplitude greater than 0.2 μSiemens occurring within a window of 5 seconds (Dawson et al., 2000).

Our first response variable $S(t)$ is determined by computing the number of these SCRs that occur in the interval $(t-30, t]$ for each $t = 30..209$. This may be described as a discrete 30-second sliding window with a resolution of 1-second.

4.7.2 Scanpattern Response Variables

The second response variable is designed to reflect the eye-head movements over a relatively short period of time. This is achieved through the analysis of the participant's line-of-sight in three-space, which is traced as it moves around the scene, and is calculated from the composite eye-tracking and head-tracking data.

Modelling the Scanpattern

Scanpatterns are modelled using transition frequency matrices. The transitions are those occurring between intersected regions-of-interest (ROIs) and the line-of-sight. We a-priori define our regions-of-interest, by segmenting the scene into r regions, using a geodesic grid (described below). A transition of the line-of-sight from one region to another is assumed only once the new region has been foveated for a period of longer than 267ms (Buswell, 1935). Other temporal thresholds were tried (Duchowski, 2003; Yarbus, 1967), but these made little difference, as there were only negligible changes in the resulting foveation sequences. A state-state transition frequency matrix may then be constructed from the sequences, and from this our measures may be computed.

Icosahedrons are a typical polyhedron used as a geodesic grid, and may be used as a starting point to create more refined grids. We use such grids to define our (potential) regions-of-interest (ROIs) around an IVE observer. Although the icosahedron has only 20 sides, it may be easily and regularly subdivided, to increase the number of faces by a factor of 4. By placing an IVE observer in the center of such a grid, the environment is subdivided into regular regions. For our purposes, there is a trade-off when deciding on the number of subdivisions to be applied. Should we choose too few, then each region will represent a large area of the environment, meaning that any resulting transition model will be a less detailed model than it could have been. In contrast, too many subdivisions (and hence too many faces on the geodesic grid) will result in a very large transition frequency matrix. This is not only because the number of faces increases as a geometric progression, but also because the number of elements in the transition frequency matrix is of order O(r^2), which compounds this increase. This is important when considering the population of the matrix with sample data: how many samples are required to sufficiently characterise the matrix? With just a few more subdivisions of the grid, the number of elements in the transition matrix increases greatly.

Once the required level of subdivision is decided and applied, the resulting geodesic grid has each of its vertices normalised, so that the distance of each vertex from the center of the whole structure is 1m, so that each vertex is a point on a sphere. This structure is then placed around the observer to segment the scene, with each face (triangle) acting as an invisible window onto each region of the environment.

Each face is numbered, and so as the line-of-sight passes from one region (face) to another, the transition is recorded in the transition frequency matrix. From this transition frequency matrix, a transition conditional-probability matrix is then constructed. This is the final step of data processing before we compute our measures.

Eye-Head Movement Response Variables

Two functions were utilised to form the scanpattern response variables, the conditional entropy, and the Euclidean distance. We explain each of these now.

(i) Conditional Entropy Rate

The conditional entropy rate (CER) acts as our first Eye-Head movement response variable, and may be computed from a single transition conditional-probability matrix. Given a particular matrix (model), the function produces a single value that quantifies the extent to which, in general, a transition to a new region is conditional upon the currently intersected region of interest. According to our operational hypothesis (Section 4.2), we expect that the conditional entropy would be lower due to more structure in the scanpattern once a meaningful environment is perceived, as the scanpattern should stabilise as a selected hypothesis is settled upon.

This measure has been termed statistical-dependency, and our use of it is inspired by the work of Ellis and Stark (1986), wherein it was also used with respect to eye movements[1]. Although we were not attempting to reproduce their results through this experiment, we find in their work this metric that characterises a transition matrix in exactly the way we require. It should be noted that our study has quite different conditions, for instance, using head tracking our scene extends 360° around the subjects, and so we are using larger regions to cover the elements of the environment. In contrast, Ellis and Stark's 1986 paper utilised a spatially fixed image that subtends an acute solid angle.

From the transition frequency matrix, it is not entirely trivial to compute the transition probability matrix. Although it appears simple to obtain the probability of transitioning from region a to region b

---

[1] Originally, the measure was mathematically defined by Brillouin (1962).

using the equation

$$p_{a \to b} = \frac{transitions_{a \to b}}{\sum_x transitions_{a \to x}} \tag{4.1}$$

problems occur when there are no sampled transitions from region a at all. The problem is twofold, firstly because the formula then yields $\frac{0}{0}$, which is undefined, and secondly because we later wish to take the log of these probabilities, so we cannot substitute a (token) zero[2] for them.

To overcome this problem, we must decide upon a suitable 'initialisation' value for each element of the transition frequency matrix, before we compute the transition conditional-probability matrix using the formula given. The initialisation of the matrix in this way is justified in Section 3.6.1, where it is shown that it is equivalent to the use of a Bayesian prior distribution. Rather than initialise the matrix with an entirely arbitrary value, we assign the equivalent of a 'single transition' from each region. For example, if there are 20 regions then we initialise each element of the transition frequency matrix to 1/20 hence the sum of all transitions from any particular region is 1. Once the experiment is started, then each transition made by the subject is added to the appropriate element in the matrix. Thus, after the first transition is made by the subject, one of the transition frequency matrix elements will have the value $1 + \frac{1}{20}$.

As the Bayesian prior distribution for each row element in the transition frequency matrix is set to $\frac{1}{20}$, any transition from an element in that row is initially as likely as any other. However, once we obtain a single piece of empirical data, specifically one transition that passes through an element in this row, the value of one (1) is added to that element. In terms of Bayesian inference, the prior is the distribution of $\frac{1}{20}$s within the elements, and the one (1) represents observed data. This number $\frac{1}{20}$ represents the value of a_i in Equation 3.19 (page 89), and d_i is the observed data, i.e. the value of one (1), in the same equation. Therefore, as data is obtained, the posterior distribution is computed and (conceptually) fed back into itself as the new prior, this occurring repeatedly and updating the distribution with empirical data and moving away from the original prior distribution. As shown in Section 3.6.1, to obtain a posterior distribution from the prior, all that needs to be done is to replace the prior a_i with the sum of a_i and d_i. In this way, Bayesian inference allows us to move away from the prior distribution, toward an empirically based distribution. Because the prior distribution is originally set to contain uniform values (i.e. $\frac{1}{20}$ in every element of the rows), the original transition matrix is synthetic and unrepresentative of the eventual matrix. However, the prior distribution must be initialised somehow, as discussed in Section 3.6.1, and this is accepted as a necessity of Bayesian inference. The values a_i must also be chosen somewhat appropriately though, as their value relative to the incoming empirical data will determine how quickly they become 'swamped' by that data, in other words, how quickly the distribution will favour the empirical data rather than the (original) prior. In our case, the value we use ($\frac{1}{20}$) is smaller than even a single piece of empirical data (which, as we have just said, has the value of one), and as such the empirical data is highly favoured. Although a different (original) prior distribution to the uniform one we have used might better reflect the way persons tend to look ahead for instance, using an alternative

---

[2] A probability of zero would suggest that this transition is impossible, which may seem acceptable, but is not accurate (any transition is in theory possible for an IVE observer to perform), and leads to mathematical problems as stated.

prior distribution presents its own problems, such as determining not only the distribution used (such as Gaussian, or exponential), but also its parameters which are best determined by empirical data anyway. Using a uniform prior distribution also means that if there are elements in the presented IVE that bias the direction of gaze (such as some item that subjects find particular interesting to look at, but that is off-centre), then this can be immediately made apparent in the posterior even after sampling only one data point.

From the transition conditional-probability matrix, the conditional entropy may be computed as follows where n is the number of regions-of-interest (Brillouin, 1962):

$$H_c = -\sum_{i=1}^{n} p(i) \left[\sum_{j=1}^{n} p(i \rightarrow j) log_2 p(i \rightarrow j) \right], i \neq j \tag{4.2}$$

This equation requires the definition of three terms, namely $p(i \rightarrow j)$, $p(i)$, and $p(i, j)$:

- $p(i, j)$ is the probability of transitioning from region i to j. It may be considered as the number of i to j transitions divided by the total number of transitions.

- $p(i \rightarrow j)$ is defined as the conditional probability of transitioning from region i to j, given that the LOS is currently intersecting region i:

$$p(i \rightarrow j) = \frac{p(i, j)}{\sum_k p(i, k)} \tag{4.3}$$

- $p(i)$ is the marginal probability of foveating region i. It is estimated as:

$$p(i) = \sum_k p(k, i) \tag{4.4}$$

It should be noted that a transition matrix is produced using a number of observed transitions over time, and as such, H_c is computed over a 30-second sliding window at 1-second intervals. Specifically, we compute $H_c(t)$ over the observations recorded in the interval $(t - 30, t]$, where $t = 30..209$, and this forms our first scanpattern response variable.

(ii) Euclidean Distance

The Euclidean distance between scanpatterns is used as a simple metric reflecting the difference of transitions that were made in two scanpatterns.

To calculate the Euclidean distance between two transition frequency matrices, A and B, which are the same size and have elements $A_{i,j}$ and $B_{i,j}$ respectively, we use the following equation:

$$\Delta x_{A,B} = \sqrt{\sum_{i,j} |A_{i,j}^2 - B_{i,j}^2|} \tag{4.5}$$

As our transition matrix is produced using a number of observed transitions over time, Δx is computed as the difference between the transitions of two (temporally) adjacent 30-second sliding windows, sampled at 1-second intervals. Specifically, we compute $X(t) = \Delta x_{A,B}$ over the observations recorded in the intervals $(t-30, t]$ (which we denote matrix A) and $(t, t+30]$ (denoted, matrix B), where $t = 30...209$, and this forms our second scanpattern response variable.

4.7.3 Questionnaire Response

Questionnaires were administered before and after the experimental trial. After each replication of the experiment was completed the participants were provided with a Simulator Sickness Questionnaire (Kennedy et al., 1993), and a presence questionnaire based on that designed by Slater et al. (1994). Although only initial exposures were analysed previously, questionnaires were completed after the complete trial (for SE and NE subjects) This means that these subjects' reports would be biased by their second exposure, although we expect the data to be based mostly on their initial exposure, and this should be kept in mind. The questionnaires were designed to elicit subjective response data, which would act as dependent variables in regression analyses. This is the standard method of analysis for such questionnaires (Slater, 2004), see Section 2.4.4.

The questionnaire was only modified slightly from the original SUS by appending three questions at the end, so as not to interfere with its use; two of these questions were open ended. The first of these additional questions asked which objects the subject remembered, and the second, whether there were objects that they did not recognise (the final question asked what they thought any such unrecognised objects might have been.) These questions were added since it was thought that responses to the second question could correlate with subjective presence responses, and thus be a useful predictor variable in a logistic regression on presence scores. This is because the inclusion of unrecognisable objects in an IVE may reduce subjective presence scores, as these objects could prove to be distracting to those that notice them. Alternatively, persons more prone to overlook unrecognisable objects (consciously or subconsciously) may tend to provide increased presence scores. Finally, the remaining two questions were added because they were natural extensions of the above, and would thus be analysed in solely an explanatory mode for (potentially) future use.

The main 'presence' questions were as follows, each requiring a response on a 7-point Likert scale:

(i) There were times during the experience when the virtual environment became more real for me compared to the "real world"... (rated 'at no time'=1 to 'almost all of the time'=7)

(ii) The virtual environment seems to me to be more like... (rated 'images that I saw'=1, to 'somewhere that I visited'=7)

(iii) I had a stronger sense of being in... (rated 'the real world of the laboratory'=1 to 'the virtual reality'=7)

(iv) I think of the virtual environment as a place in a way similar to other places that I've been today.... (rated 'not at all'=1 to 'very much so'=7)

(v) During the experience I often thought that I was really standing in the lab wearing a helmet.... (rated 'most of the time I realised I was in the lab'=1 to 'never because the virtual environment overwhelmed me'=7)

Variable	*(Measure type) and extents/values*
Gender	(Binary)
To what extent do you use a computer in your daily activities?	(Likert 7-point) ‘Not at all’ to ‘very much so’
I have experienced virtual reality	(Likert 7-point) ‘Never before’ to ‘a great deal’
My expertise with computer or video games is	(Likert 7-point) ‘Complete novice’ to ‘Expert’
How many hours per week on the average do you spend playing computer or video games (if any)?	(Continuous Interval)
I achieved my tasks...	(Likert 7-point) ‘Not very well at all’ to ‘very well’
While in the virtual reality I was aware of background sounds from the laboratory	(Likert 7-point) ‘Not at all’ to ‘very much’
My status is as follows	(Nominal) ‘Undergraduate Student, Masters Student, PhD student, Research Assistant/Fellow, Systems/Technical Staff, Administrative Staff, Academic Staff, Other’
How dizzy, sick or nauseous did you feel resulting from the experience, if at all?	(Likert 7-point) ‘Not at all’ to ‘very much so’

See Appendix A for the complete questionnaire.

Table 4.1: Main explanatory variables

The questionnaire may be found in full in Appendix A.

4.8 Manipulated and Explanatory Variables

The manipulated variable is the environment displayed to the subject, which had three conditions: SE, NE and RE.

The questionnaires that were administered provide several explanatory variables. The questionnaires may be found in full in Appendix A.

4.9 Results

As we look at the results of the experiment, we shall restrict ourselves to considering just those experimental participants that provided eye-tracking data (n=28). In the following chapter we consider the influence of eye-tracking on the findings, versus analysis using head-tracking data only.

For all results, the initial period (of 30 seconds) of each of the subject’s data has not been included to allow for participant habituation, and data recorded for the same duration is removed from the end of each series as well.

For all statistical results, we set our significance level as $\alpha = 0.05$ as is customary.

4.9.1 Overview

In order to obtain an overview of our results we first present several graphs that illustrate the overall responses of the subjects – with respect to each of the experimental conditions. After this we consider the data in detail, performing statistical tests between subjects' data (Section 4.9.2).

Each graph here presents response variable data over the time domain to illuminate overall changes in the responses as the experiment progresses. We first show the skin-conductance response data, then we present the eye-head data, and finally both together.

Skin Conductance Responses - Overview

A total of fourteen (14) participants provided skin-conductance data for environments SE (n=8) and NE (n=6).

The normalised mean (across participants) of the skin conductance response data is computed as follows. The skin conductance response variable $S_k(t)$ is computed for each participant k. The (mean) average is then computed as:

$$S_{avg}(t) = \frac{1}{N_{condition}} \sum_k S_k(t) \tag{4.6}$$

This average for participants experiencing environments SE and NE is shown in Figure 4.7, where graphs are normalised according to the minimum and maximum values of the stress-inducing condition to allow comparison. It may be observed that there appears to be an increase in the number of skin conductance responses (SCRs) for (approximately) the second half of the experiment under the condition SE. Under the NE condition, although a burst of SCRs is apparent in (approximately) the latter half of the experiment, it is of a lesser magnitude and appears transient. The data is analysed in more detail in Section 4.9.2.

Eye Scanpattern Responses - Overview

We investigate eye scanpatterns using the eye-tracking and head-tracking data. (In the following chapter we shall also investigate the head-tracking data in isolation, which allows us to form an opinion as to whether head-tracking alone could be used to discern the changes in perception that are under investigation.)

As described previously (Section 4.7), each participant's LOS data is stored in a frequency transition matrix that we shall denote $f_{i,j}$. Two icosahedrons are used in order to capture two levels of detail for independent analyses, one having 80 faces and the other 20 (Figures 4.8 and 4.9 respectively). Each face represents both a row and a column in the transition matrix. The matrix elements thus contain the number of transitions made between the respective faces as the participant's Line Of Sight (LOS) moves around (as the LOS intersects the faces.)

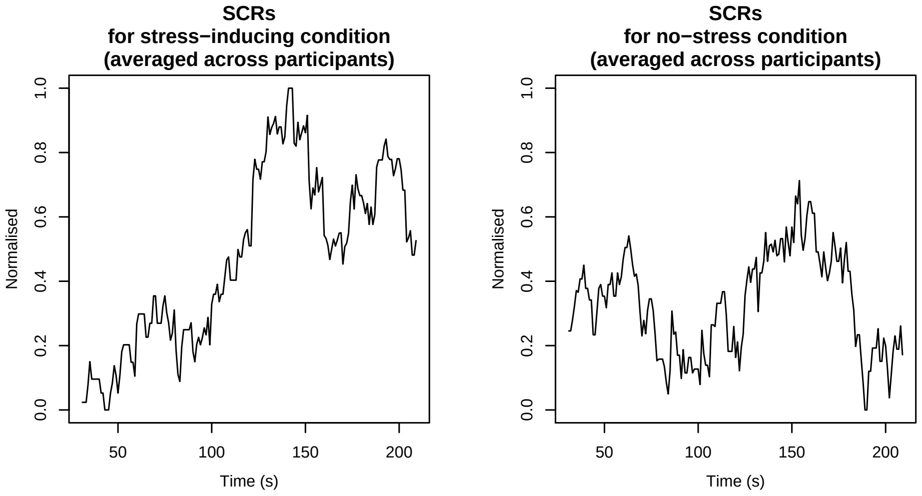


Figure 4.7: Skin Conductance Responses, SE and NE conditions (shown on same scale).

We can calculate the *approximate* number of segments found on a great circle over the icosahedron as follows. Using an eighty (80) segment icosahedron to divide the scene results in each segment having an area of $\frac{4\pi}{80}$ steradians. The relationship between the number of steradians and the half angle subtended by a solid angle is given as $\Omega = 2\pi(1 - cos\ d)$ (Henderson, 1997). Therefore, $d = arccos(1 - \frac{\Omega}{2\pi})$. Substituting for $\frac{4\pi}{80}$, we obtain an angle of $2d = 25.7$ degrees. Thus there are approximately $\frac{360}{25.7} \approx 14$ segments across any given great circle (e.g. horizontally, or vertically). In the case of a twenty (20) segment icosahedron, this becomes $\frac{4\pi}{20}$ steradians, $2d = 51.7$, and so on average we have $\frac{360}{51.7} \approx 7$ segments per great circle.

The measure we use is *Conditional Entropy Rate* (CER) as described in the response variables section (Section 4.7). In this overview, the response variable $H_k(t)$ is computed for each participant k and averaged as done for the skin conductance response (overview.) Hence, the (mean) average is computed as:

$$H_{avg}(t) = \frac{1}{N_{condition}} \sum_k H_k(t) \tag{4.7}$$

The results of this are shown for the conditions SE, NE and RE in figure 4.10.

It is useful to see how both response variables (SCRs and gaze scanpattern) develop together over time for the SE and NE conditions, and this is shown in Figure 4.11. In the second graph (NE condition), after the main scanpattern entropy drop at approximately 110 seconds, there appears to be a temporary increase in skin-conductance responses. However, in the first graph (SE) it can be seen that shortly after the main scanpattern entropy drop (at approximately 75 seconds), the skin-conductance response increases and does not return to the neighbourhood of its original distribution.

In Figure 4.11 there are differences in the initial CER measurements of the NE and SE conditions. This seems likely due to the following two facts. Firstly, that the (random) order of the triangles being introduced in a particular condition was kept the same for subjects within that condition in retrospect

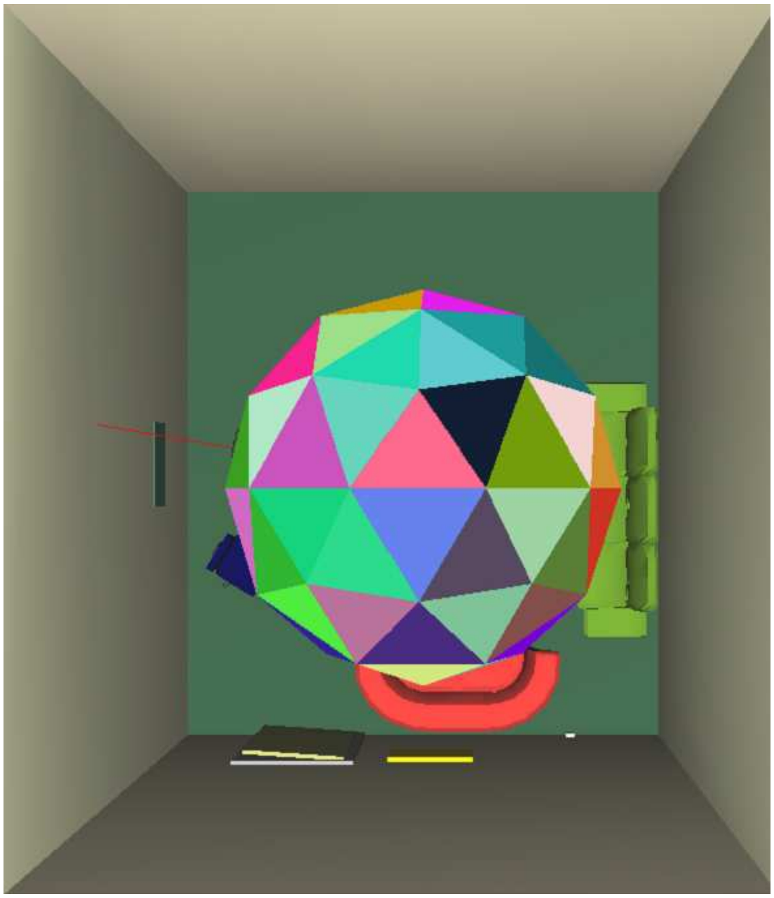


Figure 4.8: Icosahedron with 80 segments (i.e. faces).

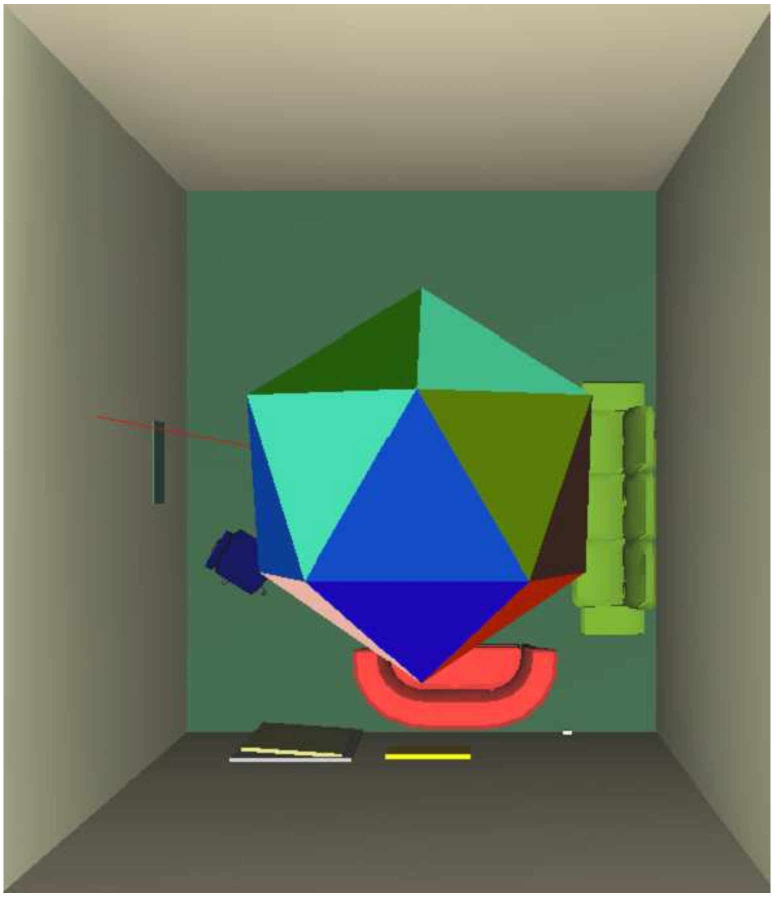
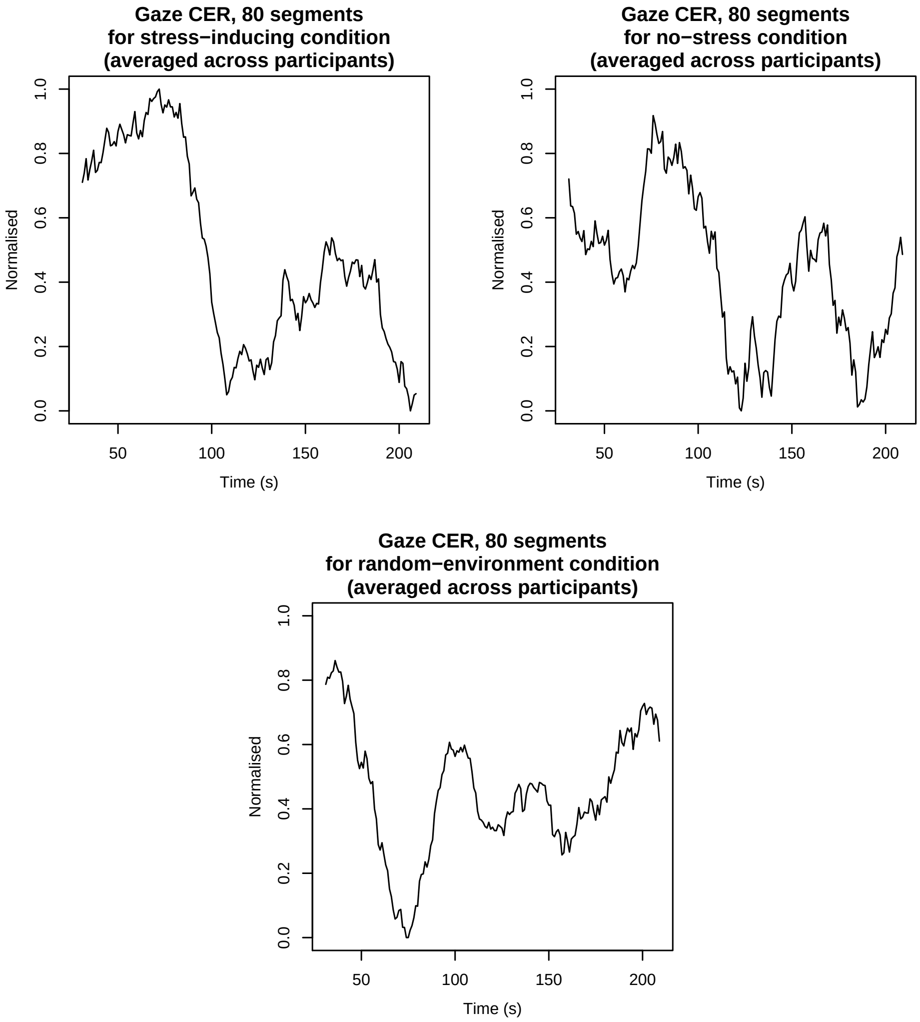


Figure 4.9: Icosahedron with 20 segments (i.e. faces).


Gaze CER, 80 segments for stress−inducing condition (averaged across participants)

Gaze CER, 80 segments for no−stress condition (averaged across participants)

Gaze CER, 80 segments for random−environment condition (averaged across participants)

Figure 4.10: Conditional Entropy Rate for SE, NE and RE (shown on same scale). Analysis performed using 80 segments.

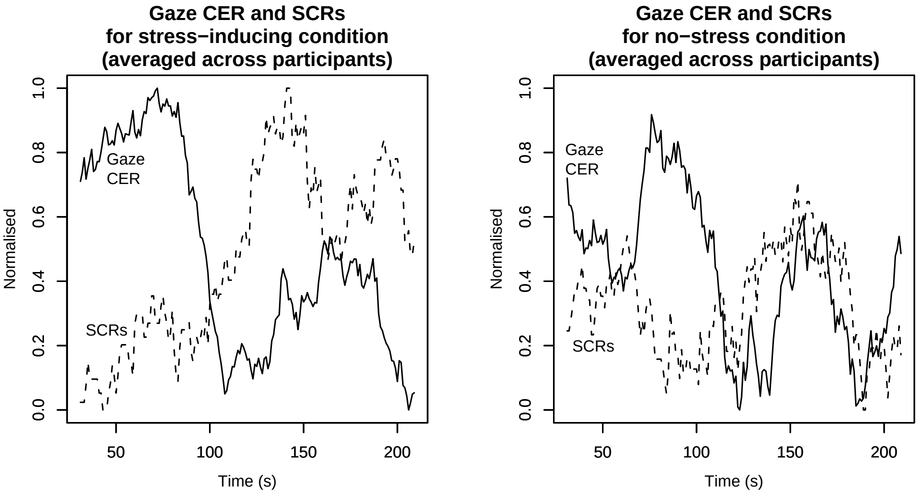


Figure 4.11: Conditional Entropy Rate and Skin Conductance Responses, for SE and NE conditions (shown on same scale). Analysis performed using 80 segments.

this not being ideal. Secondly, the NE and SE conditions had their triangles introduced a different order, with respect to the other (SE or NE) condition.

As we are interested in whether the SCRs increase whilst the scanpattern conditional entropy rate decreases, we directly plot one against the other to view any discernible change in both together, that is, occurring simultaneously. In Figure 4.12 we show the results of this for the SE and NE conditions, with the data normalised to lay in the interval [0,1] so that the two responses may be compared.

The direction of gaze for six randomly chosen subjects from each group is shown in Figure 4.13. Each white point is the end of a vector that lays on the 80 segment icosahedron. The origin of the vector is the at the Centre of Projection, which is roughly at the center of the icosahedron (not shown). The data used to produce these images are taken from times $t = 0$ to $t = 240$, that is, the entire duration of the experiment. Although the images can only be seen in two-dimensions, it is fairly clear that the distribution of gaze in the RE condition has the greatest entropy.

4.9.2 Detailed Analysis

Although we have an overview of our data, we must perform a more detailed analysis to actually test for a significant increase in SCRs for subjects in the SE condition in contrast to the NE condition, according to our alternative hypothesis. Similarly, we must test for decreases in the conditional entropy in SE and NE, in contrast to the RE condition. These tests are performed in the following two sections.

Skin Conductance Responses

Firstly, it is important to show that the increase in SCRs (under SE but not NE) is reflected by the individuals of the sample, and not only by their mean. We do this by considering differences between the response variables in the first and second halves of the experiment. The midpoint occurs at 120 seconds.

By considering the SCR response variable of the ith participant $S_i(t)$, (see Section 4.7) we may

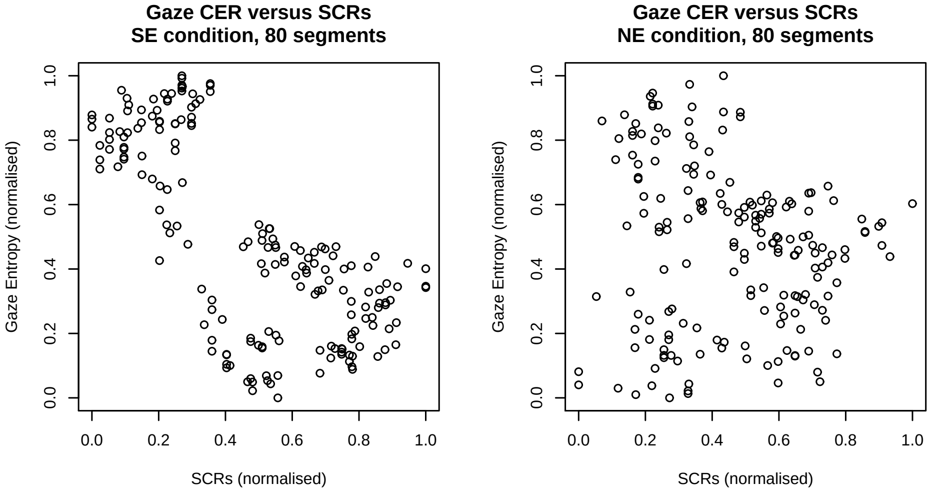


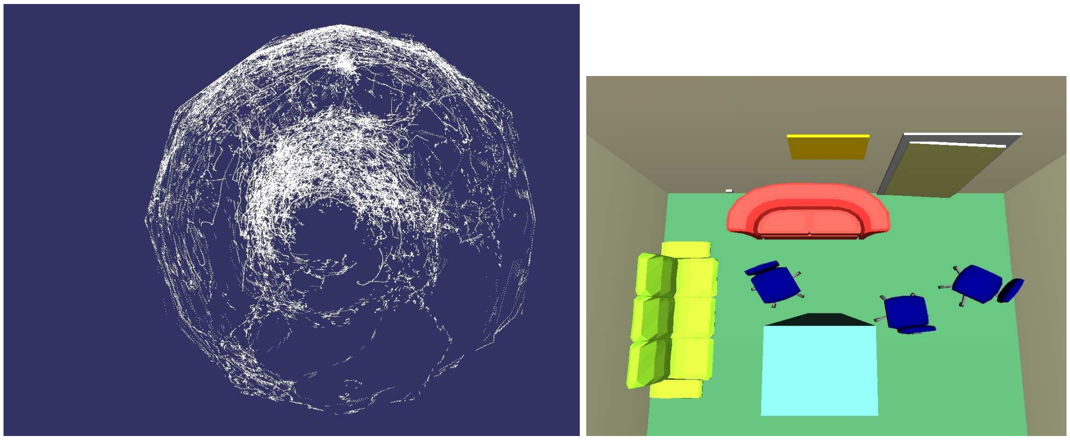
Figure 4.12: Conditional Entropy Rate versus Skin Conductance Responses, for SE and NE conditions (shown on same scale). Analysis performed using 80 segments.

compute the mean of $S_i(t-15)$ from time 60 to 119 seconds into the experiment, which we shall denote $s_{i,a}$, and similarly $S_i(t+15)$, the mean SCR rate from time 120 to 179 seconds into the experiment, which we denote $s_{i,b}$. Now, by considering all participants of an experimental condition, we may treat data from each individual as an observation from a binomial distribution, whereby if $(s_{i,b} - s_{i,a} > 0)$ we obtain a success, and a failure otherwise. In the truly random case, this distribution ought to have a probability of success with a binomial parameter $p = 0.5$.

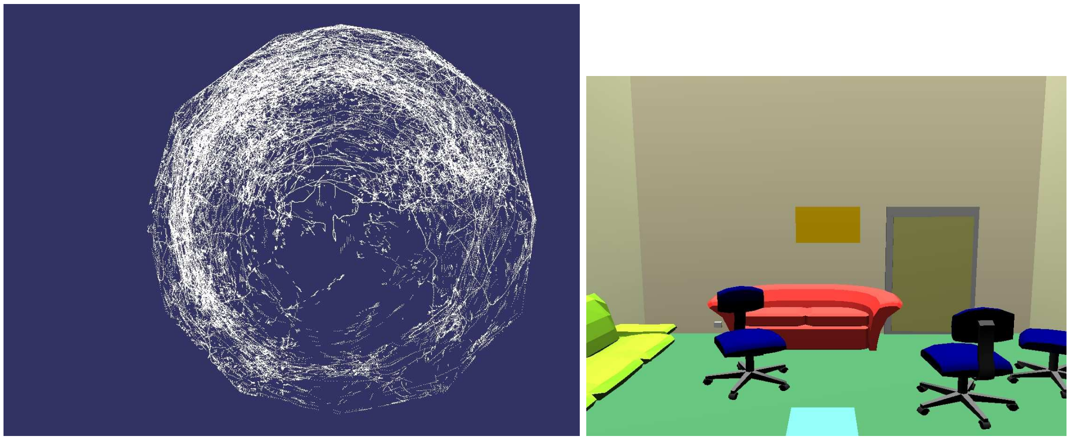
Our hypothesis under the SE condition stipulates that the SCR rate strictly increases. In practice then, we may utilise a one-sided sign test to determine whether our binary observations $(s_b - s_a > 0)$ support a binomial distribution with a parameter $p > 0.5$ (the alternative hypothesis). In doing so we obtain ($n = 8, p{-}value < 0.036$), which supports our alternative hypothesis that the SCR rate increases, thus rejecting the null hypothesis (for which $p = 0.5$).

Our hypothesis under the NE condition stipulates that the SCR rate will not increase. In effect, this means that the binomial parameter p will in this case be indiscernible from 0.5, or less than 0.5. So again we test whether the binomial parameter p is greater than 0.5, and we find that according to the data we can no longer reject the null hypothesis ($n = 6, p - value < 0.657$). In fact, we notice that the observations support a value of exactly 0.5 (the expected value of p according to the data, see Table 4.2), which if true would mean that the number of SCRs is just a likely to increase as decrease under the NE condition.

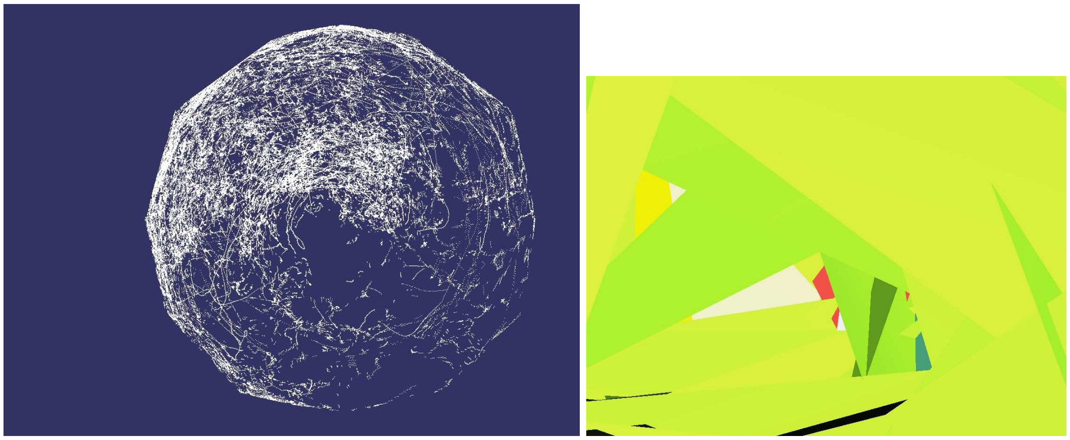
These results indicate that the NE condition had little or no effect in increasing the number of SCRs throughout the experiment, whereas there is evidence that the SE condition elicited an increased number of SCRs in the latter half of the experiment. The number of SCRs for each participant is shown in Table 4.2.



(a) SE



(b) NE



(c) RE

Figure 4.13: Aggregated gaze direction for six randomly chosen subjects from each group, SE, NE and RE.

Participant	Conditions	SCRs before mid	SCRs after mid	Difference
1	Stress-inducing	13	22	9
2	Stress-inducing	9	20	11
3	Stress-inducing	4	8	4
4	Stress-inducing	5	6	1
5	Stress-inducing	13	15	2
6	Stress-inducing	1	3	2
7	Stress-inducing	2	7	5
8	Stress-inducing	0	1	1
9	Neutral (no stress)	11	15	4
10	Neutral (no stress)	14	12	-2
11	Neutral (no stress)	11	14	3
12	Neutral (no stress)	14	16	2
13	Neutral (no stress)	6	4	-2
14	Neutral (no stress)	3	2	-1

Table 4.2: Number of SCRs for individual participants, before and after experiment's mid-point.

Eye Scanpattern Responses

Having determined that a significantly increased number of SCRs occurred between the first and second halves of the experiment under condition SE (but not significantly under NE), we may now examine the eye data we have obtained. An analysis must be performed to find evidence that there was also a difference in scanpatterns between the two halves of the experiment in condition SE (and NE for that matter) but not under RE.

As in the above skin-conductance response analysis, data falling within the two time periods on either side of the midpoint are used.

Transition frequency matrices are constructed at 1-second intervals, these being converted to conditional probability matrices using a Bayesian based algorithm, the details of which are described in Section 3.6.1.

One should note that the resulting conditional probability matrices do not qualify as fully-fledged models of the eye scanpattern per se, as the number of transitions is not deemed sufficiently large with respect to the number of matrix elements in order to justify a comprehensive model. The mean number of segment transitions made per participant is in fact shown in Table 4.3, and for clarity, graphically in the box and whisker plots, Figure 4.14. Instead measures over the matrices generated are used with the intention of showing significant commonality between participants' scanpatterns (i.e. the resulting matrices are used to construct our statistics using the methods described in Section 4.7.2).

Testing for a change in the scanpatterns (between the two time periods) is carried out using the (non-parametric) sign test, and defining a significance level of $\alpha = 0.05$. The response variable is either

		Segments ↓	
Environment	Participants, n	80	20
SE	8	186.5 +/- 48.1	119.6 +/- 29.7
NE	6	249.3 +/- 44.8	162.5 +/- 31.0
RE	14	214.5 +/- 20.2	133.3 +/- 21.6

Table 4.3: Number of segment transitions (mean +/- s.d.) made by individual participants.

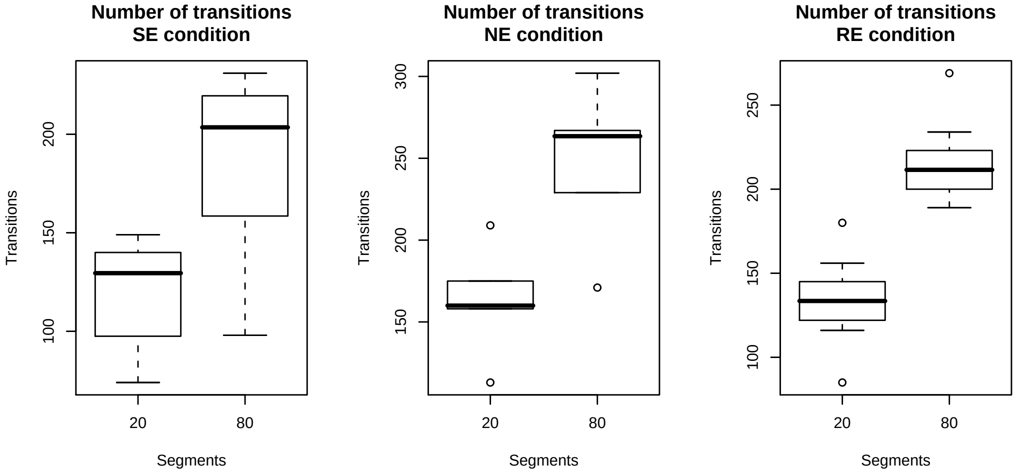


Figure 4.14: Segment transitions made by participants' line-of-sight, for SE, NE and RE conditions.

the change in Conditional Entropy Rate or in Euclidean distance (H or X respectively). We first examine the data using the Conditional Entropy Rate measure, then the Euclidean measure.

The transition probability models were tested for

1. a decrease in the Conditional Entropy Rate measures due to the independent experimental treatments SE, NE and RE, and
2. an increase in the Euclidean distance in the scanpatterns of SE and NE over those of the RE treatment.

Analyses are provided for both 20-segment and 80-segment icosahedrons.

4.9.3 The Conditional Entropy Rate measure

The CER average across participants has already been shown in Figure 4.10. It can be seen that, for the average, there is a decrease as expected for SE in contrast to the RE condition. This is also true for NE in contrast to RE, though it is less clear. In Figure 4.15 we again show the CER averaged across participants but when 20 segments are used rather than 80. Comparing these graphs with the previous ones that were computed using 80-segments, they may be seen to be similar in overall shape.

Just as in the case when analysing the Skin-conductance Responses in the previous section: it is important to show that the decrease in CER under conditions SE and NE but not RE, is reflected by the individuals of the sample, and not only by their means. We demonstrate this as follows:

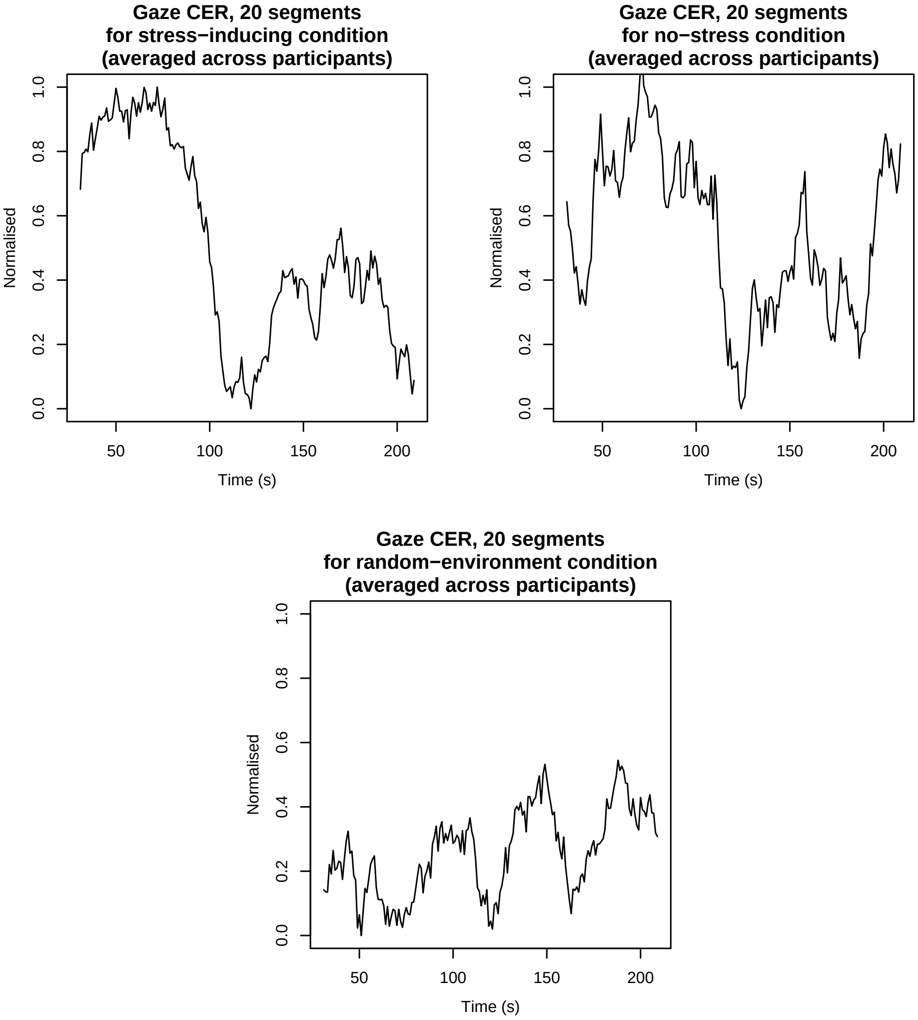


Figure 4.15: Conditional Entropy Rate for SE, NE and RE conditions (shown on same scale). 20 segments.

CER Measure H_a: Change in CER is -ve (i.e. CER decreases) Conditions: SE(n=8), NE(n=6), RE(n=14)		
	Segments ↓	
Condition ↓	80	20
SE	$n = 7/8, p - value <$ **0.036**	$n = 6/8, p - value < 0.145$
NE	$n = 6/6, p - value <$ **0.016**	$n = 5/6, p - value < 0.110$
RE	$n = 6/14, p - value < 0.788$	$n = 6/14, p - value < 0.788$

Significant p-values ($\alpha = 0.05$) in **bold**

Table 4.4: Change in CER, Eye & head data, both 80 and 20 segments.

By considering the CER variable of the i'th participant $H_i(t)$, (Section 4.7.2) we may compute the mean of $H_i(t - 15)$ from time $t = 60$ to $t = 119$ seconds into the experiment which we shall denote $h_{i,a}$. Similarly we compute $H_i(t + 15)$, the mean response rate from time $t = 120...179$ seconds into the experiment, and denote this $h_{i,b}$. Then, by considering all participants of an experimental condition, we treat each individual's data as an observation from a binomial distribution, whereby if $(h_{i,b} - h_{i,a} < 0)$ we obtain a success, and a failure otherwise. In the truly random case, this distribution ought to have a probability of success with a binomial parameter $p = 0.5$.

According to the general hypotheses for this experiment, we may now define three operational sub-hypotheses, one for each of the conditions, SE, NE and RE:

1. CER necessarily decreases under SE. (The binomial parameter p will be greater than 0.5 under SE.)

2. CER necessarily decreases under NE. (The binomial parameter p will be greater than 0.5 under NE.)

3. CER will remain at the same level or increase under RE. (The binomial parameter p will be indiscernible from 0.5, or appear to - significantly - have a value less than 0.5 under RE.)

Applying one-sided sign tests with each of these operational hypotheses, we obtain the results found in Table 4.4.

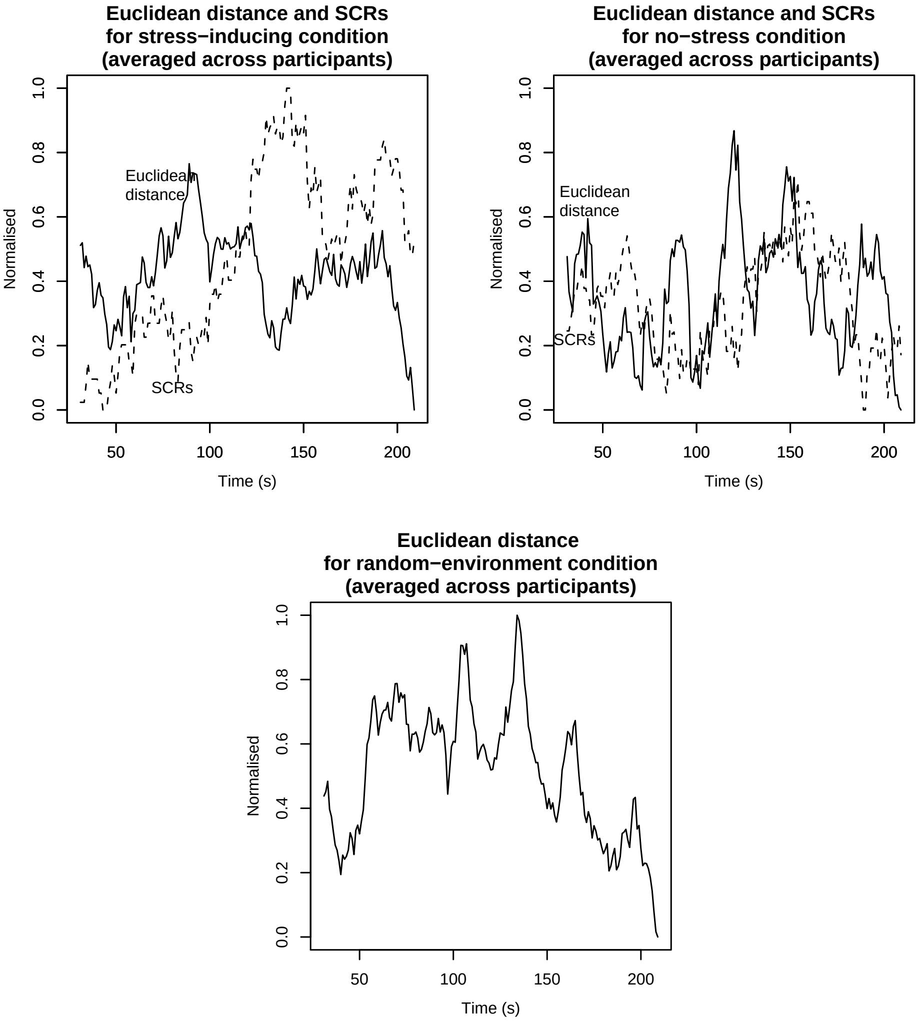
The results shown provide evidence supporting each of our sub-hypotheses, and thus the general hypothesis for this experiment. When 80 segments are used to classify the line of sight (LOS), the results are significant exactly as hoped. In contrast, using 20 segments does not provide significant results, but the data from each condition appear to be ‘approaching’ significance. That is, the pattern of ‘successes’ and/or the p-values in the table mirror those of the tests carried out when using 80 segments although they are not significant. The actual values of the changes in CER for all individuals and conditions, computed with both 80 and 20 segments, are presented in Table 4.5 and Table 4.6 respectively.

Participant	Condition	Mean CER Before Midpoint	Mean CER After Midpoint	Difference
1	SE	5.885	5.641	-0.244
2	SE	6.091	5.756	-0.335
3	SE	5.756	5.726	-0.030
4	SE	5.787	5.631	-0.156
5	SE	5.594	5.509	-0.084
6	SE	5.644	5.575	-0.069
7	SE	5.571	5.744	0.173
8	SE	6.288	5.610	-0.678
9	NE	5.450	5.408	-0.043
10	NE	5.428	5.217	-0.211
11	NE	5.300	5.263	-0.038
12	NE	5.423	5.405	-0.018
13	NE	5.497	5.356	-0.142
14	NE	5.681	5.649	-0.033
15	RE	5.518	5.379	-0.138
16	RE	5.545	5.599	0.055
17	RE	5.514	5.484	-0.030
18	RE	5.547	5.555	0.008
19	RE	5.642	5.511	-0.131
20	RE	5.408	5.384	-0.023
21	RE	5.445	5.583	0.138
22	RE	5.733	5.620	-0.113
23	RE	5.575	5.428	-0.147
24	RE	5.490	5.813	0.323
25	RE	5.245	5.278	0.033
26	RE	5.503	5.683	0.180
27	RE	5.523	5.573	0.050
28	RE	5.510	5.541	0.031

Table 4.5: The effect of ‘minimal visual cues’ in EyeHead Scanpatterns, 80 Segments.

Participant	Condition	Mean CER Before Midpoint	Mean CER After Midpoint	Difference
1	SE	3.608	3.522	-0.087
2	SE	3.963	3.702	-0.261
3	SE	3.520	3.647	0.127
4	SE	3.562	3.504	-0.058
5	SE	3.559	3.148	-0.411
6	SE	3.561	3.262	-0.299
7	SE	3.476	3.578	0.102
8	SE	4.268	3.425	-0.843
9	NE	3.209	3.262	0.053
10	NE	3.176	3.096	-0.081
11	NE	3.232	3.039	-0.193
12	NE	3.354	3.177	-0.177
13	NE	3.423	3.290	-0.133
14	NE	3.644	3.563	-0.082
15	RE	3.319	3.274	-0.045
16	RE	3.481	3.505	0.025
17	RE	3.452	3.425	-0.027
18	RE	3.406	3.490	0.085
19	RE	3.465	3.425	-0.040
20	RE	3.292	3.274	-0.018
21	RE	3.426	3.502	0.076
22	RE	3.697	3.555	-0.142
23	RE	3.476	3.302	-0.174
24	RE	3.478	3.567	0.089
25	RE	3.196	3.308	0.112
26	RE	3.354	3.576	0.222
27	RE	3.392	3.576	0.184
28	RE	3.392	3.567	0.176

Table 4.6: The effect of ‘minimal visual cues’ in EyeHead Scanpatterns, 20 Segments.


Euclidean distance and SCRs for stress−inducing condition (averaged across participants)

Euclidean distance and SCRs for no−stress condition (averaged across participants)

Euclidean distance for random−environment condition (averaged across participants)

Figure 4.16: Participant Means: Euclidean Distances for SE, NE and RE conditions (shown on same scale). 80 segments.

4.9.4 The Euclidean distance measure

Unlike the measure in the previous section, the Euclidean distance is computed *between two* scanpattern transitions, using (temporally adjacent) sliding windows, each window containing transitions from a period of 30 seconds — see Section 4.7.2 for further details. The resulting distances may thus be graphed (Figures 4.16 and 4.17), using both 80 and 20-segment icosahedrons for line of sight classification into potential regions of interest as before.

The changes in the Euclidean distance as presented in the figures show wide variation, and provide no clear indication that this response variable relates to the SCR variable. This is true for both the 80 and 20-segment cases. However, in order to have some confidence that there is little if any value of this measure, we must apply the appropriate statistical tests pertaining to our hypothesis.

By considering the Euclidean distance variable of the i'th participant $X_i(t)$, (see Section 4.7.2) we may determine the time t at which there is a maximum value of $X_i(t)$ from the potential values $T = 60...179$ seconds into the experiment. Then, by considering all participants of a single experimental

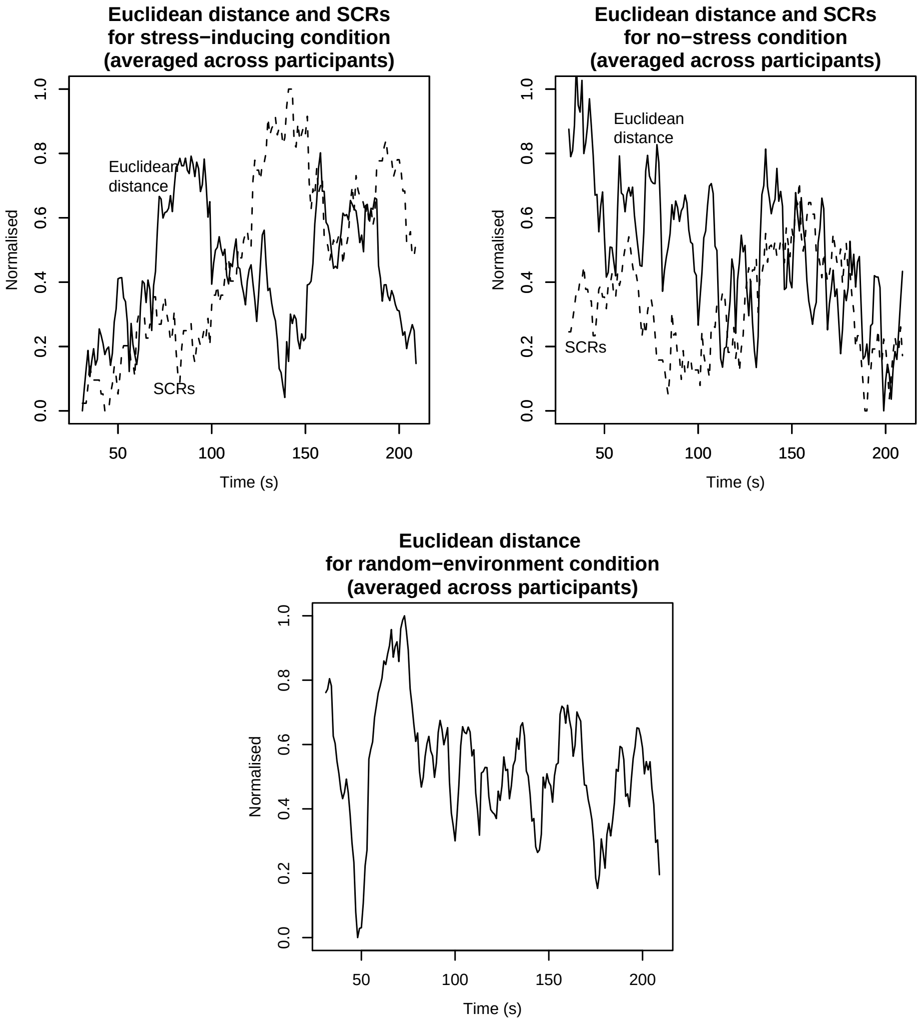


Figure 4.17: Participant Means: Euclidean Distances for SE, NE and RE conditions (shown on same scale). 20 segments.

Euclidean Measure H_a: Maximum $X_i(t)$ occurs at $t < 120$ (i.e. Gestalt occurs before experiment's midpoint) Conditions: SE(n=8), NE(n=6), RE(n=14)		
	Segments ↓	
Condition ↓	80	20
SE	$n = 6/8, p - value < 0.145$	$n = 4/8, p - value < 0.637$
NE	$n = 3/6, p - value < 0.657$	$n = 3/6, p - value < 0.657$
RE	$n = 5/14, p - value < 0.911$	$n = 8/14, p - value < 0.396$

Table 4.7: The effect of 'minimal visual cues' in EyeHead Scanpatterns.

condition, we treat each individual's data as an observation from a binomial distribution, whereby if $t < 120$ we obtain a success, and a failure otherwise. In the truly random case, this distribution ought to have a probability of success with a binomial parameter $p = 0.5$.

We may now define operational sub-hypotheses for each of the conditions, SE, NE and RE:

1. The greatest distance in scanpatterns will occur before the experiment's midpoint under SE. (i.e. The binomial parameter p will be greater than 0.5 under SE.)
2. The greatest distance in scanpatterns will occur before the experiment's midpoint under NE. (i.e. The binomial parameter p will be greater than 0.5 under NE.)
3. The greatest distance in scanpatterns will occur randomly throughout the duration of the experiment under RE. (i.e. The binomial parameter p will be indiscernible from 0.5 under RE.)

Utilising a one-sided sign test, we obtain the results presented in Table 4.7. However, the results of the tests are not significant throughout, and as such the measure appears to be of no value for rejecting our overall hypothesis as the Euclidean is unable to discern between the conditions. As a check, the test was re-applied after smoothing the data using a moving average filter (filter lengths 2, 4, and 8-seconds were all tested), should there be any noise in the data to bias the outcome. This filtering had negligable effect on the results, the only difference being when the 8-second filter was used: under the 80-segment analysis, this resulted in one less success in the SE condition, and one additional success in the NE condition. Under the 20-segment analysis there was one more success in the NE condition.

4.9.5 Clustering

We have obtained evidence that supports our hypotheses utilising the SCR and CER measures, but now we return to the data plotted earlier, which we present again but with graphs for both 20 and 80 segments shown, Figure 4.18 and Figure 4.19 . The two measures are plotted against each other, first normalising the CER and SCR data such that they all belong to intervals [0,1]. Under the SE condition, we observe the potential presence of two tenuously connected but distinct clusters.

To formally dissect the two clusters of the SE condition we may apply a two-dimensional hierarchical clustering method. The method is characterised as having 'complete-linkage', see Section 3.6.2. The

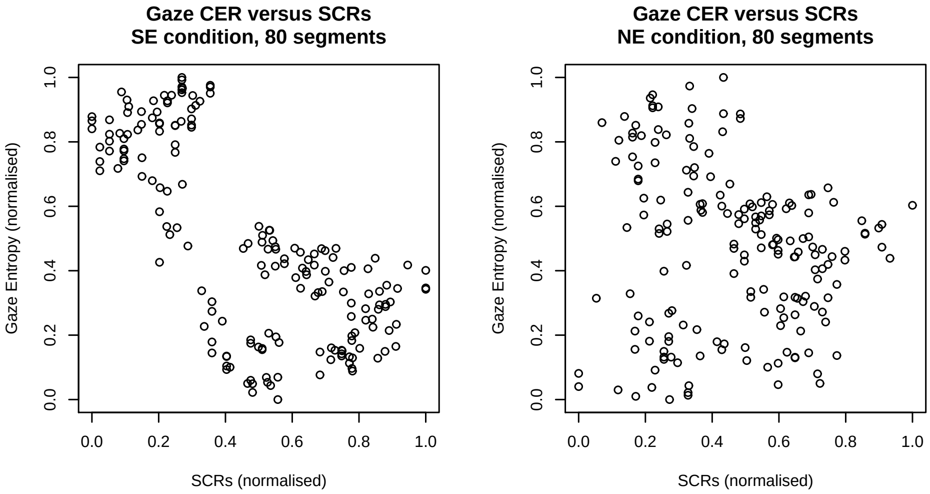


Figure 4.18: (Normalised) CER versus SCRs, SE and NE conditions (shown on same scale). 80 segments.

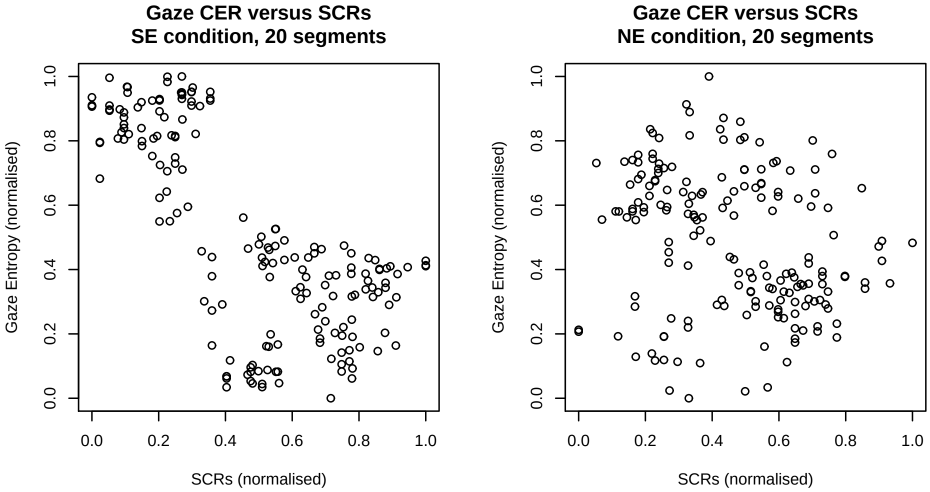


Figure 4.19: (Normalised) CER versus SCRs, SE and NE conditions (shown on same scale). 20 segments.

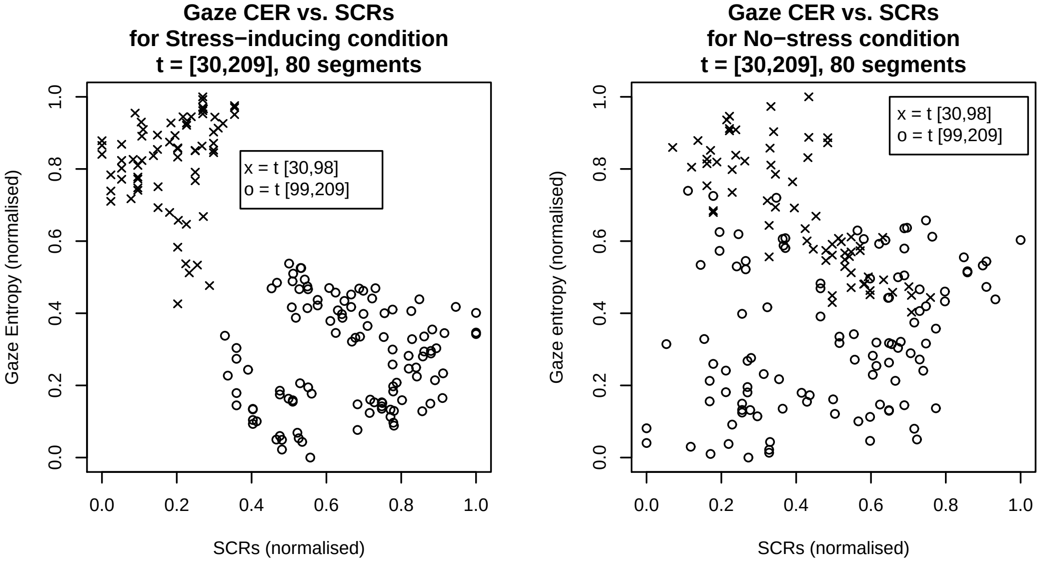


Figure 4.20: *Left:* (Normalised) CER versus SCRs with SE divided into two groups using Hierarchical Clustering (complete-linkage). *Right:* The same graph is shown but for condition NE. The same division of points is used (found by clustering the SE data), allowing comparison.

reason for doing this is to objectively split the data into two groups in order to find the greatest discontinuity occurring (simultaneously) in the measures. Although clusters are apparent in 20 and 80-segment analyses, we restrict ourselves to consider the 80-segment case, as this is likely more accurate as we found significant differences between the first and second halves of the experiment, but not quite in the 20-segment case.

The Hierarchical Clustering algorithm is applied, after which the final two clusters remain, and these are shown in Figure 4.20. Unsurprisingly, the two clusters computed reflect the two apparent clusters we observed earlier (Figure 4.18 and Figure 4.19). The time of the last observation in the first group is 98 seconds, and the earliest observation in the second group is 99 seconds, and so the groups are mutually exclusive. That there are these two distinct groups supports our operational hypothesis that the CER would decrease as the SCRs increase.

4.9.6 Questionnaire Results

All participants completed questionnaires prior to and subsequent to the experimental trial. The data collected included firstly general demographical information that could be used to form explanatory variables in the analysis phase, and secondly data to gauge the participants' (immediate) state using the Simulator Sickness Questionnaire (Kennedy et al., 1993). The remaining data collected by questionnaire was used to construct a subjective response measure of presence. Together with the manipulated variable (the experimental condition) these would be used to perform regression analyses.

The responses to the presence-specific questions are presented for overall comparison in the box and whisker plots of Figure 4.21. Carrying out a Kruskal-Wallis test to analyse the presence scores between questions and conditions, we find no significant difference between them. However, if the responses are

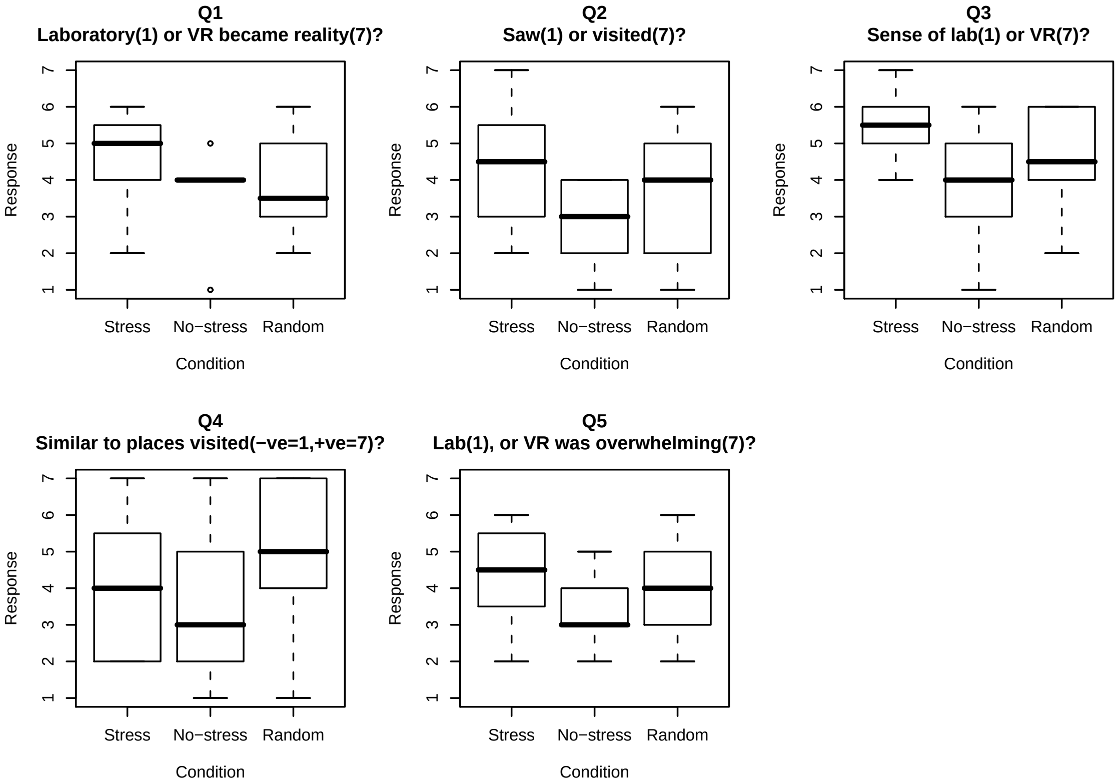


Figure 4.21: Questionnaire: Presence Responses

combined to produce a single, overall, 'presence score' for each person, we find a significant difference between the conditions ($\chi^2 = 6.09$, $dof = 2$, $p < 0.048$). The actual medians and interquartile ranges for the responses to the individual questions are provided in Table 4.8.

Next, a regression is performed to analyse how the presence questionnaire responses vary with respect to:

1. the manipulated variable (i.e. the condition)
2. the demographic factors recorded by the general questionnaire
3. the objective eye-scanpattern responses measured previously.

In order to perform the regression analysis with the questionnaire response variable, which is initially in ordinal form, it is transformed to a binomial response with which a logistic regression may be applied. To do this, the ordinal (Likert) scales are transformed by counting the response to a question that is above (or below) a threshold as a success in a Bernoulli trial. Specifically, the binomial count response variable represents the number of answers to the questionnaire items having a response greater than 4 (which was the neutral response.) Although this transformation leads to an inevitable loss of information (and thus) statistical power, it is straightforward and deemed an acceptable method to perform a regression using the ordinal response variable — indeed this method has been a de-facto standard for some researchers analysing such data in presence research (Slater and Garau, 2007).

Presence Questionnaire — Summary of Likert responses			
Question/Likert Statement	Condition	Median	Interquartile Range
Q1. VR became the reality?	Stress Environment	5.0	0.75
	Neutral Environment	4.0	0.00
	Random Environment	3.5	2.00
Q2. Saw or visited?	Stress Environment	4.5	2.25
	Neutral Environment	3.0	1.50
	Random Environment	4.0	2.50
Q3. Sense of VR or lab?	Stress Environment	5.5	1.00
	Neutral Environment	4.0	2.00
	Random Environment	4.5	2.00
Q4. Similar to places visited?	Stress Environment	4.0	3.25
	Neutral Environment	3.0	2.25
	Random Environment	5.0	2.50
Q5. Lab, or was VR overwhelming?	Stress Environment	4.5	1.50
	Neutral Environment	3.0	0.75
	Random Environment	4.0	2.00

Table 4.8: Summary of Presence Questionnaire Responses on a 7-point Likert scale.

A backward and forward stepwise (AIC) regression selection procedure was used to determine the independent variables of the logistic regression. This was performed in R, using the stepAIC function (of the MASS library). Potential explanatory variables were: all items on the presence questionnaire (barring the presence specific questions themselves), items from the demographic questionnaire, and the result of the Simulator Sickness Questionnaire. The potential, and then selected independent variables are shown in Table 4.9. The overall fit of the selected model was reasonable, explaining over 64% of the variation ($R^2 = 0.64$).

As would be hoped, the manipulated variable was found to affect the subjective presence scores. The SE and RE conditions elicited greater presence scores than the NE condition, being significantly increased by over 1 point. Presence scores from the SE and RE conditions were however mutually indiscernible, neither of them affecting the presence score any more or less than the other. This finding is inconsistent with the results of the objective eye-scanpattern response, whereby the SE and NE conditions were found to elicit similar responses in contrast to the RE condition. It is of note that when asking subjects about their experience in terms of presence, that greater scores are allocated to the more unusual environments than the more typical scene (NE). Thus the sense of novelty could be provoking an increased perception being ‘present’ in the IVE. More likely though (if the objective results are taken into account) the novelty of an environment is biasing the way in which the experience is being reported.

While it might seem that students were prone to scoring higher on the presence related questions,

Potential and Selected Independent Variables				
Variable	Selected?	Estimate	Std Error	p-value
(Intercept)	Y	-0.9553	0.8548	0.2638
SE condition	Y	1.5369	0.6986	0.0278
RE condition	Y	1.3084	0.6489	0.0438
Current status: Masters Student	Y	0.5930	0.5830	0.3090
Current status: PhD Student	Y	0.3931	0.7898	0.6187
Current status: Academic Staff	Y	-1.5651	0.8980	0.0814
Current status: Other	Y	-1.4281	0.6706	0.0332
Computer game hours per week?	Y	-0.1710	0.0847	0.0435
Felt nauseated after VR?	Y	-0.5159	0.2526	0.0412
Saw unrecognisable objects in VR	Y	1.1738	0.4677	0.0121
Gender	N			
Extent of previous VR experience?	N			
Extent of daily computer use?	N			
Computer game expertise	N			
SSQ score	N			
Subjectively self-scored task performance	N			
Measured change in CER	N			

Table 4.9: Independent-variable selection for logistic regression. The baseline condition was NE, and the ‘Current status’ baseline is set as ‘Undergraduate’.

only those from a non-academic background provided data that significantly affected the presence score, decreasing it.

While it is unsurprising that the number of computer-game-hours affected the presence score, the size of the effect seems low compared with the other factors of the regression. The direction of the effect is as expected, with more time spent playing computer games leading to a small decrease in the expected presence score.

It is interesting that the SSQ score did not appear in the regression, while the question “How dizzy, sick or nauseous did you feel resulting from the experience, if at all?” does. It could be that this single question appears to desire a more subjective and informal response than the direct and solely symptomatic SSQ. One would otherwise expect these two variables to have a similar effect on the presence score.

Positive responses to the question whether subjects did not recognise any objects in the environments had a significant positive effect on the presence score. In fact, the strength of this effect is only comparable to that of the manipulated variable itself. Although purely retrospective conjecture, it seems that a person with a lower threshold for the ‘suspension of disbelief’ may be more suggestible and thus less likely to consciously question oddities within the environment. It seems unlikely that only some people noticed the less defined elements and therefore had a reduced sense of presence, because there were ambiguous elements clearly present in every environment. For instance, the ‘empty picture frames’ on the wall, and of course all the elements in the random environment condition. How people interpret badly defined elements of an otherwise perceptible environment seems an interesting, and open, problem for future research (see Section 2.5.5.)

Although there was no hypothesis regarding the questions of which objects subjects remembered, the data has been collected and classified. The objects reported were: sofa, chairs, door, wall, ceiling, wall-socket, picture, virtual-body, column, and triangles.

As might be expected, there was no (overall) significant difference between the ‘everyday’ objects reported in SE and NE conditions ($N_{SE} = 28$, $N_{NE} = 27$, Fisher’s Exact Test, two-sided, $d.o.f. = 9$, $p = 0.890$), while subjects viewing RE only reported seeing triangles[3] and the virtual-body. However, there was a single notable difference in reporting the virtual-body between RE ($virtualbody_n = 7$) the SE ($virtualbody_n = 2$) and NE ($virtualbody_n = 1$) conditions. Thus, it would seem that the familiarity of seeing one’s own body made a more striking impression when there was little else to recognise however, this is just conjecture. It is also somewhat interesting that five (5) of the fourteen (14) respondents in the RE condition remarked that they did not recognise what the displayed triangles were. It is not clear what this might indicate, but the fact there appears to be such a subset within the respondents could be of interest to researchers investigating abstract environments.

---

[3] Of course, no SE and NE subjects reported seeing triangles.

4.10 Conclusions

This investigation set out to find evidence for the concept of minimal visual cues, and secondarily to test a novel perceptual-presence measure based on gaze. Recapitulating, our operational hypotheses (Section 4.2) were:

1. “Once the minimal visual cues are presented, a perception of a meaningful virtual environment is achieved, and a stabilised perception should commence. We hypothesise that this will be reflected by a significant drop in the entropy of eye-in-head movements.”

2. “Once minimal visual cues are presented, a perception of a meaningful virtual environment is achieved, and there should be a physiological stress response given the premise that the environment is designed and is able to provoke such a physiological response.”

The evidence as to whether or not there is a threshold at which minimal visual cues are present is borne firstly by our physiological (EDA) results. The increase in the SCRs due to stress is marked, as evidenced by the graphs (Figures 4.7 and 4.20) and the statistical tests showing a significant increase in the skin conductance responses after our expected point of minimal visual cues in a stress-inducing environment, and this contrasts with the stress-neutral (NE) environment.

The results support the hypothesis that entropy will decrease when there is high-level perception, that is, after a point of minimal visual cues. Although only the 80-segment analysis supported our hypothesis with statistical significance, the 20-segment analysis followed the same trend, and so it is suspected that given a larger sample size, significance may have been reached under this analysis parameter as well. If we consider just the 80-segment analysis, then whilst the CER measure showed decreases in CER around the midpoint of the experiment for the SE and NE conditions, there was no discernible change in CER for the RE condition. This concurs with the thesis of Gregory, since in the RE condition there could hardly be successful perceptual selection, or to put it another way, the stimulus could not be interpreted as something meaningful. Hence, if presence involves a perceptual selection response, then CER appears to be a potential presence indicator. These findings are in-line with the results of Yarbus (1967), Buswell (1935), and Stark et al. (1992) (see Section 2.5.4), who predict that the scanpattern over a perceived stimulus has repetitive components. But they are also aligned with Gregory’s (1977) theory that until a meaningful perception is achieved, the evidence (i.e. stimuli) is continually examined to develop hypotheses. This examination process that occurs before perceptual selection would produce a scanpattern with greater entropy, and this is evidenced by our results.

The cluster analysis also supports the hypothesis, providing us with the most likely time of the onset of increased SCRs, between 98 and 99 seconds according to the 80-segment analysis. At this point in time only 65% of the environment’s polygons are visible (Figures 4.22, 4.23 and 4.24 show the state of the environment for each condition at this point in time). Hence there was evidence of presence before all the polygons of the SE condition were displayed, suggesting that minimal cues were already displayed. This estimated time at which the transition seems to occur concurs with our hypothesis, because it is before the environment is fully developed, implying that a minimal cues threshold has been exceeded.

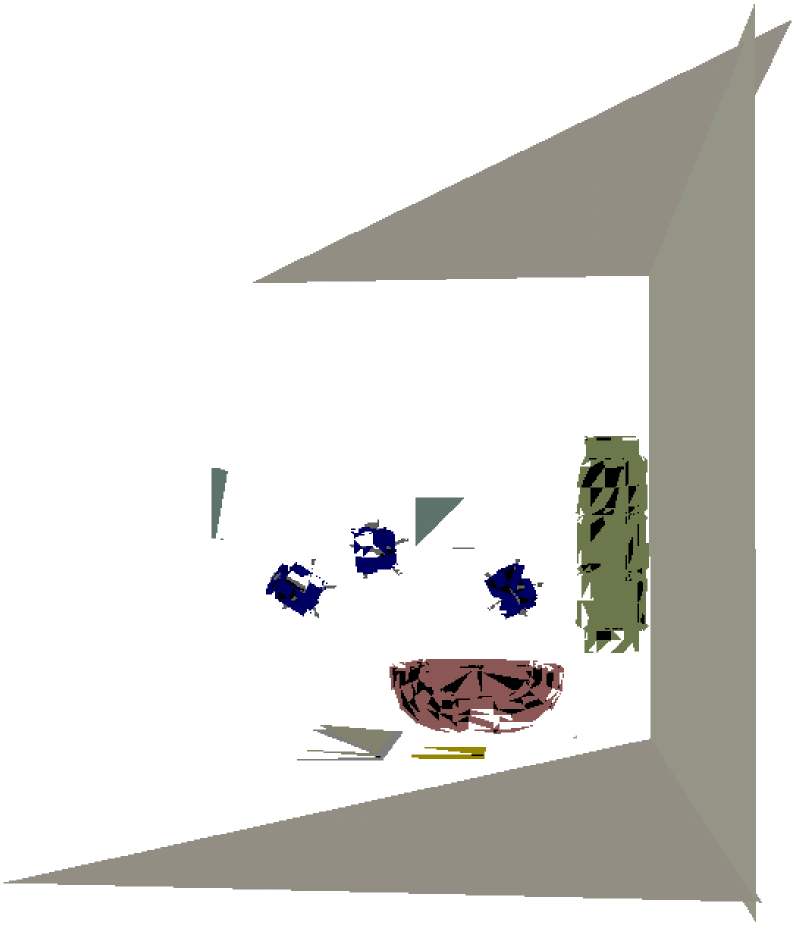


Figure 4.22: Top down view of condition SE as displayed between $t = [96, 100)$, showing 65% of the total environments' polygons.

Figure 4.23: Top down view of condition NE as displayed between $t = [96, 100)$, showing 65% of the total polygons.

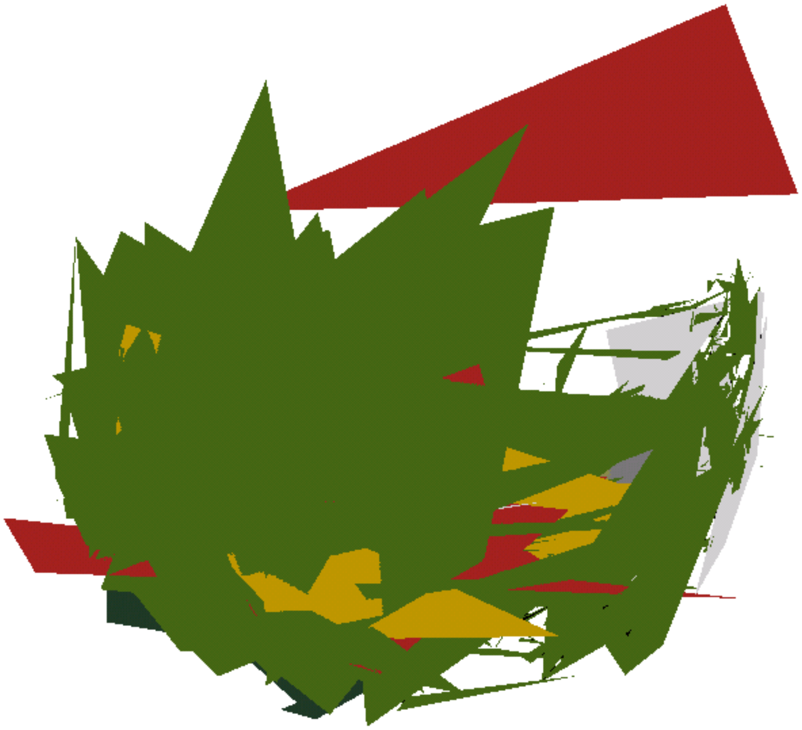


Figure 4.24: Top down view of condition RE as displayed between $t = [96, 100)$, showing 65% of the total environments' polygons.

The fact that the series of observations in the two clusters are consecutive with no overlap between clusters implies a strong correlation between the two measures, as there is a one-way transition between states. The duration of this transition appears to be of up to a maximum of 11 seconds, as estimated from the number of apparent observations made between the two clusters in Figure 4.20. This 'trail' between the clusters is likely to be an artefact of the process used to produce the graph itself, the averaging across participants, and because the response would occur at slightly differing times for each participant. This change is also indicative of a discontinuity that would accompany the expected Gestalt, or sudden shift in perception. Again this supports the minimal cues theory as there would otherwise be a linear transition between the states, forming a single cluster stretched in the expected direction of the CER and SCR changes.

The eye scanpattern measures, the Euclidean distance and the CER, did not perform as expected. Unlike the CER measure, the Euclidean distance was in fact developed post-hoc (with respect to the experiment) to determine whether a more simplistic measure might produce similar results to the CER. Due to the much simpler and direct approach of using a Euclidean measure, it was thought that it would provide a weak but significant result. In contrast, the CER measure reflects the transitions of the eye scanpattern less directly (not comparing specific transitions in a like-for-like manner) — although it much better reflected the concepts and rationale behind our hypothesis. In practice though, whilst the Euclidean measure could not distinguish between the former and latter scanpatterns, the CER measure exceeded expectations and appears to have been able to discern between the immersive virtual environments presented. This result supports the concepts that underlie our scanpattern hypotheses, the concepts suggested by Ellis & Stark, and Gregory. The Euclidean Measure does not feature in the remainder of this thesis.

Regarding the number of ROIs, in theory, the CER could be computed for a 2x2 transition model, which could represent exactly two (2) ROIs. In practice, Ellis and Stark (1986) was able to successfully use the CER measure to differentiate scanpath entropy values using just eight (8) ROIs. Given the extensive stimulus of the IVE (i.e. that it entirely surrounds the observer), this number could certainly be considered as an initial lower bound on the number of ROIs for which we can usefully compute the CER. On the other hand, the number of segments (ROIs) that we use should not be too large, as we would have to sample an exponentially increasing number of transitions for our model, as explained in Section 3.6.1. The number of segments we have used (80 and 20) was been relatively successful, but the ideal number would undoubtedly be dependent upon the specific IVE to some extent. However, there appears to be a fair degree of tolerance, given that the effects we expected to see were apparent to some extent using either 80 or 20 segments, and there is no reason to believe that these values could not be used in other similar experiments. The most important characteristic of the segmentation of the scene is that it divides regions that contain salient locations, and our original “guesstimates” of 80 or 20 segments appear to have been successful.

In considering the questionnaire responses, it has been pointed out by Slater and Garau (2007) and Gardner and Martin (2007) that the utilisation of ordinal scale response data in a regression context must be carefully considered. As such we utilised the considered approach of the SUS analysis method that entails a logistic regression analysis.

It is interesting that the questionnaire could discern differences in the presence score between {SE,NE}, and {RE,NE} but could not discern a difference in presence scores between SE and RE — whilst the objective CER measure discerned both SE and NE to be distinct from the RE condition. Since the RE condition was a *meaningless* random-polygon environment, this raises further doubts about sole reliance on ‘presence questionnaires’ that are based on questions that are not sufficiently defined. This backs up the finding of Slater (2004) that people will always find a way to interpret a questionnaire as meaningful, even if it is not. Although just conjecture, it could be that the questionnaire is eliciting a subjective score of vividness in terms of novelty, as this is the most obvious way in which the NE condition differs from SE and RE: the NE condition presents the most plain or unsurprising environment and is associated with the lowest presence scores.

It seems that the most influential (significant) regression predictor variable was the manipulated variable: the environment condition. After this, there are four other explanatory variables that appeared in the regression model: ‘Unrecognised objects’, ’Dizzy/nauseated’, ’Computer game hours’, and ‘Status’.

Of course it is not surprising that the reporting of ‘Unrecognised objects’ was found to lead to a decrease in the presence score as this would likely detract from the experience. However, the RE condition was scored higher than NE, even though it had no recognisable objects at all. But in this case it is probable that the question over unrecognisable objects might be interpreted differently, as it may be considered to be asking whether the subject recognised the triangles. In other words, it is possible that the use of 'Unrecognised Objects' as a predictor variable could be related to the propensity of the participant for cognitive dissonance (keeping their memory of the experience consistent with how they think they should answer the questionnaire). However, this requires further investigation. Because of this, the presence score and the 'Unrecognised objects' variable were tested for correlation for the SE and NE conditions. Both environments still included objects that were, by design, ambiguous and unlikely to have been recognised. The ambiguous items were clearly displayed blank 'picture frames' placed on the wall, which were in fact just singly coloured (yellow) cuboids. The correlation was slightly too weak to be significant (Pearson $r = 0.52$, $dof = 13$). Nevertheless, the overall finding is that the acknowledgement that one did not recognise objects in the environment was related to lower scoring of presence.

Of the demographic questions that remained in the stepwise regression, the 'Current Status' variable had the greatest effect on the presence score. It had a significant negative relationship to the presence score when the response was 'Other'. This particular response indicated that the subject was not a student, nor an academic, nor a researcher, nor part of academic support staff. Why this should reduce the presence score is unclear though of interest, but to probe this question a larger study would be required to obtain more data. Such a study should investigate the possibility that bias was due to differences in how the questionnaire is understood and/or approached, or whether due to a more objective factor.

Subjects were asked whether they felt dizzy, sick or nauseated after the experience. This predictor variable was found to significantly decrease the presence score. As noted earlier however, the Simulator Sickness Questionnaire was not included in the regression, which is somewhat odd. Either the importance of individual elements of the SSQ was lost in the computation of the overall SSQ score, or the SSQ is being rated differently to this more direct question. One might conjecture that because the SSQ directly asks about particular symptoms, it may be less prone to some bias that appears in the responses to this questionnaire item. Alternately, perhaps there is some subtle factor that it is missing from the SSQ. It seems that this is another avenue for further research.

The last item that appeared in the regression regards the number of hours per week that the subject spends playing computer games. This had the weakest although significant effect on the presence score of all the independent variables. The variable reduced the presence score by about 0.17 points, which is rather negligible. Given that the environments were in no way as compelling (in terms of dynamics or richness of content) as a commercial three-dimensional computer game this is not too surprising.

Regarding the apparent conflict between the resulting gaze-entropy (CER) measures and the presence-questionnaire scores, it must be kept in mind that results have to be interpreted according to the measures used. The presence questionnaire results must be regarded as reflecting the subjective, cognitive, and contemplative experience of the subjects. However, the gaze entropy (CER) measured how the subject's body reacts in a more autonomous sense. As these measure different phenomena, it is not especially surprising that their results may seem conflicting. Such differences are to be expected in many instances, and similar differences between subjective reports and behavioural measures have been found previously (Freeman et al., 2000; Slater, 2002). If all measures pointed in the same direction and were highly correlated this could indicate a very strong experience of presence. However, in any investigation of presence, an a-priori definition of presence must be provided and in this experiment, the focus was firmly upon the gaze entropy (CER) and skin conductance responses as being RAIR measures, and so we defined our operational hypotheses in terms of these.

More recently, such ‘presence questionnaires’ have been used alongside other types of measure such as behavioural and physiological measures, rather than being relied upon as the sole measure. This change has been due to difficulties in defining and communicating the concept of presence through a questionnaire format. The results of this experiment might be said to demonstrate this problem and support this trend away from the sole use of ‘presence questionnaires’. However, it certainly cannot be said that questionnaires are problematic *per se* within the field of presence research. As an example, Mania et al. (2005) used a well-defined questionnaire to probe experimental subjects concerning their memory of objects within an IVE. Asking subjects whether they remember objects is a better defined question than asking about presence directly, and if answered appropriately could indicate the construction of as-if-real memory-schema based on the IVE. This, in effect, could provide a *post-hoc* RAIR measure of presence.

It is hoped that the findings herein will be useful to presence research in the future, in particular to enable the determination of states of presence when experiencing immersive virtual environments. Having the ability to compare empirical scanpatterns to those expected in IVEs could be an important facility for presence research analysis, especially due to the nature of IVEs which is typically biased toward a visual experience.

Chapter 5

Post-hoc Analysis of Experiment I

5.1 Introduction

In the previous chapter a novel measure was developed to investigate minimal cues and the presence phenomenon using eye scanpatterns. The results thus obtained lend credibility to that method (computing the Conditional Entropy Rate of a scanpattern). In this chapter we investigate one way in which we might improve the practicability of the methodology, by considering whether eye-tracking is really necessary to use this method. This is related in our third research question, Research Question 3, "**Could our gaze-scanpattern methodology be useful, given only an approximation of the line-of-sight?**"

5.2 A Problem Encountered

In practice, the research described in this thesis lead to a reconsideration of the eye-tracking method that was being employed. This was due to difficulties experienced in executing Experiment I. Our first-hand experience in using the eye-tracking equipment demonstrated that even with years of experience at least some systems (such as ours, an Applied Science Labs 501 eye-tracking device) can be very difficult to operate. The particular problem that we encountered entails the use of an eye-tracking device set in a confined space; specifically, our device was attached by the manufacturer to a head-mounted display and head-tracking system designed to allow head movements for viewing virtual environments. The eye-tracker system is based on an infra-red (IR) light source and video camera. The device locates the user's line-of-sight from the retinal IR reflection and the (first Perkinje) corneal reflection when the IR source is shone toward the eye (see Figure 5.1). The IR source is directed to the eye from behind a half-silvered mirror, and then reflected from a beam splitter to the eye. The camera views the eye as reflected back via both the beam splitter and the mirror (Figure 5.2.) Due to the confined space in which the device is fitted, some users found the device would push against the face, and this could move the device if not secured tightly. For the majority of people, the device would require adjustment to align the camera and infra-red light source with the eye, to get it to sit right, and so that it would give the best possible result when calibrating. There were certain persons for which pupil recognition and identification of the corneal reflection were more difficult or impossible to obtain. The infra-red light source had a variable output, and the lighting conditions in the lab never changed: there were no windows of any kind. There appeared to be two main sources of the problem. First, there seemed to be differences in the pigment or material

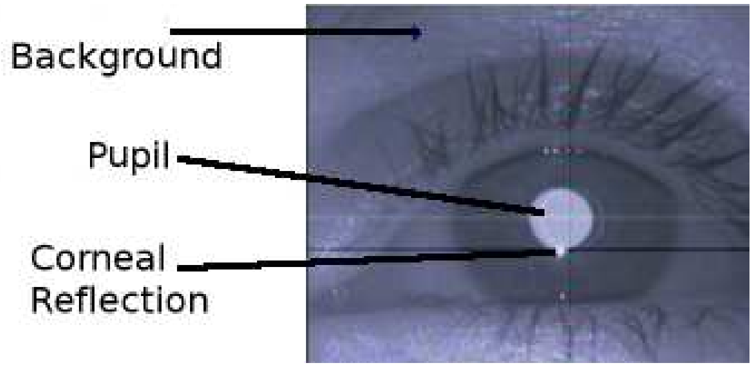


Figure 5.1: Perkinje based infra-red eye-tracking. Adapted (for clarity) from the Applied Science Labs 501 Manual.

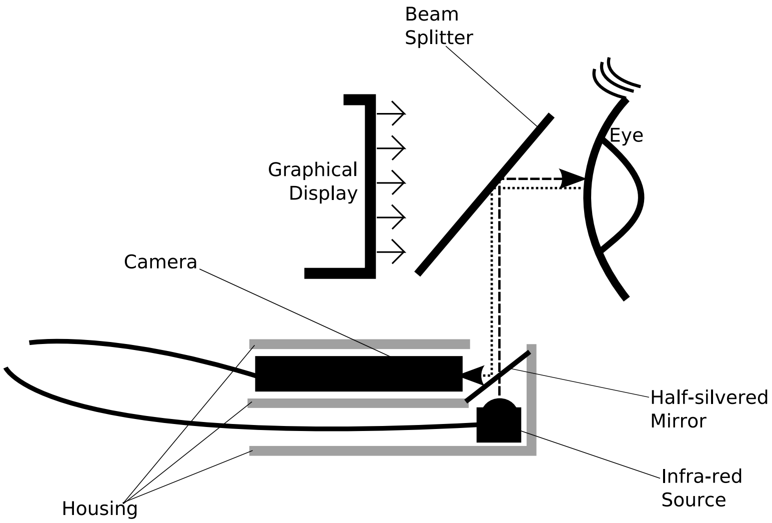


Figure 5.2: ASL 501 eyetracker configuration. (Head-Mounted Display not shown.)

surface of the eye that affected the recognition of the pupil and corneal reflection. Secondly, the facial shape and the seating of the system on subjects' heads could differ sufficiently to cause a detrimental effect to recognition, as the path between the eye-tracker and the eye was adjusted for each person. Due to the complexity of the problem it was not possible to study the problem sufficiently to determine its exact cause. Reasons for this were the fact that it was uncomfortable to adjust the equipment, and to wear the equipment for long periods of time, for instance, longer than 15 minutes or so.

The eye-tracker problems were not only the experience of this author, but were at the time acknowledged by support staff of the manufacturer who were unable to improve the system either, even though they had full physical access and were given a demonstration of the problem first-hand. One person of their staff explained, only by way of personal face-face communication, that the problem lay in the fact that the eye-tracker was in fact a customisation of a non-HMD but head mounted system, and so not originally engineered for this purpose. The same problems have been experienced elsewhere, and a paper was specifically published on this topic (Schnipke and Todd, 2000). Despite this obstacle, Experiment I was completed although with fewer participants than were originally hoped. Some other researchers have also published results with the same equipment, although the number of participants is typically fewer than the number in Experiment I, it being the second largest study in the following list of publications using this equipment: Triesch et al. (2002); Hayhoe et al. (2003); Triesch et al. (2003); Jovancevic et al. (2006); Mennie et al. (2007). It is therefore not so surprising then that eye-tracking has been largely limited to research contexts to-date, although new technology is emerging.

For the above reasons, it seemed both sensible and interesting to investigate whether we may obtain similar results to those presented in the experiment of the previous chapter by using only head-tracking data. This lead to the following analysis. Throughout this chapter, when eye-tracking data is composed with head-tracking data to determine the line-of-sight, it is referred to as EyeHead data. When the line-of-sight is estimated solely from the head-tracking data, then the data is termed HeadOnly data. The line-of-sight in the latter case is computed identically to that of the eye-tracked method but the axis of the eye is always taken to be looking directly ahead of the subject.

5.3 Transition Analysis

To attain a sense of the extent to which EyeHead and HeadOnly data differ, we first consider them in general, by looking at number of segment transitions that occur under each source of data. Figures 5.3, 5.4, and 5.5 demonstrate that the differences between the two sources do not differ greatly. This is a little surprising, as it was thought that the additional data from the eye might lead to a higher variability in the data, increasing the overall number of transitions. Whilst this is case in general, the paired samples are not significantly different from each other at the $\alpha = 0.05$ level (tested using Wilcoxon rank-sum non-parametric two-tailed test) with the exception of the RE 20-segment condition ($W = 141.5, p-value = 0.049$).

Another way in which we can compare the two data sources is by plotting both of their values over time. Plots were examined for each participant, the data in each case being processed to provide the angle of rotation around the vertical axis, over time. An exemplar (plot) is provided in Figure 5.6

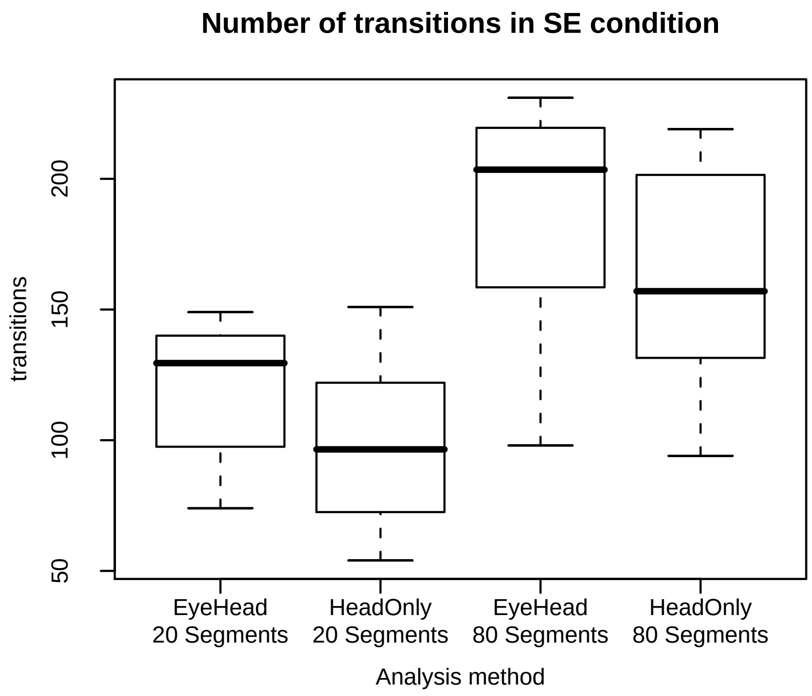


Figure 5.3: SE: Comparison of segment transitions made when utilising eye-tracker data (EyeHead), and when discarding eye-tracker data (HeadOnly).

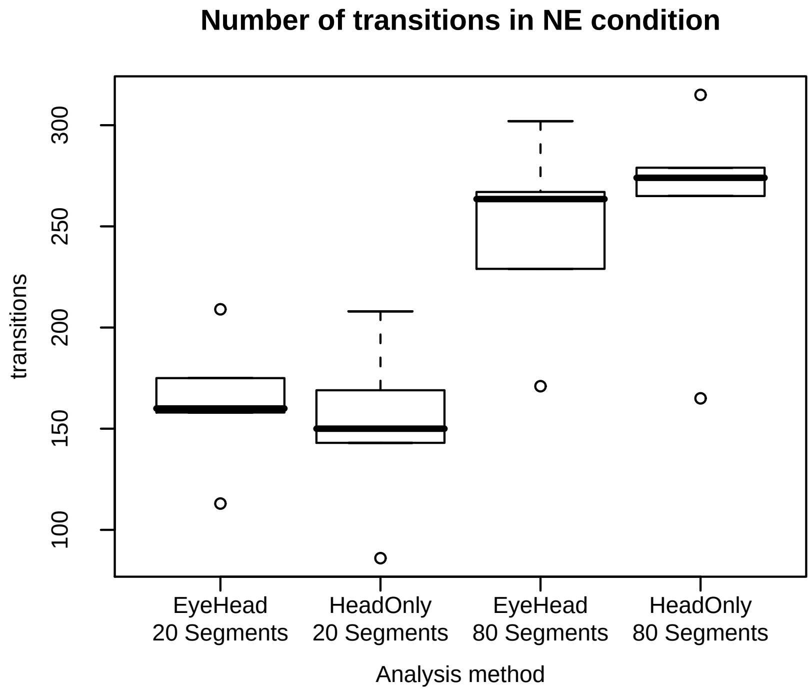


Figure 5.4: NE: Comparison of segment transitions made when utilising eye-tracker data (EyeHead), and when discarding eye-tracker data (HeadOnly).

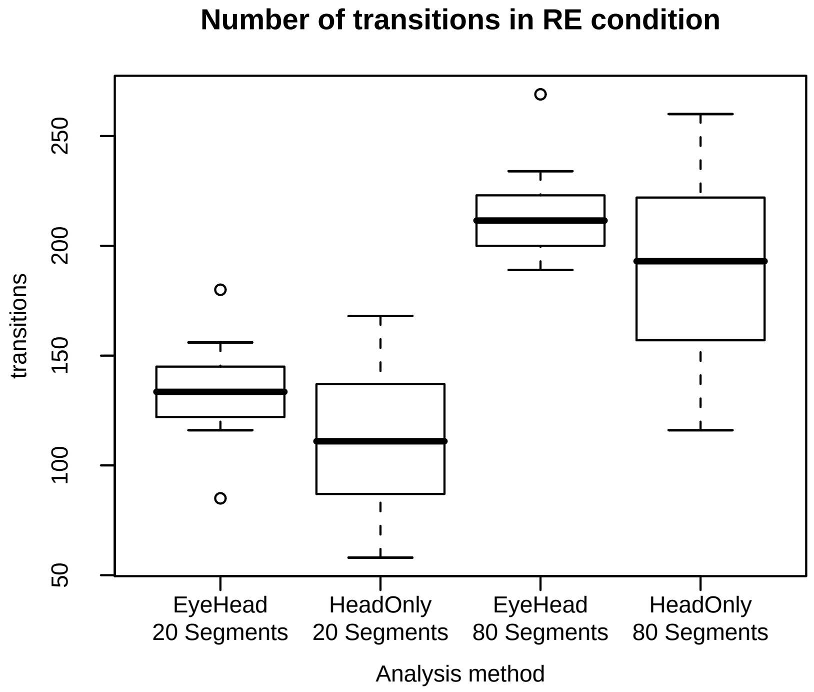


Figure 5.5: RE: Comparison of segment transitions made when utilising eye-tracker data (EyeHead), and when discarding eye-tracker data (HeadOnly).

(see also Figure 5.7.) From this figure, it is clear how the additional data available when using eye tracking modulates the data arising from the head-tracking device. However, the EyeHead data tracks the HeadOnly data rather closely.

We shall call the angular difference between the EyeHead and HeadOnly values the *divergence*. The divergence of the EyeHead line-of-sight from the HeadOnly line-of-sight (in degrees about the vertical axis) was sampled and is presented in Figure 5.8. Five thousand (5000) randomly selected point-samples of this divergence were taken from an exemplar subject, and from the figure the sample appears to be somewhat normally distributed (when applying the Shapiro-Wilks test of normality, the null hypothesis that the data is normally distributed was rejected with $\alpha = 0.05$, though not too surprising given the large n). The divergence between the two line-of-sight variables has a mean of (all following figures in this paragraph are given to 2 d.p) 2.35 degrees and standard deviation of 7.87 degrees. We can compare this divergence with the 5000 samples of the LOS angle itself, as taken from the HeadOnly data. We find this LOS angle to have a mean of -27.94 degrees, and a standard deviation of 30.55 degrees dwarfing the divergence due to the modulating effect of eye-tracking. For the EyeHead data the mean LOS angle is -25.07 degrees, and the standard deviation is 29.96 degrees. Alternatively, we may show the relationship between the EyeHead and HeadOnly data line-of-sight by correlation. By randomly selecting just twenty (20) data points from the larger (5000 element) sample, the Pearson correlation of the EyeHead and Headonly data results in a statistically significant value of $r = 0.97$ ($t = 18.13, p - value < 0.01$).

The divergence between the EyeHead and HeadOnly data is also apparent when viewing both the

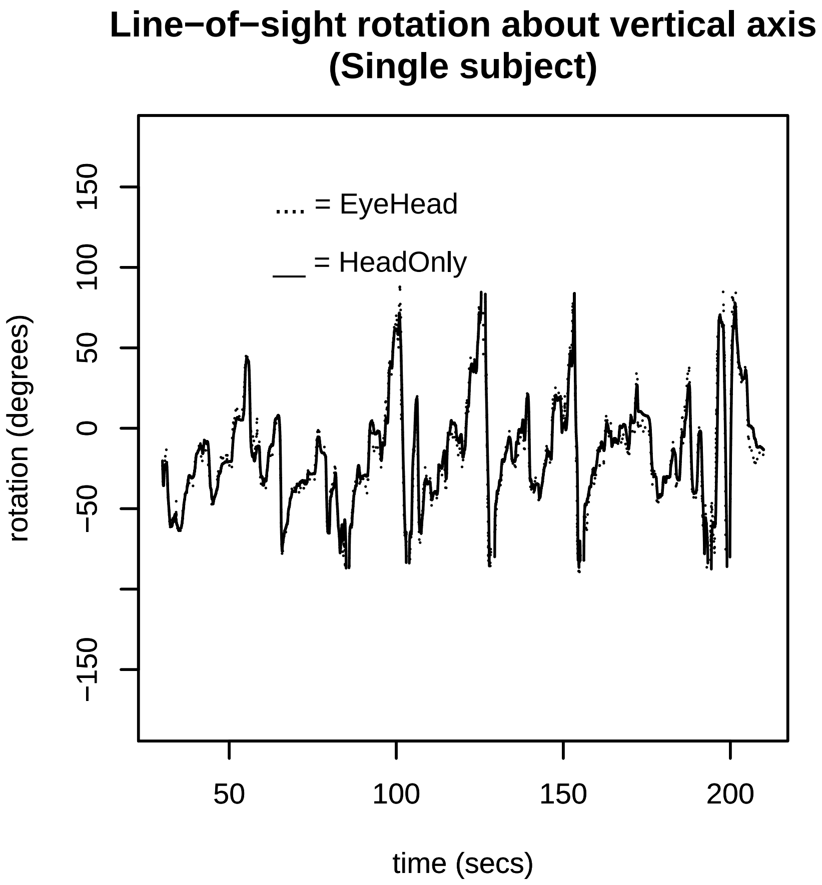


Figure 5.6: Illustrating Line-of-sight Divergence: EyeHead tracking and HeadOnly tracking plotted on the same graph.

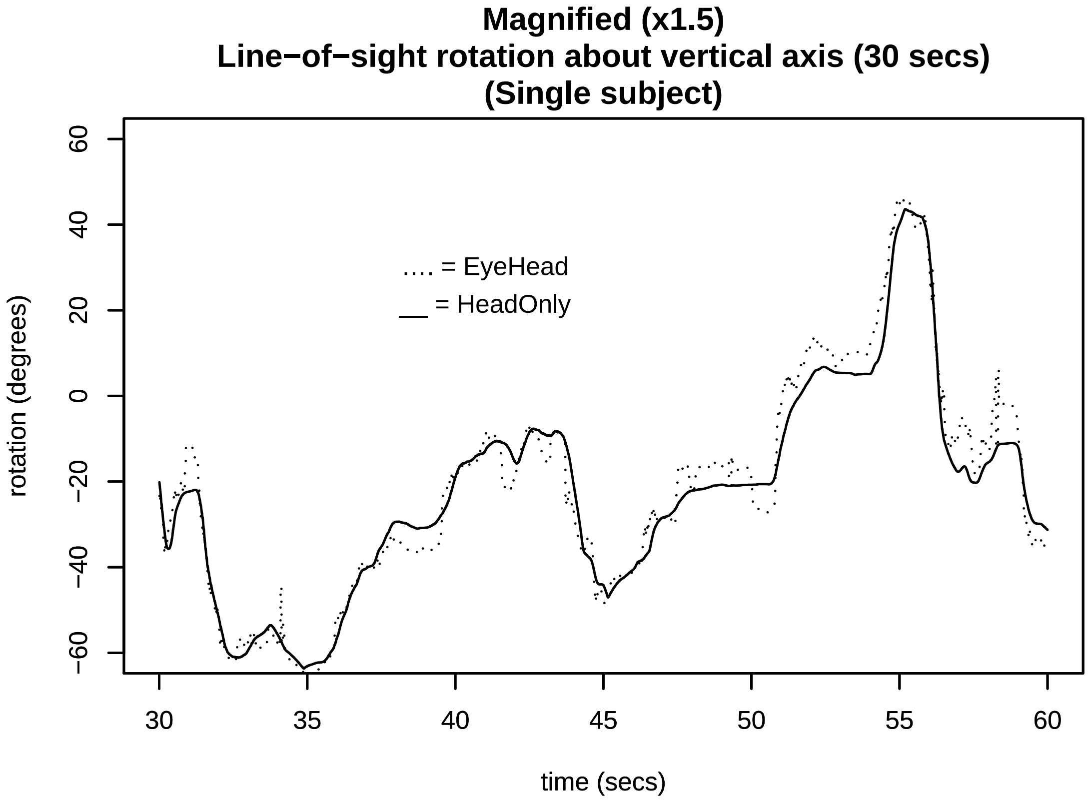


Figure 5.7: (Magnified) Illustrating Line-of-sight Divergence: EyeHead tracking and HeadOnly tracking plotted on the same graph.

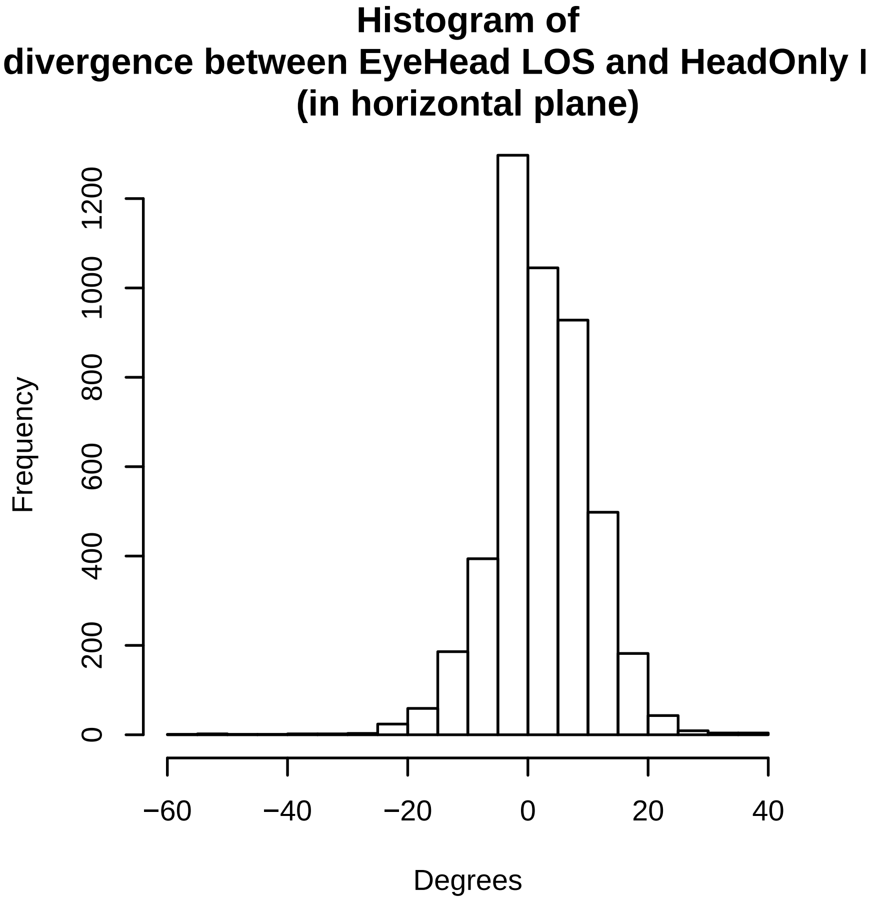


Figure 5.8: Histogram of 5000 samples of the divergence of the EyeHead line-of-sight from the Head-Only line-of-sight.

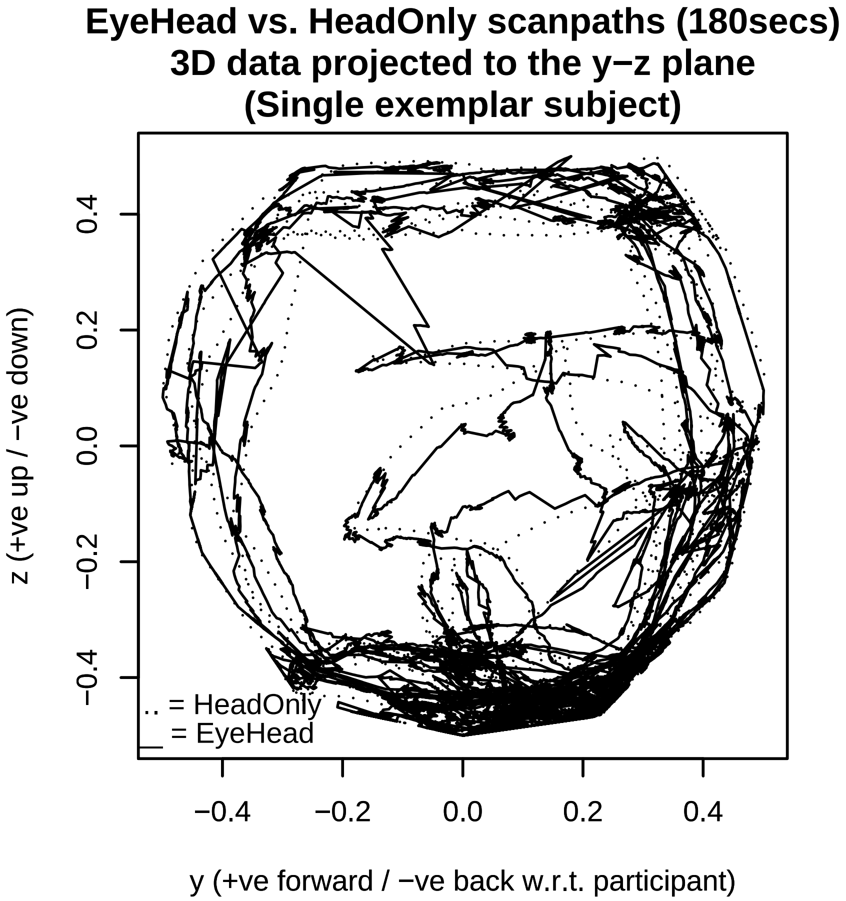


Figure 5.9: Scanpatterns with EyeHead, and HeadOnly data.

vertical and horizontal vectors are composed. A two-dimensional plot of y against z values is shown in Figure 5.9. It might be noted from this figure that the fixation points are not directly observable from the HeadOnly data, which could be a disadvantage for some applications where eye-tracking is not available. The HeadOnly data also seems quite a general approximation of the line-of-sight in this plot, as compared the previous Figures 5.6 and 5.7. On the other hand, the head-tracked data does not suffer from recurrent eye-tracking loss and artefacts, an instance of such an artefact is obvious in the upper left quadrant of the figure, where a few large movement artefacts appear. The addition of eye-tracking data provides strong indicators of the points where fixations were made, which appear looking somewhat like ‘bird nests’ dotted along the scanpattern. Because eye-tracking data modulates the head-tracked signal, we might expect the entropy of the composed eye and head data to be different to that of just the head data, either adding more variation leading to greater entropy levels, or possibly reducing entropy levels if strongly affected by top-down cognition. By calculating the CER, the rate of entropy per transition arising from our scanpattern ‘models’, and taking samples of the CER from experimental participants, we may compare the distributions of samples for both EyeHead and HeadOnly data. Forty (40) instantaneous samples were taken at random points in time for all subjects in the NE condition.

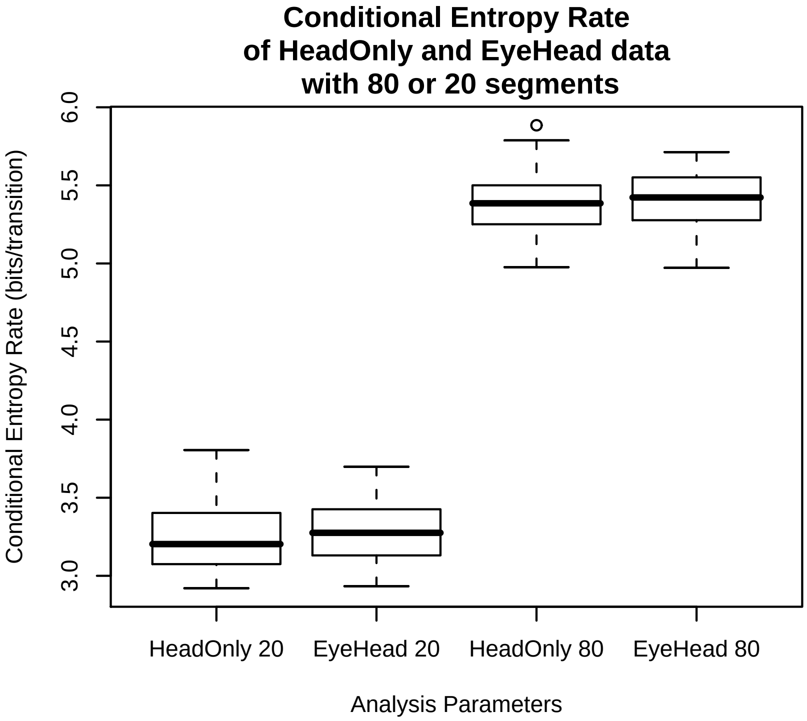


Figure 5.10: Conditional Entropy Rate for each configuration of HeadOnly, EyeHead, 80 segments, and 20 segments.

All conditions produced similar results. The result is shown in Figure 5.10 for both 20 segment and 80 segment ROIs.

There is surprisingly little difference between the data of the 80- or 20- segment pairs, and if we test for a difference (for which we use a two-sided non-parametric Wilcoxon rank-sum test), we find that the sample distributions are statistically indiscernible in both cases (80-Segments: $n = 40, W = 750, p-value < 0.636$, and 20-Segments: $n = 40, W = 866, p-value < 0.531$ respectively.) It ought to be remembered that the Wilcoxon rank-sum test should show up differences not just in the location and variation of the distributions, but in their form. It is not surprising that the entropy is generally greater in the 80 than the 20 segment case. Before data is recorded, the transition matrices are initialised with equal transition probabilities between pairs of segments as explained in Section 4.7.2.

The reason that the differences between EyeHead and HeadOnly data are clear in our 2D plot (Figure 5.9) but not in our analyses and statistics is likely due to our method of data processing. Although fixations are clearly seen on the plot when using the eye-tracker, our data processing algorithm estimates these points by assuming that a fixation has occurred once the line-of-sight has settled upon a segment for longer than a temporal threshold (as explained in Section 3.4.1). In essence, the relatively large segments act has a low-pass filter for information gleaned from the eye-tracker, meaning that EyeHead and HeadOnly data are similar. Although some information from eye-tracking is inevitably lost, the use of temporal thresholds to estimate fixation times is commonly applied in eye-tracking analyses

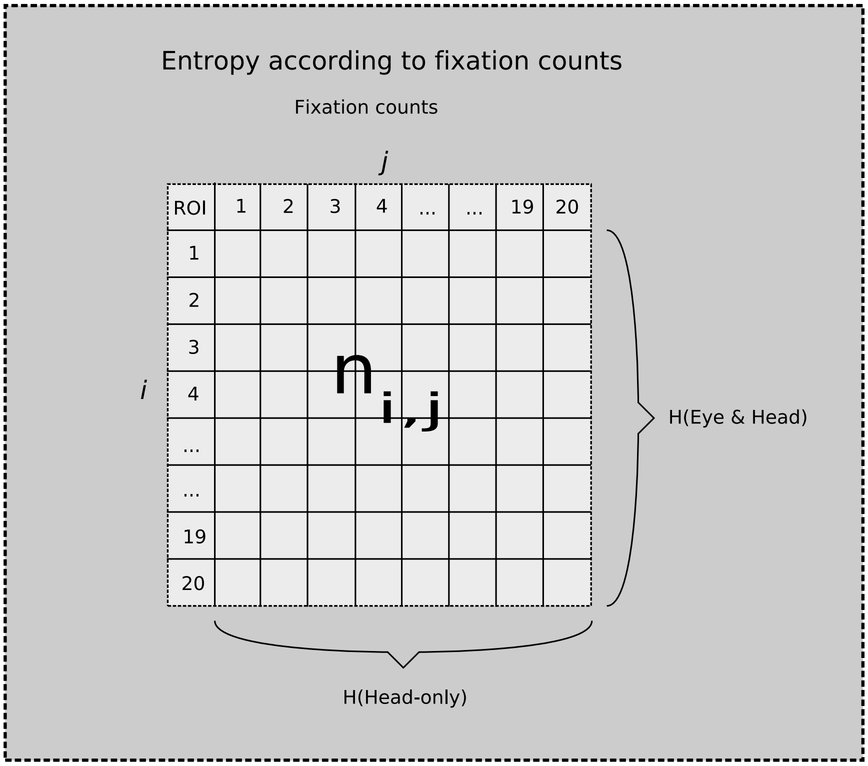


Figure 5.11: Fixation contingency matrix for EyeHead versus HeadOnly tracking illustrated with 20 segments.

(Duchowski, 2003). However, the segmentation method we use to determine regions-of-interest is less common.

5.4 Fixation Analysis

Although we have been considering transitions, some types of analyses require a greater data density than the relatively sparse transition matrix. If we consider our participants as sources of information, we may view the fixations of the two sets of data together (i.e. the EyeHead and HeadOnly data) as receivers of that information, and we can tabulate a contingency matrix (Figure 5.11). To populate the matrix, we take samples at random intervals throughout the experimental data to ensure that each is independent of the others. If we were to take each consecutive sample of our data we would find that there is dependency between the samples because, for instance, lag between the eye and head arriving at the same ROI would introduce correlations in adjacent samples. To avoid this, we wait for a (uniformly) random number of seconds between each sample, with the time between being no less than 10 seconds, and no more than 15 (i.e. 12.5+/-2.5) seconds. This procedure provides us with a total number of $n = 500$ samples to be analysed in the following section.

Statistical independence

It is natural for us to expect that the EyeHead and HeadOnly line-of-sight data are correlated, and so we may test this by applying the chi-square independence test, where the statistic is calculated in the standard way:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{5.1}$$

Where O represents the observed frequencies of the joint event whereby EyeHead data determines a fixation of ROI i and HeadOnly data determines a fixation on ROI j. The expected frequencies, denoted by E, are computed by multiplying the marginal probabilities of the EyeHead and HeadOnly events respectively, and multiplying this value by the total number of observations made. As usual, we require that each of the elements has a value of at least 5 in order that we may apply a chi-square test. To achieve this we consolidate the data by reducing the matrix size to 4x4. This requires us to re-classify the sampled fixations to these much wider bins. Each of the 80 segments is therefore randomly mapped to one of 4 new pseudo-segments, and the fixations are mapped to the new 4x4 matrix. We include, impartially, the data from all three conditions (SE, NE and RE). The matrix obtained is:

$$\begin{pmatrix} 102 & 5 & 25 & 26 \\ 6 & 42 & 7 & 12 \\ 15 & 5 & 78 & 13 \\ 22 & 8 & 22 & 112 \end{pmatrix}$$

(Number of samples, $n = 500$)

Applying the chi-square test of independence after this procedure results in a large value of χ^2 ($X^2 > 467$). The 4x4 contingency matrix, and the number of participants from which the samples were drawn determine the number of degrees-of-freedom as:

$$d.o.f = (4-1) \times (4-1) = 9 \tag{5.2}$$

These values lead us to reject the null hypothesis that the distributions are independent ($X^2 = 467$, $d.o.f = 9$, $p - value < 0.0001$). This is not too surprising as the EyeHead data is indeed derived from HeadOnly data, and one might expect this relationship to remain intact to some degree.

Information theory independence

We may also turn to information theory, which can be used to better describe the relationship between the information content of the two data sets. As illustrated in Figure 5.12, based on an explanatory conceptual diagram attributed (by Attneave, 1959) to Quastler (1953), from the matrix we can calculate the entropy estimates: $\hat{H}(EyeHead)$, $\hat{H}(HeadOnly)$, $\hat{H}(EyeHead, HeadOnly)$ (joint entropy), and $\hat{T}(EyeHead; HeadOnly)$ (the mutual entropy). The joint entropy represents the entropy of the events at the two receivers viewed as a single simultaneous event. The mutual information is the information *common* to both the EyeHead and HeadOnly data. From sample data, estimates for each of these values may be computed using the equations 5.3, 5.4, 5.5, and 5.6 (Attneave, 1959). (It should be noted that the base of log operations in the context of information theory is taken to be 2 unless stated otherwise.)

$$\hat{H}(EyeHead) = \log n - \frac{1}{n} \sum_i n_i \log n_i \tag{5.3}$$

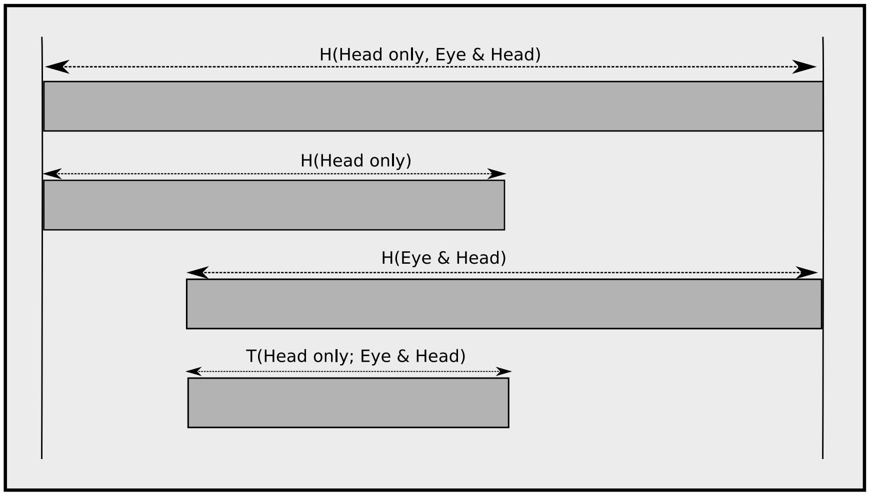


Figure 5.12: Entropy of two sources, their joint information, and mutual information.

$$\hat{H}(HeadOnly) = \log n - \frac{1}{n}\sum_{j} n_j \log n_j \tag{5.4}$$

$$\hat{H}(EyeHead, HeadOnly) = \log n - \frac{1}{n}\sum_{i,j} n_{i,j} \log n_{i,j} \tag{5.5}$$

$$\hat{T}(EyeHead; HeadOnly) = \hat{H}(EyeHead) + \hat{H}(HeadOnly) - \hat{H}(EyeHead, HeadOnly) \tag{5.6}$$

By applying the first three of these equations to our sampled data, we obtain the following values. Each value represents the expected number of bits of information per fixation on a group of segments:

$$\hat{H}(EyeHead) = 1.919 \tag{5.7}$$

$$\hat{H}(HeadOnly) = 1.923 \tag{5.8}$$

$$\hat{H}(EyeHead, HeadOnly) = 3.313 \tag{5.9}$$

We can see that EyeHead and HeadOnly fixations have a very similar level of entropy. We may now calculate the mutual information as:

$$\hat{T}(EyeHead; HeadOnly) = 0.530 \tag{5.10}$$

The computed mutual information is small in comparison to the information arriving at the receiver EyeHead (0.530 versus 1.919 respectively). So although the HeadOnly and EyeHead methods appear to be operating similarly (evidenced by their similar entropy levels at approximately 1.92 bits to 2 d.p.), the actual information content differs by a relatively large amount. This indicates that the HeadOnly line-of-sight moves between ROIs roughly to the same extent as EyeHead, but not always falling upon the same region. We might presume that the differences in where the LOS falls is because the HeadOnly data would be less accurate. It is also likely that if the sizes of the segments were smaller, then there would be a difference in entropy between the HeadOnly and EyeHead data. But these last two thoughts must be considered no more than conjecture. That the entropy values are similar means that although possibly less accurate than the EyeHead data, the HeadOnly data may well suffice for our purposes (entropy comparison) and even more so for a 20 rather than 80-segment analysis.

The computed mutual information has a chi-square distribution and the χ^2 statistic may be approximated directly from its value (Attneave, 1959) as:

$$\chi^2 \approx 2(\log_e 2)nT = 1.3863nT \tag{5.11}$$

Where n is the number of observations and T is the mutual information. The resulting value may be used to test the null hypothesis that the true mutual information value is in fact zero. The statistic is thus calculated from our sample as being:

$$\chi^2 \approx 1.3863n \times T(EyeHead; HeadOnly) = 1.3863 \times 500 \times 0.530 = 367 \tag{5.12}$$

This value encourages us to reject the null hypothesis, providing evidence that the quantity of mutual information is non-zero ($X^2 \approx 367$, $d.o.f = 9$, $p - value < 0.0001$), which confirms our result (above) when we tested for independence of the EyeHead and HeadOnly data (see the previous Section 5.4).

Having noted that the EyeHead and HeadOnly entropy values $H(EyeHead)$ and $H(HeadOnly)$ appear to be very similar, we examine the difference between them and find that in fact it is statistically indiscernible ($X^2 \approx 2.773$, $d.o.f = 9$, whereas the critical value would be 16.918 to 3 d.p). The working is shown here:

$$X^2 \approx 1.3863n \times |H(EyeOnly) - H(HeadOnly)| = 1.3863 \times 500 \times (1.923 - 1.919) = 2.773 \tag{5.13}$$

20-segments

Increasing the size of the segments and recomputing each of these measures should lead to an increase in similarity between the EyeHead and HeadOnly data. This is because on average the eye then has to diverge from the ahead-axis further in order to put it out of step with the ROI determined by the HeadOnly data. Therefore, the mutual entropy should increase relative to the entropy $H(EyeHead)$.

Sampling as before, we obtain the following 20-segment matrix:

$$\begin{pmatrix} 77 & 10 & 14 & 8 \\ 20 & 137 & 5 & 6 \\ 9 & 11 & 83 & 8 \\ 26 & 5 & 10 & 71 \end{pmatrix}$$

(Number of samples, $n = 500$)

As a routine check: testing for the Chi-Square independence of EyeHead and HeadOnly data (the null hypothesis), we reject the null hypothesis that the variables are independent having obtained $X^2 >$ 629 ($d.o.f = 9$, $p - value < 0.0001$).

The information theoretic values are as follows:

$$\hat{H}(EyeHead) = 1.969 \tag{5.14}$$

$$\hat{H}(HeadOnly) = 1.973 \tag{5.15}$$

$$\hat{H}(EyeHead, HeadOnly) = 3.161 \quad \text{(joint entropy)} \tag{5.16}$$

$$\hat{T}(EyeHead; HeadOnly) = 0.782 \quad \text{(mutual information)} \tag{5.17}$$

If we compare the mutual information under the 80 and 20 segment cases, we find that under the 80-segment analysis the mutual-information represents 27.6% of the EyeHead entropy, whereas under the 20-segment analysis it is 39.7%. This is a notable difference, and as expected the HeadOnly and EyeHead methods have more commonality when using 20 segments rather than 80.

5.5 Head-tracked data in isolation

After finding through the above analysis that EyeHead and HeadOnly data are a good deal alike as regards entropy, we now turn to put this knowledge to use. In this section we apply the tests used in the experimental analyses from Experiment I, to HeadOnly data. Our aim is to find evidence supporting whether in practice we may be able to substitute HeadOnly data for EyeHead data, thus potentially eliminating the need to use an eye-tracking device in future experiments. We use data from the same subjects that took part in the experiment of the previous chapter.

We first test whether the data of head movements alone (without eye-tracking) will reflect the minimal cues threshold as before (Experiment I). The CER averaged across participants is shown in the Figures 5.13. The data obtained from head-tracking only results in the response variable values shown in Tables 5.1 and 5.2. The same statistical analysis is carried out as is done in Section 4.9, and as before we use the same one-sided sign test. The hypothesis-test results are shown in Table 5.3.

From a subjective point of view, the graphs and the results follow the same trend as before. The CER appears to decrease under the SE condition (and the NE condition though to a much smaller extent), and CER does not appear to decrease under the RE condition. The only differences are that for 20 segments the change in CER under the SE condition is significant, and the change in CER for the NE condition no longer reaches significance, having one less binomial ‘success’. The results here are *almost* identical between the 20 and 80-segment cases. The results are also similar to those found in Chapter 4. (Experiment I).

We also plot CER against SCRs, using the clustering technique to (objectively) find a point at which to dissect our clusters. As before, we find there does indeed appear to be two distinct clusters within the SE condition and not with the NE condition (Figures 5.14 and 5.15).

Perhaps the most noticeable difference between these plots and the ones from the original experiment analyses is the amount of time over which the change appears to occur. This suggests that the use of an eye-tracker is of particular value for increasing the temporal resolution when monitoring the entropy of eye scanpatterns, though perhaps not necessary for detecting an overall difference between (or change in) states. In the 80-segment case the two clusters are split between [98,99) seconds, which is the same as our previous finding. In the 20-segment case there is overlap between the two clusters, though only of 2 seconds which seems rather negligible given that we are using low temporal resolution. From our response variable definitions, the maximum temporal resolution we could expect is 1 second.

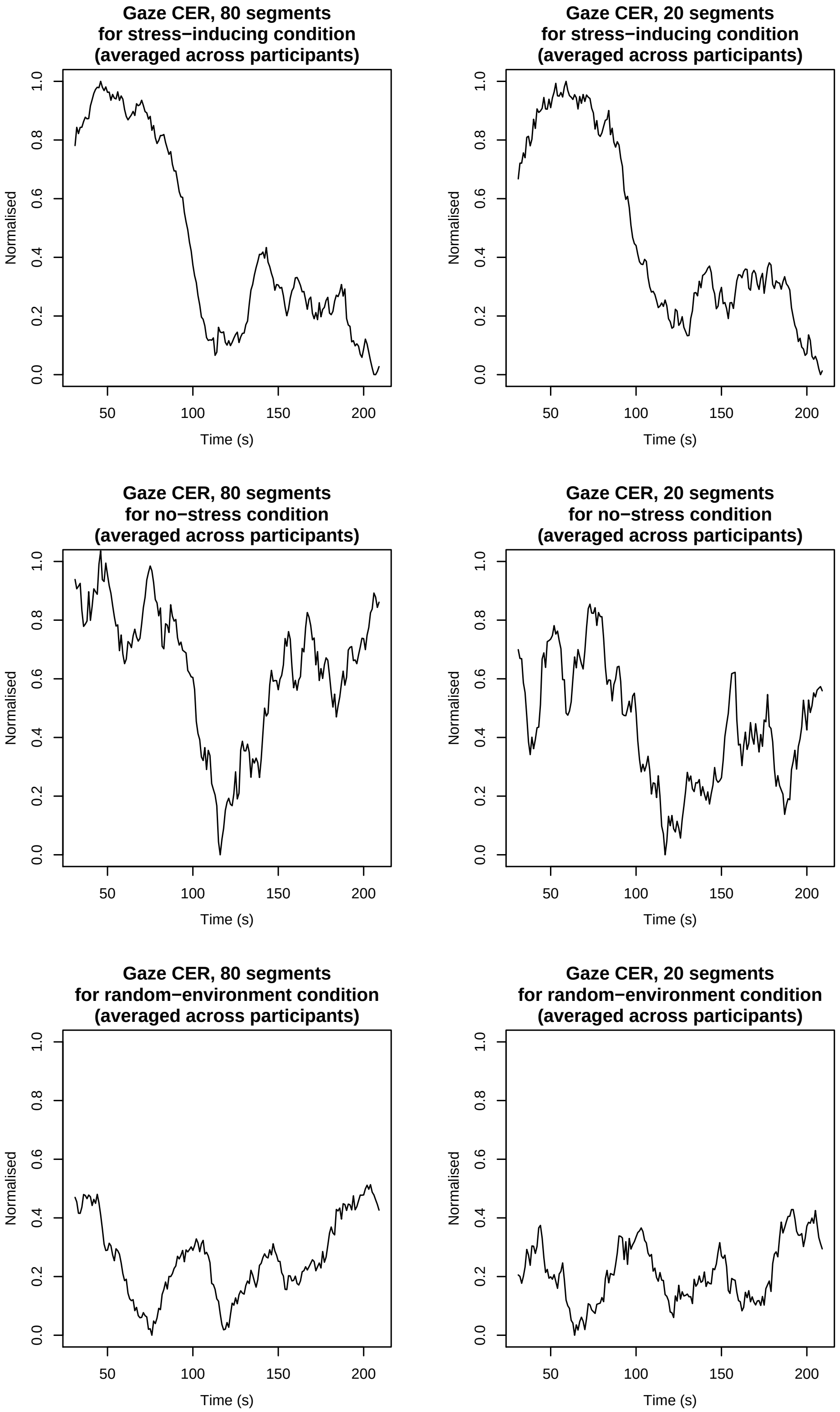


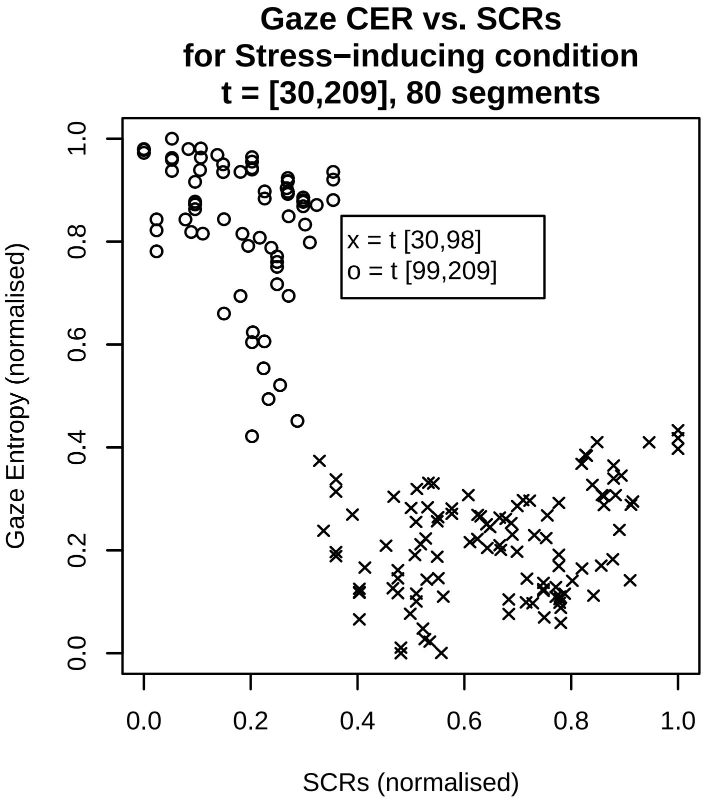
Figure 5.13: Conditional Entropy Rate for HeadOnly data, averaged across participants ($n = 28$). NB: An even clearer picture is given if one considers the larger group of participants by including those for whom eye-tracking was not calibrated ($n = 42$). This is provided in Appendix B.

Participant	Condition	Mean CER Before Midpoint	Mean CER After Midpoint	Difference
1	SE	5.991	5.703	-0.288
2	SE	6.184	5.853	-0.331
3	SE	5.983	5.764	-0.219
4	SE	5.881	5.731	-0.150
5	SE	5.646	5.392	-0.254
6	SE	5.750	5.444	-0.307
7	SE	5.598	5.731	0.133
8	SE	6.167	5.549	-0.618
9	NE	5.312	5.331	0.019
10	NE	5.367	5.087	-0.280
11	NE	5.245	5.199	-0.046
12	NE	5.345	5.329	-0.016
13	NE	5.493	5.356	-0.137
14	NE	5.680	5.668	-0.012
15	RE	5.536	5.486	-0.051
16	RE	5.767	5.734	-0.033
17	RE	5.459	5.471	0.012
18	RE	5.629	5.569	-0.060
19	RE	5.633	5.468	-0.166
20	RE	5.402	5.419	0.017
21	RE	5.590	5.653	0.063
22	RE	5.905	5.774	-0.131
23	RE	5.567	5.506	-0.061
24	RE	5.521	5.822	0.301
25	RE	5.270	5.279	0.009
26	RE	5.666	5.899	0.234
27	RE	5.571	5.622	0.051
28	RE	5.518	5.508	-0.010

Table 5.1: The effect of ‘minimal visual cues’ in HeadOnly scanpatterns, 80 segments.

Participant	Condition	Mean CER Before Midpoint	Mean CER After Midpoint	Difference
1	SE	4.149	3.806	-0.342
2	SE	3.480	3.654	0.174
3	SE	3.857	3.717	-0.140
4	SE	3.757	3.276	-0.481
5	SE	4.288	3.567	-0.721
6	SE	3.796	3.501	-0.295
7	SE	3.730	3.564	-0.166
8	SE	3.612	3.062	-0.550
9	NE	3.488	3.341	-0.147
10	NE	3.649	3.680	0.031
11	NE	3.296	3.286	-0.010
12	NE	3.256	3.126	-0.130
13	NE	3.225	3.010	-0.215
14	NE	3.310	3.120	-0.190
15	RE	3.368	3.421	0.053
16	RE	3.812	3.717	-0.096
17	RE	3.400	3.432	0.032
18	RE	3.594	3.635	0.041
19	RE	3.493	3.354	-0.140
20	RE	3.371	3.321	-0.051
21	RE	3.590	3.612	0.021
22	RE	3.871	3.718	-0.153
23	RE	3.596	3.412	-0.185
24	RE	3.429	3.665	0.236
25	RE	3.261	3.216	-0.045
26	RE	3.536	3.809	0.273
27	RE	3.492	3.528	0.036
28	RE	3.480	3.494	0.014

Table 5.2: The effect of ‘minimal visual cues’ in HeadOnly scanpatterns, 20 segments.



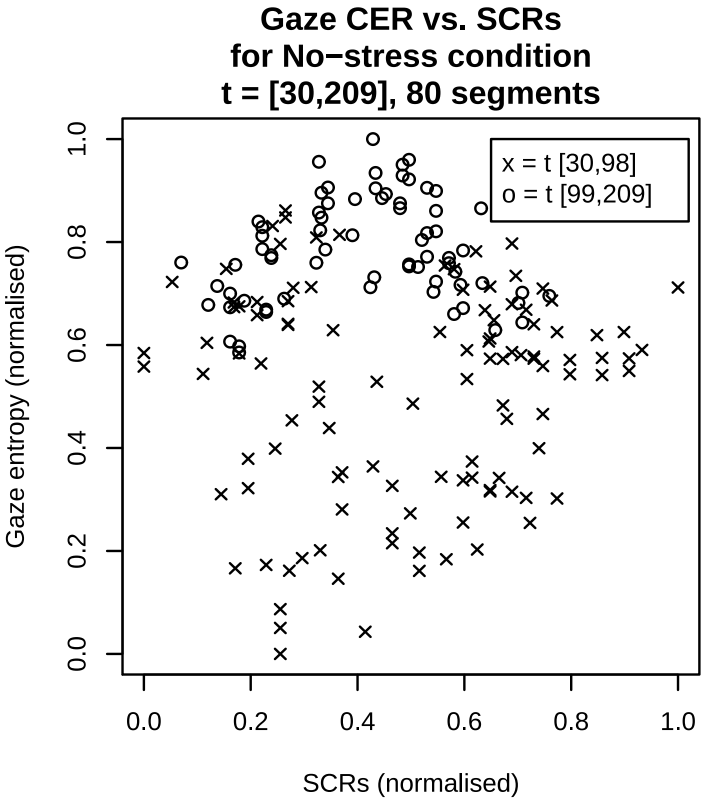
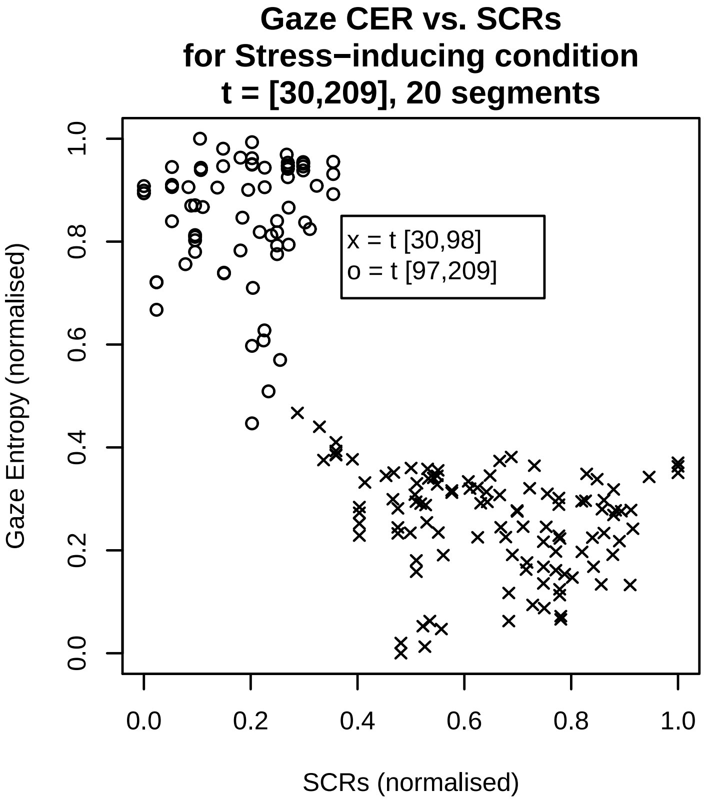


Figure 5.14: HeadOnly CER versus SCRs, SE and NE (80 Segments). SE and NE data divided into two groups using Hierarchical Clustering (complete-linkage.)



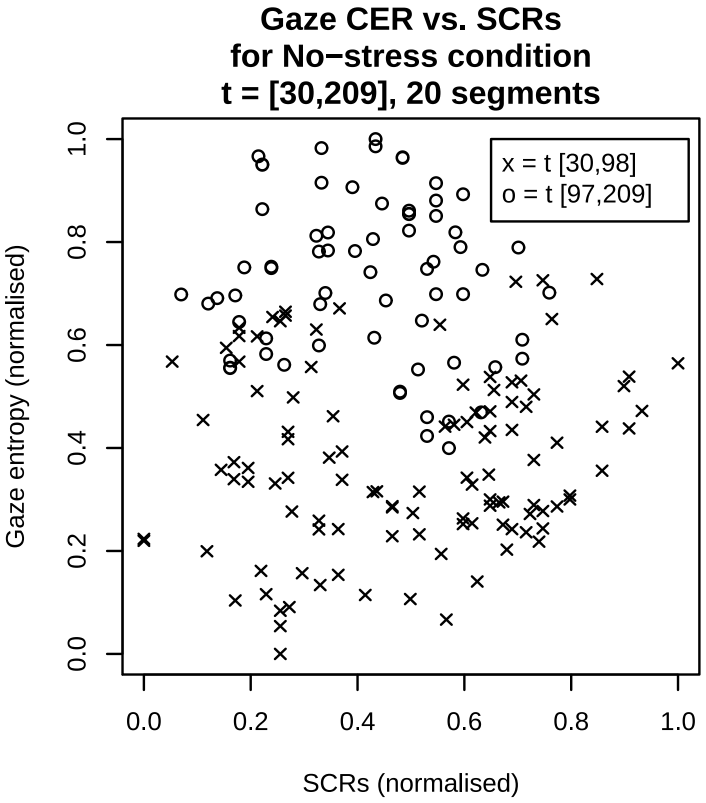


Figure 5.15: HeadOnly CER versus SCRs, SE and NE (20 Segments). SE and NE data divided into two groups using Hierarchical Clustering (complete-linkage.)

CER Measure H_a: Change in CER is -ve (i.e. CER decreases) Conditions: SE(n=8), NE(n=6), RE(n=14)		
	Segments ↓	
Condition ↓	80	20
SE	$n = 7/8, p - value <$ **0.036**	$n = 7/8, p - value <$ **0.036**
NE	$n = 5/6, p - value < 0.110$	$n = 5/6, p - value < 0.110$
RE	$n = 7/14, p - value < 0.605$	$n = 6/14, p - value < 0.788$

Significant p-values ($\alpha = 0.05$) in **bold**

Table 5.3: Change in CER, HeadOnly data, both 80 and 20 segments.

Overall the graphs look remarkably similar to the original analysis, and the difference in the lack of definition that now appears is, from a conceptual point of view, not too surprising due to the lack of eye-tracking.

5.6 Conclusions

It has been very useful to have carried out analysis of HeadOnly data, particularly as eye-tracking had proven to be an especially difficult technology to use with participants that are expected to move around while using an HMD. As well as making the rest of the equipment further uncomfortable for the wearer, the eye-tracker set up and calibration for each participant was frequently too difficult to complete, which led to an excess of potential participants for whom the equipment did not work sufficiently well.

It was not expected that the HeadOnly analyses would provide results almost on par with those from the eye-tracked *and* head-tracked subgroup, although they were not quite significant in some cases. The positive outcome of this is that we believe that HeadOnly data with a slightly larger subject pool will lead to significant results using a simpler, more robust technology, and thus widens the population of participants that qualify for future studies.

Our analysis provides evidence that the quantity of mutual information found between the EyeHead and HeadOnly data is significant, and that their level of entropy appears to be very similar — at least when the ROIs are as large as ours (using 80 segments or fewer). As such, to some extent EyeHead data may be substituted with and approximated by HeadOnly data.

The analysis in Section 5.5 suggests that for similar experiments head-tracking alone should be considered as a potentially viable and more efficient method for experimentation than by employing an eye-tracker. This is further tested empirically in Experiment II (Chapter 6).

Throughout our analyses we have seen that in both an 80 and 20-segmentation of the environment reasonably similar results arise. The only major exception found was that under a 20-segment analysis the mutual information between HeadOnly and EyeHead data increases from 27.6% to 29.7%. This is not so surprising, as using head-tracking data alone to estimate the line-of-sight with a higher resolution (i.e. 80 instead of 20 segments) is likely to lead to greater divergence[1] values.

---

[1] As explained at the start of the chapter (Section 5.3) divergence is defined as the angular distance between the actual eye line-of-sight and the line-of-sight as estimated by using head-tracker data alone.

Chapter 6

Experiment II: Comparing a Real World Environment with IVEs using Gaze Tracking

6.1 Introduction

In this chapter we investigate the fourth and fifth research questions of this thesis. The first of these is Research Question 4, "**Does the gaze scanpattern in an IVE correlate with that of the real world, when present?**" An experiment is designed to provide evidence that an IVE is perceived in a similar way to a real world environment. We are interested in this because in order to measure the success of an IVE to induce presence we require some frame of reference, and believe that the ideal reference would be the real world, as it is a concrete instance of what we may model. The real world then, may be seen as providing an *absolute* frame of reference. To investigate our first question (Research Question 4), we would like to make relative measurements of the 'success' of various IVEs' ability to provoke a real world response, in order to compare the IVEs. We attempt to derive a measure of the 'distance' of an IVE from the real world using scanpatterns. Such a device will enable us to test whether IVEs that provide representative models of a real world environment lead to identifiably shorter 'distances' between the them and the real world than inferior models.

We also consider Research Question 5, "**If visual cues are provided over and above the minimal visual cues will they affect the gaze scanpattern, and if so would this be indicative of a greater (or perhaps lesser?) presence response?**" Regarding this question, we investigate whether we can detect an increase in the realism of an IVE due to specific improvements in the rendering/presentation; in our case by applying radiosity computations. If minimal cues exist, then under the assumption that a minimal cues threshold is reached, a more realistic IVE will not necessarily elicit a scanpattern response that is any closer to that generated while viewing the real world. There is an implied caveat, because the meaning of the environment could in fact be changed by improving its realism. As an example, an IVE containing mirrors that are in one instance rendered as a basic grey surface, and in another are rendered as specular-reflective surfaces would result in substantially different environments.

6.2 Conceptual Hypotheses

In the following experiment then, we shall consider the following conceptual hypotheses:

A) “Experiences in IVE and real world environments lead to similar scanpatterns when the IVE reflects the real world.”

B) “In the *general case*, once minimal visual cues have been established, further visual cues that improve rendering realism do not significantly change the perception of being present within the immersive virtual environment.”

6.3 Assumptions

There are three major assumptions that we hold. Firstly, we assume that the scanpattern will change if the environment is changed and is thus subsequently perceived as a different environment (this is in fact what is tested.) Secondly, we assume that scanpattern transitions will not only reflect change in perception, but that such changes in the scanpattern will also be statistically discernable. Thirdly, we assume that the scanpattern characteristics (Section 3.5.1) will not be significantly inhibited by the head-mounted display equipment.

6.4 Methods and Materials

6.4.1 Participant Population

Advertisements were placed in and around the University College London campus. Persons were invited to take part in a ‘Virtual Reality Study’ that lasted approximately 20 minutes, whereby they would be paid a total of £5 if not employees of the university. The advertisement stated that participants could only take part if they had good uncorrected eyesight, precluding contact lens and glasses wearers. The intention of this restriction was to reduce any variances due to differences in eyesight, as well as to facilitate the wearing of the head-mounted-display equipment that could be physically intrusive. Willing participants were instructed to email the author in order to book an appointment. In addition, participants were approached on the university campus, shown the advertisement, and given the opportunity to take part immediately, or at some later time or day. A total of seventy-four (74) participants took part, each being randomly allocated to one of the experimental conditions.

6.4.2 Environments

The testing of the hypotheses is carried out by recording scanpatterns whilst viewing a real world environment, an IVE representation of that real world environment, a severely diminished IVE representation of that environment, and finally an improved representation of that environment having had its radiosity computed and applied to the model. We now describe these in more detail.

The real environment (Real World Environment, RWE) is a physical room containing: one chair, a desk, lamp, telephone, a computer ‘tower’ case under the desk, keyboard and monitor, skirting board, vents, room sensors for the lights and fire alarm, pipe work, a structural ‘brace’ that is part of the building and that diagonally spans one wall of the room, ceiling lights, a cabinet, two cardboard boxes, the room’s door, the IVE equipment, and a sheet of paper on the desk.

The room was recreated as a virtual environment using the 3D modelling software ‘Blender’ (version 2.3). The room and every object within it apart from the computer-keyboard were modelled by hand

(a) View from door.

(b) View towards door.

(c) View from corner of the room.

Figure 6.1: IVE_{Full}: IVE rendered using standard OpenGL lighting. (Phong computed lighting but without Radiosity computations applied to the vertices.)

using a tape-measure, with geometric accuracy to within 1cm. The model of the keyboard was sourced from a public domain database, but was the same size and style as the keyboard in the actual room. Colours were, however, approximated by eye only. Graphical textures were not used. We use the term IVE_{Full} to refer to this model. Images of IVE_{Full} are shown in Figure 6.1.

Blender was also used to produce the ‘improved’ version of IVE_{Full}, which we term IVE_{Rad}. The original model was improved by having Blender increase the number of vertices in the model and compute radiosity values, applying them to the model’s vertex colours. When this new model is rendered by the computer it results in more realistic lighting of the scene. Images of this version of the environment are shown in Figure 6.2.

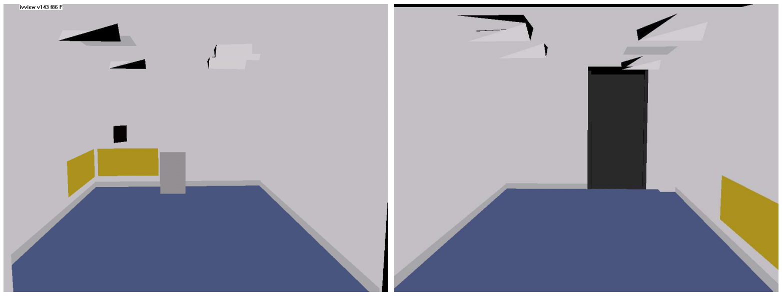
Finally, a much reduced version of IVE_{Full} was created using the software QSlim (v2.1), which reduces the level of detail in a three-dimensional model. QSlim was used to reduce the number of polygons from 15689 to just 86. We term this environment IVE_{Dim}. Almost all the polygons apart from the floor, walls, skirting board, door and walls were eventually removed. The remaining handful of polygons were remnants of the ceiling lights, desk and computer monitor but were largely unidentifiable. The room remained the same size and had the door in the same location as in the alternative experimental conditions, but the room was otherwise unidentifiable as being the same place (see Figure 6.3.)

(a) View from door.

(b) View towards door.

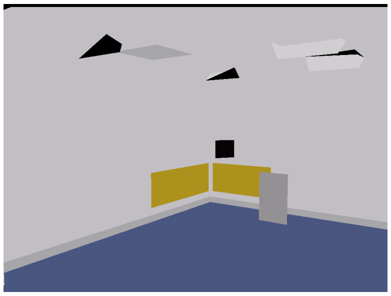
(c) View from corner of the room.

Figure 6.2: IVE_{Rad}: IVE rendered using OpenGL Phong lighting and Radiosity values computed at the vertices.



(a) View from the door.

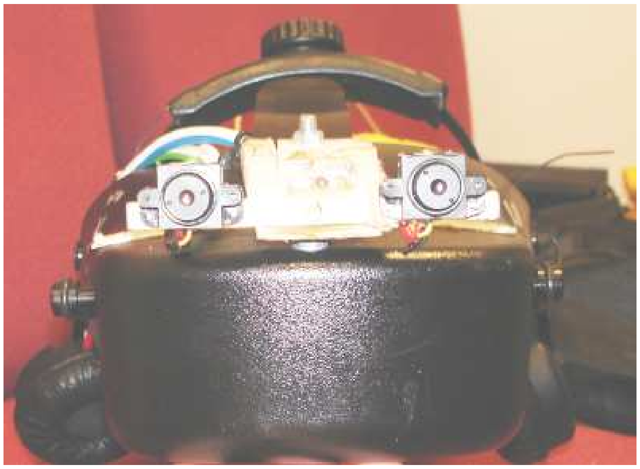
(b) View toward the door.



(c) View from the corner of the room.

Figure 6.3: IVE_{Dim}: IVE with most of the polygons removed and rendered using standard OpenGL Phong lighting without Radiosity computations. The remaining polygons include at least the walls, ceiling, door and floor.

(a) Side view



(b) Front view

Figure 6.4: Head-Mounted Display with attached cameras for 'see-through' sterescopic vision.

All environments were displayed via a HMD. The RWE was displayed on the HMD through the use of two small video cameras securely attached to the HMD, with separate video signals being used to drive the two displays, one for each eye. This results in a 'see through' HMD. The field of view (FOV) of the cameras was the same as that of the display, thus the FOV was the same across all conditions and treatments. The IVE centres of projection were translated to the position of the two video cameras to ensure the displayed images of the real environment and IVE were registered. The cameras were attached to the HMD and remained there for all participants to ensure that the additional weight and the set-up was the same under all conditions; conditions were as similar as we could make possible. Where space or software limitations have constrained us, we have used the terms Dim, $Full$, and Rad instead of IVE_{Dim}, IVE_{Full}, and IVE_{Rad} respectively.

6.4.3 Equipment

The equipment used for the experiment was as follows:

- A 1.8GHz PC was used to drive the graphical main application (using an AGP nVidia GeForce4 ti4600 graphics card.) A separate SGI O2 machine was used to record all data using VRPN software.
- A Virtual Research VR8 head mounted display (HMD) was used to display stereo images. A black piece of material was attached to the HMD to keep light out from the two displays, and to keep the participants from peeking at the real world that would otherwise be visible at the HMD extremities.
- A 6DOF Polhemus Fastrak was used to track the head position and orientation.
- Two small and lightweight (60-degree-diagonal field-of-view) colour cameras and two Sony scan-converters (model DSC-1024). The two (identical) cameras output a composite PAL signal. The PAL signal has 576 (interlaced) lines vertically, and is analog. Thus, the output of the cameras had a greater resolution than the HMD's displays.

  The HMD with the two attached cameras is shown in Figure 6.4.

6.5 Experimental Design

In the experiment, after a set period of time viewing one of the environments, there was an instantaneous switch in the participant's display, such that the participant views another environment for a second period of time.

As each treatment consists of presenting two environments, one after the other, not only do we compare between-subjects using just the first exposures to the environments, but we also compare within-subjects to investigate the effects from switching between the two environments.

The experiment has a single factor design, having four levels and a total of six treatments, as shown in Figure 6.5.

Each participant experiences one treatment only.

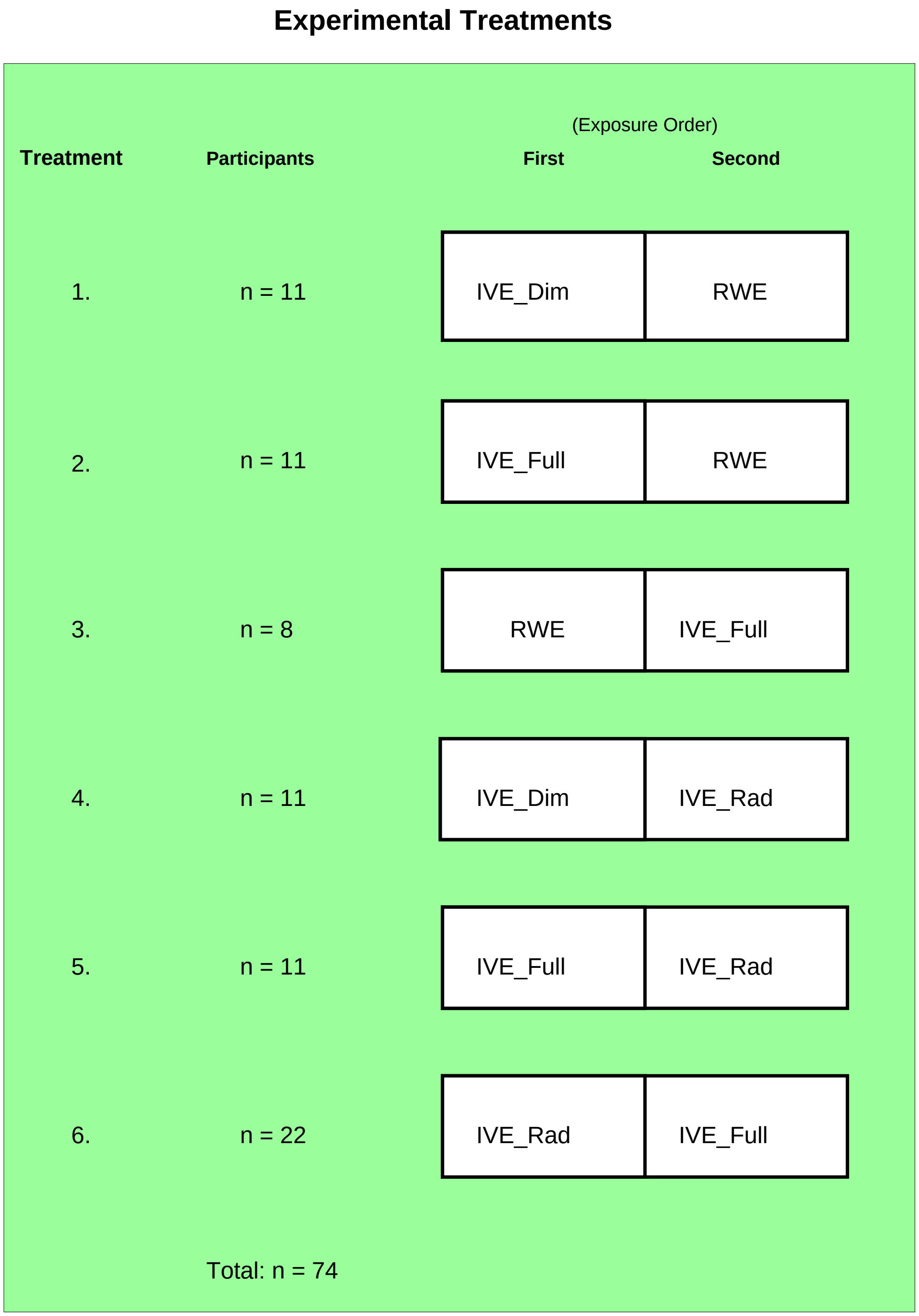
6.5.1 Experimental Treatments

The six treatments are presented in Figure 6.5. The first treatment consists of participants whom view IVE_{Dim} followed by RWE, the second consists of participants whom view IVE_{Full} followed by RWE, the third participants whom view RWE followed by IVE_{Full}, and so on. The number of experimental subjects, per treatment and in total, are shown on the diagram. Due to technical difficulties, only 8 participants could experience treatment (3).

When we analyse between-subjects, we group participants from treatments (1) and (4), and (2) and (5), giving the effective experimental design shown in Figure 6.6. When analysing within-subjects we have 11 participants per group and only use data from treatments (1), (2), (4) and (5), as shown in Figure 6.7.

6.6 Procedure

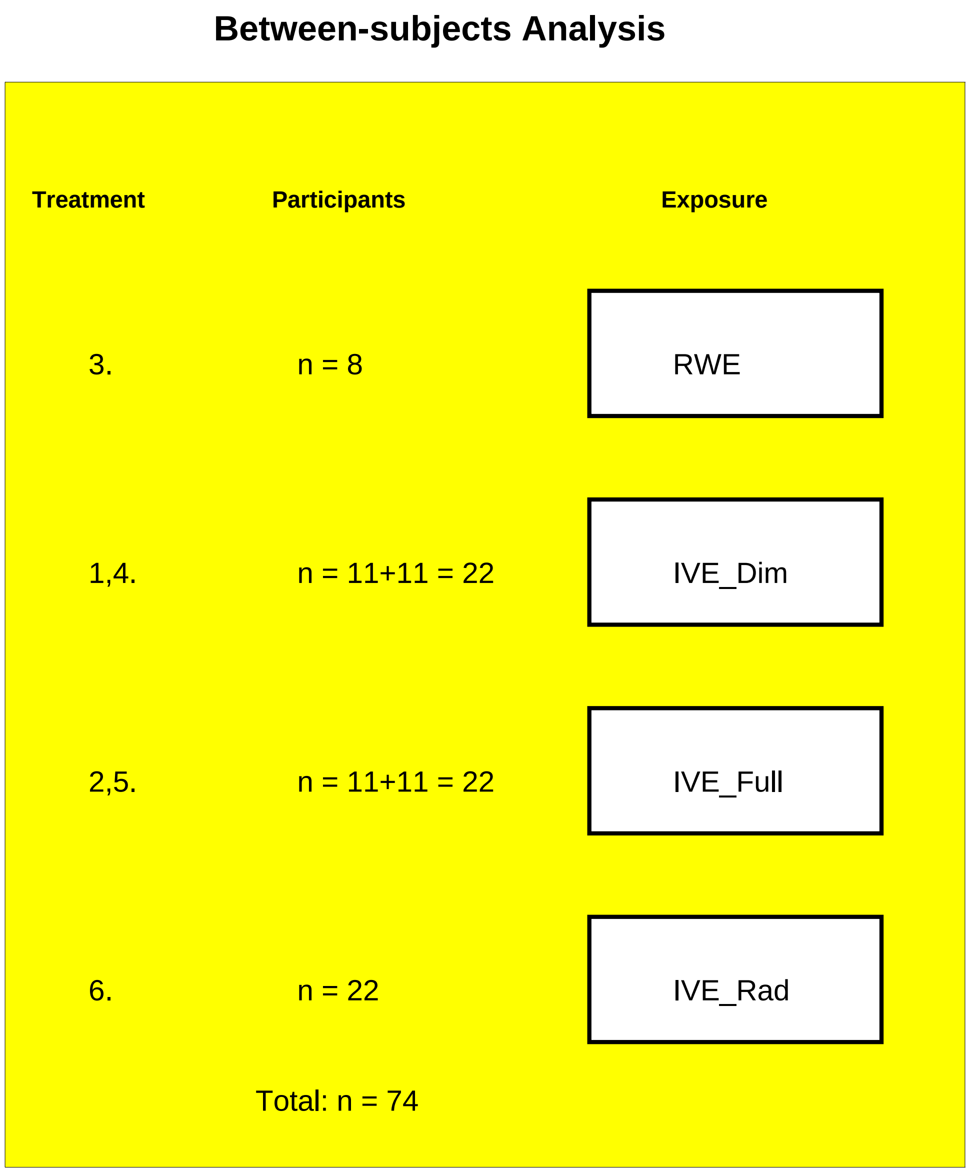
The procedure for executing the experiment was the same for all treatments:



Experimental Treatments

Treatment	Participants	(Exposure Order) First	Second
1.	n = 11	IVE_Dim	RWE
2.	n = 11	IVE_Full	RWE
3.	n = 8	RWE	IVE_Full
4.	n = 11	IVE_Dim	IVE_Rad
5.	n = 11	IVE_Full	IVE_Rad
6.	n = 22	IVE_Rad	IVE_Full
	Total: n = 74		

Figure 6.5: Experimental Treatments used to create the Between-subjects and Within-subjects groups.

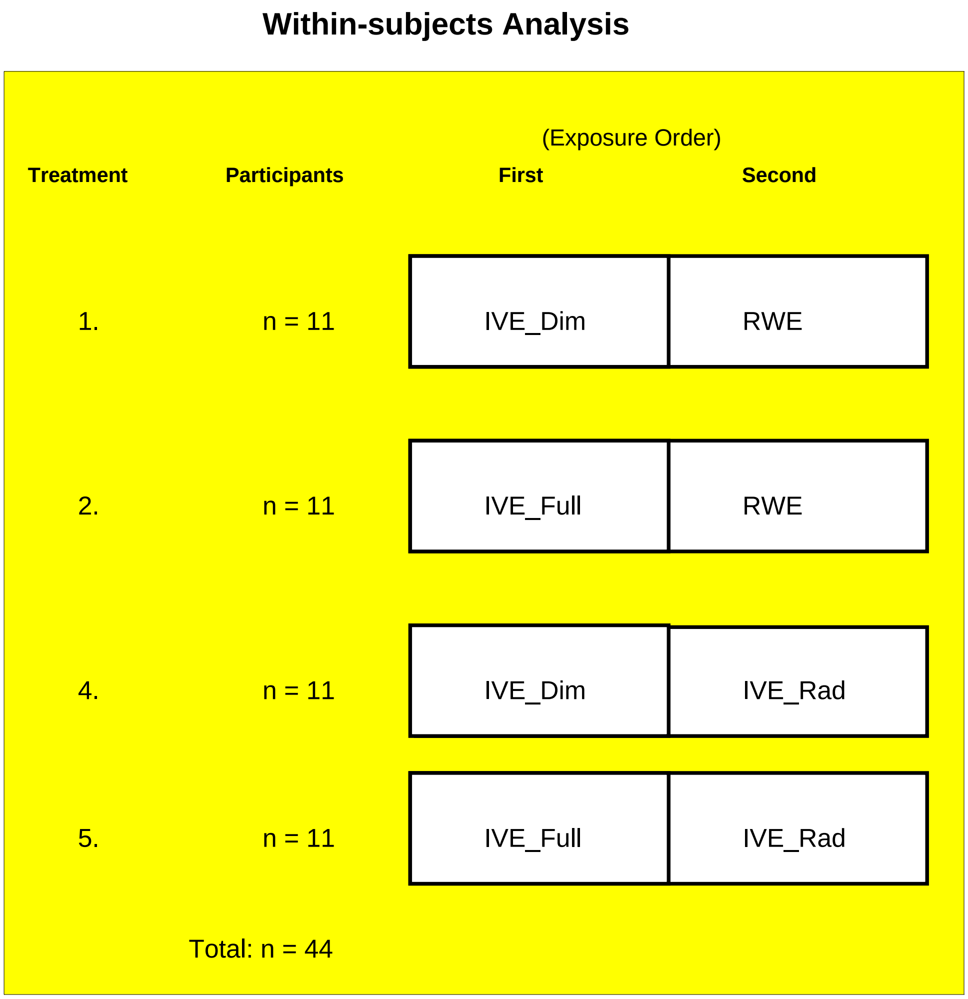


Between-subjects Analysis

Treatment	Participants	Exposure
3.	n = 8	RWE
1,4.	n = 11+11 = 22	IVE_Dim
2,5.	n = 11+11 = 22	IVE_Full
6.	n = 22	IVE_Rad

Total: n = 74

Figure 6.6: Between-subjects Analysis. Only the first exposure/condition is used.



Within-subjects Analysis

		(Exposure Order)	
Treatment	**Participants**	**First**	**Second**
1.	n = 11	IVE_Dim	RWE
2.	n = 11	IVE_Full	RWE
4.	n = 11	IVE_Dim	IVE_Rad
5.	n = 11	IVE_Full	IVE_Rad
	Total: n = 44		

Figure 6.7: Within-subjects Analysis.

The subject was first asked to complete a demographic questionnaire and a Simulator Sickness Questionnaire (Kennedy et al., 1993).

Upon completion, each participant was shown the equipment and was told that the experiment would take place in the adjacent room. It was explained that once in the room they must not move their feet from the spot where they were placed for the duration of the experiment, but that they could move their body otherwise, such as twisting the torso, neck and head to look around, or bending down.

The subject donned the HMD and was lead blind into the experiment lab to a predetermined spot from which they would observe the environments. This spot was also the position of their virtual body that was rendered in the virtual environment. The virtual body was included to aid the inducement of a sense of presence (see Section 2.3.2).

The subject was given the instructions: "I would like you to view some environments for a total of 2 minutes, and your task is to determine to what uses the environments might be put. For instance, you may think that an environment might be used as a kitchen, office or study. You can write your answers on the questionnaire afterwards."

To commence the experiment proper, a scan-converter was used to then switch the HMD to either display the view through the RWE cameras, or the application that renders the static virtual environment (environment IVE_{Rad}, IVE_{Full} or IVE_{Dim}.) The recording of measured data was begun.

Halfway through the experiment (after 1 minute), the environment was quickly switched between the two environments, according to the treatment. This switch was facilitated by the experimenter through an almost instantaneous and only just discernible change in the video signal by his pressing a button on the scan-converter when using the video cameras, or a button on the computer otherwise.

After the two minutes had elapsed, the recording of data was stopped, and the experimenter entered the room and instructed the participant to remove the HMD. The participant left the room with the experimenter, and was then asked to complete the final SSQ and presence questionnaire, after which questions were allowed to be asked of the experimenter.

6.7 Operational Hypotheses

In this section, we map the conceptual hypotheses described at the beginning of the chapter to respective operational hypotheses that may be tested experimentally. We perform the tests using the mathematical models and methods originally described in Section 3.4 to Section 3.7.

Our first hypothesis is that **"Experiences in IVE and real world environments lead to similar scanpatterns when the IVE reflects the real world."** We translate this into the operational hypothesis that **"The scanpatterns arising from the $\mathbf{IVE_{Full}}$ condition will have a greater likelihood of having arisen from the RWE model (i.e. a scanpattern model based on the RWE scanpatterns), than the $\mathbf{IVE_{Dim}}$ condition scanpatterns."**

A second way to test this is to examine the number of ROI transitions that are made subsequent to a sudden change in the environment displayed. Thus, we also test the operational hypothesis that **"The number of transitions after the switch in environments will be greater for the switch from $\mathbf{IVE_{Dim}}$ to RWE, than those in the $\mathbf{IVE_{Full}}$ to RWE condition."** Our premise is that IVE_{Dim} is less similar to RWE than IVE_{Full} is. In fact, we can further modify the definition of *transitions*, so that we only count *new transitions* that occur after the environment switch (i.e. transitions not having occurred prior to the switch), rather than counting *any* transition.

We now move on to the second conceptual hypothesis that **"In the *general case*, once minimal visual cues have been established, further visual cues that improve rendering realism do not significantly change the perception of being present within the immersive virtual environment."** To operationalise this, we assert that **"The probability that radiance condition ($\mathbf{IVE_{Rad}}$) scanpatterns could have arisen from the RWE scanpattern model will not be significantly greater than the probability that scanpatterns of the $\mathbf{IVE_{Full}}$ condition could have arisen from the RWE scanpattern model."** Likewise, we also hypothesise that the number of (new) transitions after the switch in environments will not be significantly fewer for the IVE_{Rad}, than for the IVE_{Full} condition.

6.8 Response Variables

Both objective and subjective measures are used, specifically: head-tracking based measures and questionnaire based measures respectively.

6.8.1 Scanpatterns

Scanpatterns are created by recording the direction of gaze, as it traverses pre-defined regions-of-interest. The regions are defined as in Experiment I, using a subdivided icosahedron that has a radius of 1 meter, and that is placed around the observer. We use a 20-segment icosahedron as justified by the analysis in the previous chapter (Section 5.6), whereby using 20 or 80 segments led to similar results, but with 20 segments the mutual information was greater between the EyeHead and HeadOnly data. Just as in Experiment I, a transition is recorded whenever the region intersected by the direction of gaze changes, and the new region has been intersected for longer than 267 milliseconds (see Section 4.7.2.)

We define two ways to compare scanpatterns in Section 6.7. Each implies a different analysis method:

1. Given two sets of scanpatterns, they may be compared, one against the other, by computing the probability of each set arising from a third reference set of scanpatterns. This third set is used to build a scanpattern model, and is used as the reference model. The greater the computed probabilities, the less the 'distance' between the set under consideration and the reference model.
2. Scanpatterns over the two environments of a treatment will be said to be similar if there are relatively few new ROI transitions made subsequent to the switch between the environments.

The hypotheses are tested under both methods.

6.8.2 Scanpattern distances

Our first definition of similarity above (1) requires a complex response variable: that of a measure of 'distance' between the observed transitions of a participant and a model of transitions that represent the RWE. We construct the RWE model using the data from participants that view the RWE in their first exposure (i.e. their first environment). A conditional-transition-probability matrix is constructed, starting with a uniform distribution over all possible transitions from each node. We update it using the data from each participant to provide a posterior distribution over the transitions (see Section 3.6.1.)

Now, for *any* given participant's scanpattern, such as those arising while viewing alternative environments, we may then trace each transition made through the RWE model. As we do this, we multiply out the probabilities of the individual transitions and then compute the geometric mean, effectively normalising the overall transition probability. This results in a final probability value that reflects the 'average' likelihood of the scanpattern. Not surprisingly, this value would be extremely small, so in practice we work in log space, summing the logs of each probability and finally dividing the final sum by n to obtain the log likelihood of the overall path. The log likelihood is thus used as the response variable.

6.8.3 Transition counts

Under the second definition (2) (above), we are interested in the number of transitions made in the second exposure. The response variable is therefore simply the count of these transitions. We can either count all transitions or just new transitions (i.e. transitions that had not occurred when the first environment was being viewed.) We shall analyse our data both ways, in an exploratory mode.

6.8.4 Questionnaire

The remaining response variable is constructed from responses to presence questions in our SUS questionnaire. In order to respond, the participants select a value on a Likert scale, after reading the respective Likert statement. There are five such statements:

i) There were times during the experience when the virtual environment became more real for me compared to the "real world"... (rated 'at no time'=1 to 'almost all of the time'=7)

ii) The virtual environment seems to me to be more like... (rated 'images that I saw'=1, to 'somewhere that I visited'=7)

iii) I had a stronger sense of being in... (rated 'the real world of the laboratory'=1 to 'the virtual reality'=7)

iv) I think of the virtual environment as a place in a way similar to other places that I've been today.... (rated 'not at all'=1 to 'very much so'=7)

v) During the experience I often thought that I was really standing in the lab wearing a helmet.... (rated 'most of the time I realised I was in the lab'=1 to 'never because the virtual environment overwhelmed me'=7)

The analysis of the responses is described in the results section of this chapter (Section 6.10.) The complete questionnaire is provided in Appendix A.

As in Experiment I, the same three further questions were added to the SUS questionnaire regarding the recall of objects in the environment. The first of these questions asked which objects the subject remembered, and the second, whether there were objects that they did not recognise, with the final question asking what they thought any such unrecognised objects might have been.

Variable	*(Measure type) and extents/values*
Gender	(Binary)
To what extent do you use a computer in your daily activities?	(Likert 7-point) 'Not at all' to 'very much so'
I have experienced virtual reality	(Likert 7-point) 'Never before' to 'a great deal'
My expertise with computer or video games is	(Likert 7-point) 'Complete novice' to 'Expert'
How many hours per week on the average do you spend playing computer or video games (if any)?	(Continuous Interval)
I achieved my tasks...	(Likert 7-point) 'Not very well at all' to 'very well'
While in the virtual reality I was aware of background sounds from the laboratory	(Likert 7-point) 'Not at all' to 'very much'
My status is as follows	(Nominal) 'Undergraduate Student, Masters Student, PhD student, Research Assistant/Fellow, Systems/Technical Staff, Administrative Staff, Academic Staff, Other'
How dizzy, sick or nauseous did you feel resulting from the experience, if at all?	(Likert 7-point) 'Not at all' to 'very much so'

Table 6.1: Main explanatory variables

These three questions were added for the same reason as in the previous experiment: It was thought that responses to the second question could correlate with subjective presence responses, and thus be a useful predictor variable in a logistic regression on presence scores. This is because the inclusion of unrecognisable objects in an IVE may reduce subjective presence scores, as these objects could prove to be distracting to those that notice them. Alternatively, persons more prone to overlook unrecognisable objects (consciously or subconsciously) may tend to provide increased presence scores. Finally, the remaining two questions were added because they were natural extensions of the above, and would thus be analysed in solely an explanatory mode for (potentially) future use.

6.9 Manipulated and Explanatory Variables

The only manipulated variable is the experimental treatment. The questionnaires that were administered provide several explanatory variables, many of which are shown in Table 6.1. The full questionnaire is provided in Appendix A.

Log Probability Measure $H_a : P(SP_x \vert RWE) > P(SP_{Dim} \vert RWE)$ (one-sided Wilcoxon R.S.) Number of participants: $n_{Dim} = 22, n_{Full} = 22, n_{Rad} = 22 (n_{RWE} = 8)$	
IVE_{Full} vs. IVE_{Dim}	W=142, p-value $<$ **0.010**
IVE_{Rad} vs. IVE_{Dim}	W=170, p-value $<$ **0.047**

Significant p-values ($\alpha = 0.05$) in **bold**.

Table 6.2: Log probabilities that scanpatterns arise from RWE (with IVE_{Dim} as the control.)

6.10 Results

In the following two subsections we first analyse the objective results. Specifically, we look at scanpattern probabilities, transitions, and new transitions. In the final subsection we examine the subjective (questionnaire) results.

6.10.1 Objective Data

Scanpattern Log Probability Measure

In this analysis we only consider data from an individual's first exposure to their treatment, the experimental design is as shown in Figure 6.6.

We first test whether scanpatterns arising from the IVE_{Dim} condition match scanpatterns from the RWE condition to a lesser extent than those recorded from exposure to the IVE_{Full} or IVE_{Rad} conditions. We may state this notationally as:

$$H_a : P(SP_x|RWE) > P(SP_{Dim}|RWE) \tag{6.1}$$

Where SP_x represents scanpatterns from environment x which is either IVE_{Full} or IVE_{Rad}, which we write as SP_{Full} and SP_{Rad} respectively. Likewise, SP_{Dim} represents the scanpatterns from the IVE_{Dim} condition.

If we compute the (log) probabilities of $P(SP_x|RWE)$ for each individual that experienced condition x, and those of $P(SP_{Dim}|RWE)$ for each individual, then we obtain a distribution for each environment. These distributions may then be compared using a statistical test. Although the computed probabilities belong to a ratio scale, we transform them into log probabilities, which must be viewed as belonging to an interval scale. These log probabilities are unlikely to be normally distributed, and so we use a nonparametric statistical test.

We may now quantify the evidence as either for or against the hypothesis dependent on the values of $P(SP_x|RWE)$ being greater than $P(SP_{Dim}|RWE)$. The results are presented in Table 6.2, where a Wilcoxon Rank-Sum (one-sided) nonparametric test has been applied.

As we can see, the results in Table 6.2 (and presented graphically in Figure 6.8) show that the data of the IVE_{Full} and IVE_{Rad} conditions provide for significantly better (at the $\alpha = 0.05$ level) representations of the RWE with respect to the scanpattern measure. We may therefore reject the null hypothesis in favour of our alternative hypothesis (A). The IVE_{Rad} however has a greater variance, which might reflect the increased graphical information available in the presented scene.

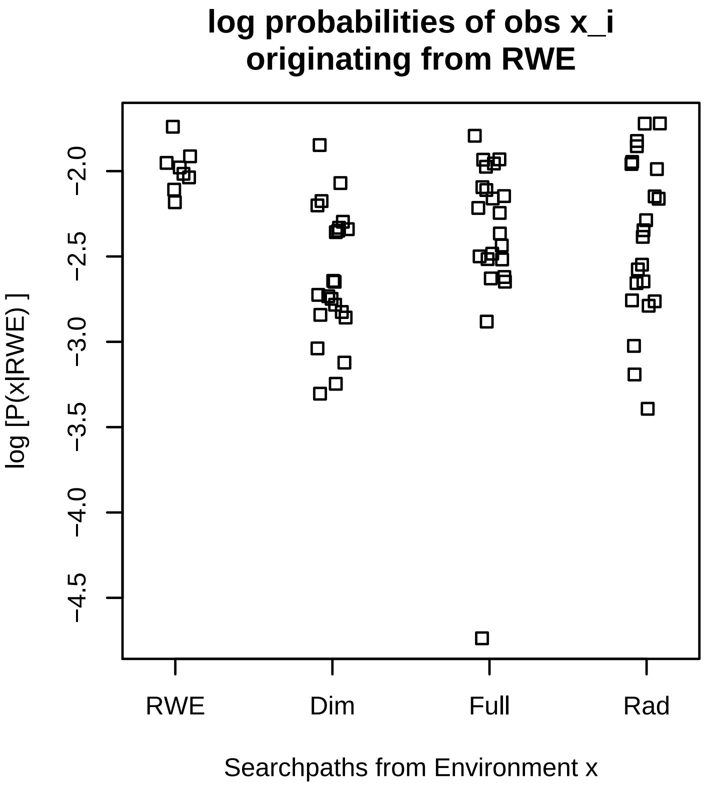


Figure 6.8: Probabilities of scanpatterns being induced by viewing the real world environment.

Log Probability Measure $H_a : P(SP_{Full}\|RWE) - P(SP_{Rad}\|RWE)$ is zero (two-sided Wilcoxon R.S.) Number of participants: $n_{Full} = 22, n_{Rad} = 22(n_{RWE} = 8)$	
IVE_{Full} vs. IVE_{Rad}	W=261, p-value <0.668

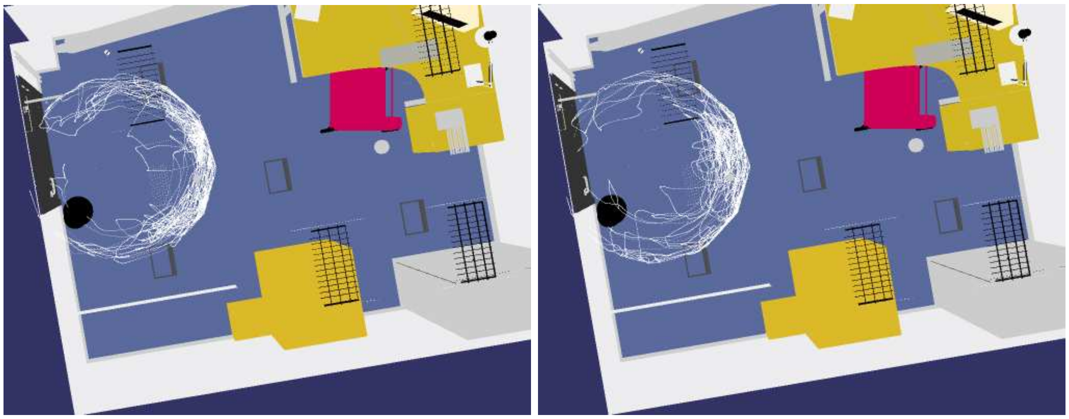
Significant p-values ($\alpha = 0.05$) in **bold**.

Table 6.3: IVE_{Full} vs IVE_{Rad}: Log probabilities that scanpatterns arise from RWE.

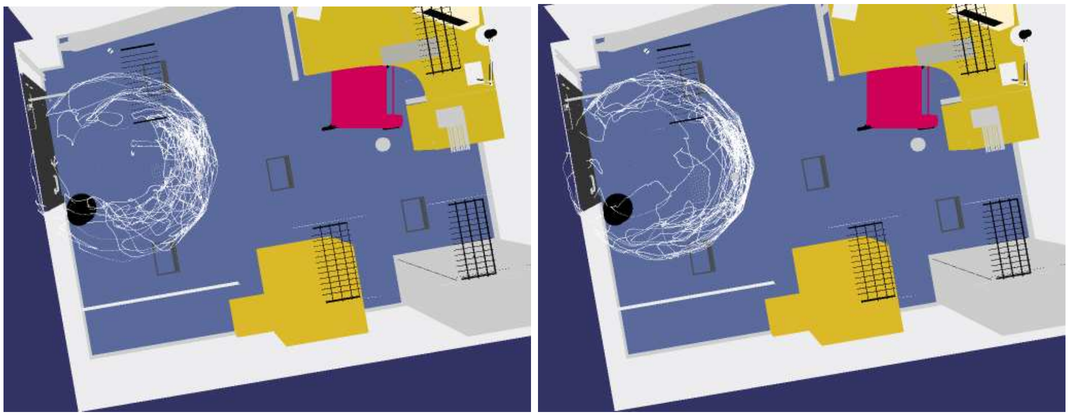
We also wish to investigate whether the IVE_{Full} and IVE_{Rad} are distinct from each other, with respect to our measure. Hence the next test is applied to determine whether there is a significant difference between the two conditions, with respect to our frame of reference, the RWE scanpatterns.

The results as provided in Table 6.3 show no discernable difference between the conditions, in accordance with our alternative hypothesis, (B).

Gaze direction is shown for eight randomly chosen experimental subjects for comparison in Figure 6.9. Each white point is the end of a vector that lays on the 20 segment icosahedron. The origin of the vector is the at the Centre of Projection, which is roughly at the center of the icosahedron (not shown). The data shown is an aggregation of the subjects' head direction vectors for a single exposure, the first 60 seconds of their experimental trial. Gaze direction for each of the four versions of the environment is



(a) RWE (b) IVE_{Dim}



(c) IVE_{Full} (d) IVE_{Rad}

Figure 6.9: Gaze Direction for eight subjects for each environment, Real World Environment, Diminished, Full, and Radiance.

shown, albeit with the same background image for reference, but there is little to differentiate between the sets of points by eye alone.

Transition counts

Our second operational hypothesis determines that we measure the differences between treatments by counting the number of transitions subsequent to the switch in environments. It suggests that when testing for statistical differences there will be a significantly greater number of subsequent transitions made when the presented environments are dissimilar. Hence, we expect that the transition counts occurring subsequent to viewing the IVE_{Dim} condition and then viewing either RWE or IVE_{Rad} will be significantly greater than transition counts from the second half of the alternate treatments. The full set of treatments are shown in Figure 6.5.

We define a function T, which simply counts the number of transitions made by each participant before and then after the switch in the displayed environment has occurred, and divides the latter value by the former. By applying T to the scanpatterns of each individual in a treatment group, we obtain a distribution for that group. These distributions are mutually tested using a one-sided Wilcoxon Rank

Transition Count Measure	
$H_a : T(SP_{Full \rightarrow RWE}) < T(SP_{Dim \rightarrow RWE})$ (one-sided Wilcoxon R.S.) Number of participants: $n_{Full \rightarrow RWE} = 11, n_{Dim \rightarrow RWE} = 11$	
$IVE_{Full} \rightarrow RWE$ vs. $IVE_{Dim} \rightarrow RWE$	W=76, p-value <0.162

Significant p-values ($\alpha = 0.05$) in **bold**.

Table 6.4: Transitions: $IVE_{Full} \rightarrow RWE$ distances, versus $IVE_{Dim} \rightarrow RWE$.

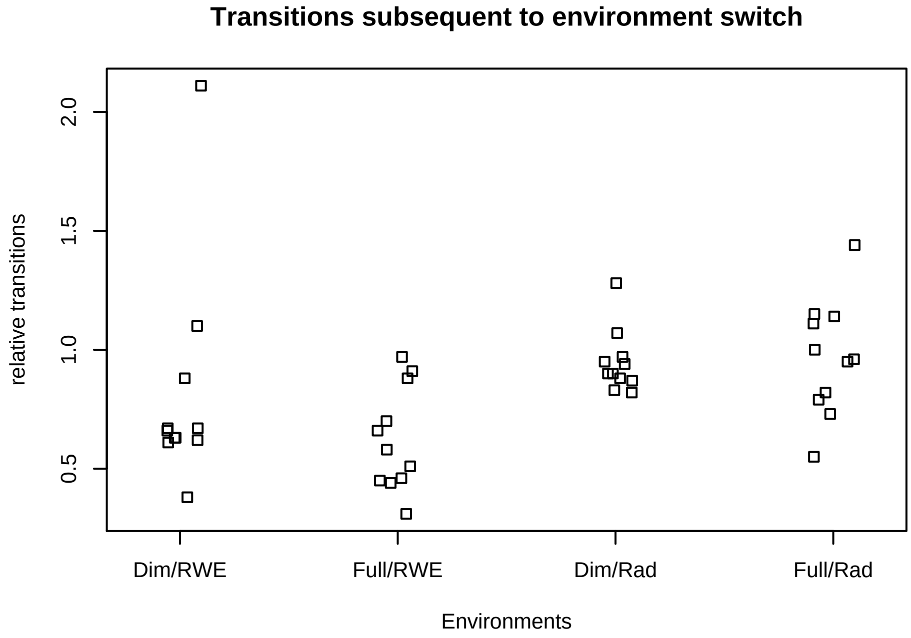


Figure 6.10: Transition Counts for each Treatment

Sum test, and the results are presented in Table 6.4 and 6.5, with the complete set of data being plotted together in Figure 6.10 allowing comparison.

From Table 6.4 we see that the relative transitions under the $IVE_{Full} \rightarrow RWE$ condition are indiscernible from the values under the $IVE_{Dim} \rightarrow RWE$ condition. This was unexpected, and does not support the rejection of the null hypothesis that the difference in transitions in the $IVE_{Full} \rightarrow RWE$ treatment is the same or greater than the difference found in the $IVE_{Dim} \rightarrow RWE$ treatment.

We may also consider whether the difference in viewing actual video (RWE) rather than a computer generated display confounds the results, and so we compare the results to subsequently viewing the IVE_{Rad} exposure. The results are shown in Table 6.5, however, we again find that there is no significant difference between the treatments. Thus, according to this measure, our operational hypotheses regarding transition counts cannot be preferred over the null hypothesis.

New Transitions

We now examine the number of *new* transitions, which occur subsequent to the switch in the environments. While T counts transitions subsequent to the switch in the environments, we now define the

Transition Count Measure $H_a : T(SP_{Full \rightarrow Rad}) < T(SP_{Dim \rightarrow Rad})$ (one-sided Wilcoxon R.S.) Number of participants: $n_{Full \rightarrow Rad} = 11, n_{Dim \rightarrow Rad} = 11$	
$IVE_{Full} \rightarrow IVE_{Rad}$ vs. $IVE_{Dim} \rightarrow IVE_{Rad}$	W=55, p-value <0.654

Significant p-values ($\alpha = 0.05$) in **bold**.

Table 6.5: Transitions: $IVE_{Full} \rightarrow IVE_{Rad}$ distances, versus $IVE_{Dim} \rightarrow IVE_{Rad}$.

New Transition Count Measure $H_a : T_N(SP_{Full \rightarrow RWE}) < T_N(SP_{Dim \rightarrow RWE})$ (one-sided Wilcoxon R.S.) Number of participants: $n_{Full \rightarrow RWE} = 11, n_{Dim \rightarrow RWE} = 11$	
$IVE_{Full} \rightarrow RWE$ vs. $IVE_{Dim} \rightarrow RWE$	W=89.5, p-value <**0.029**

Significant p-values ($\alpha = 0.05$) in **bold**.

Table 6.6: New Transitions: $IVE_{Full} \rightarrow RWE$ distances, versus $IVE_{Dim} \rightarrow RWE$.

function T_N, which is similar to T but only counts transitions that occur in latter half of the trial *that have not already occurred in the former half*. The counts are again normalised by dividing by the number of transitions made in the first half (i.e. first condition) of the treatment.

We apply exactly the same tests as previously were used, the results being presented in Tables 6.6 and 6.7, and again the complete set of data is presented for comparison across treatments in Figure 6.11.

Overall, under either of the transition count analysis methods, the null hypotheses could not be rejected in favour of the alternate hypotheses. The data as plotted in Figures 6.11 and 6.10 do however show some tendency of the treatments that begin with the IVE_{Dim} condition to have a greater mean value than the alternate treatments, as per hypothesis (A). It seems possible that the analysis methods would distinguish between treatments given larger samples.

6.10.2 Questionnaire

We next consider the responses to our questionnaires. After computing the presence scores of each participant we build a regression model using the explanatory demographic questionnaire variables, and the manipulated variables. Doing this should provide insight into the factors affecting subjective presence. As pointed out in Section 2.4.4, a regression analysis over ordinal response data cannot be immediately performed, so instead the data is transformed by counting the number of positive responses to 'presence' questions (measured on the 7-point Likert scale) to obtain a presence score that has a binomial distri-

New Transition Count Measure $H_a : T_N(SP_{Full \rightarrow Rad}) < T_N(SP_{Dim \rightarrow Rad})$ (one-sided Wilcoxon R.S.) Number of participants: $n_{Full \rightarrow Rad} = 11, n_{Dim \rightarrow Rad} = 11$	
$IVE_{Full} \rightarrow IVE_{Rad}$ vs. $IVE_{Dim} \rightarrow IVE_{Rad}$	W=64, p-value <0.422

Significant p-values ($\alpha = 0.05$) in **bold**.

Table 6.7: New Transitions: $IVE_{Full} \rightarrow IVE_{Rad}$ distances, versus $IVE_{Dim} \rightarrow IVE_{Rad}$.

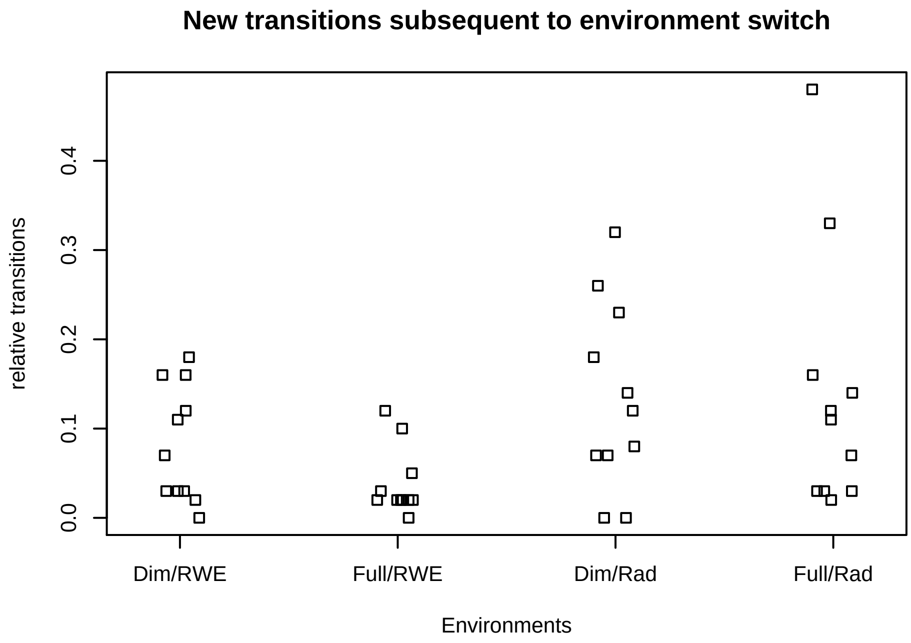


Figure 6.11: New Transition Counts for each Treatment

bution. The presence scores for each treatment are shown in Figure 6.12. We conservatively count the values of 6 or 7 as being positive responses, and use a logistic regression model.

The logistic regression was applied to the data from each of the treatments [1] (see Figure 6.5) to illuminate factors that appear to lead to differences in the attributed presence scores. Each of the explanatory and manipulated variables are used in a stepwise regression (R statistical software) to model the changes in the presence scores between the first and second exposures. The following model is produced:

$$logit(PresenceScore) = 0.753 + 0.037 GameHours - 0.130 GameExpertise - 0.110 AchievedTasks \tag{6.2}$$

That this appears to be a weak model (e.g. the manipulated variables do not appear) suggests that there are few factors that lead to major differences in presence scores. To test the goodness of fit of the model, we compute the coefficient of correlation, finding that $R^2 = 0.118$. The model is not a particularly good fit.

However, the difference in presence scores attributed to the two exposures in a treatment are significantly different across treatments. This may be shown using a Kruskal-Wallis (nonparametric ANOVA): $\chi^2 = 13.563, df = 5, p - value < 0.019$. Applying a paired Walch-Sutterworth t-test to the treatments produces the results shown in Table 6.8. These results show that subjective presence scores reflect greater differences in the treatments that include the IVE_{Dim} environment. It might be remembered that the test violates the assumption that the data is normally distributed, which leads to a less powerful test. However, the data appears convincing (see Figure 6.13), and the results shown may at least be viewed as

---

[1] Only 11 of the 22 subjects are retained from the final treatment $IVE_{Full} \rightarrow IVE_{Rad}$, to produce a balanced data set. The 11 subjects are of course selected randomly from the larger group.

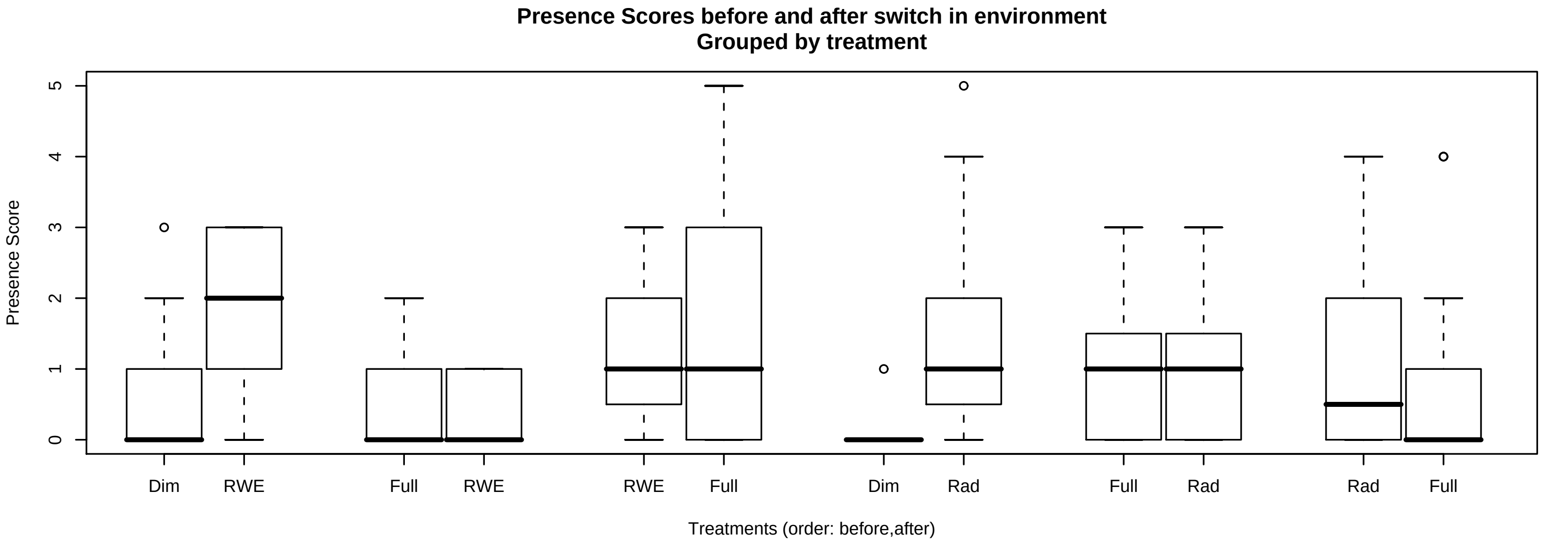


Figure 6.12: Presence scores grouped by treatment

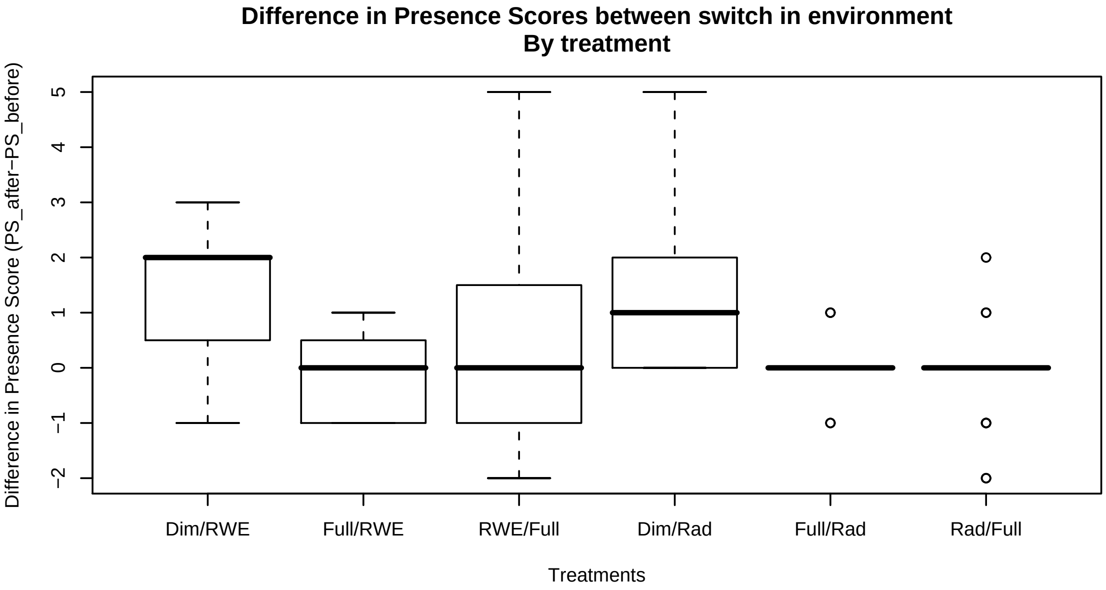


Figure 6.13: Change in Presence scores, according to treatment

Difference in the Presence Score after switch in environment		
(i) $IVE_{Dim} \rightarrow RWE$	$t = -3.545$, $df = 10$	$p - value <$ **0.006**
(ii) $IVE_{Full} \rightarrow RWE$	$t = 0.424$, $df = 7$	$p - value < 0.685$
(iii) $RWE \rightarrow IVE_{Full}$	$t = -0.713$, $df = 10$	$p - value < 0.493$
(iv) $IVE_{Dim} \rightarrow IVE_{Rad}$	$t = -3.024$, $df = 10$	$p - value <$ **0.013**
(v) $IVE_{Full} \rightarrow IVE_{Rad}$	$t = 0$, $df = 10$	$p - value < 1$
(vi) $IVE_{Rad} \rightarrow IVE_{Full}$	$t = -0.319$, $df = 10$	$p - value < 0.756$

Significant p-values ($\alpha = 0.05$) in **bold**.

Table 6.8: Tests for difference in the Presence Score after switch in environment

weak evidence for the hypothesis (A), that responses to a real environment may be similar to responses to a representative immersive virtual environment.

Finally, to test whether reported-presence scores are similar to the objective measures used, we correlate change in presence score between the first and second exposures with the changes in transitions, and new transitions. Neither of the correlations were significant.

There was no hypothesis regarding the questions of which objects subjects remembered, so the following analysis consists of purely exploratory findings. The reported (recognised) objects were collected and classified, but it must be noted that the memory-recall questions did not ask the subject to distinguish between the exposures so reporting of objects was done regarding the entire trial (i.e. over the two exposures together.) The objects reported were: computer, cupboard, box, stand, desk, door, doorstop, chair, lamp, phone, paper, structural-brace, lights, vents, smoke-alarm, and virtual-body.

Due to the experimental design and the fact that subjects were not asked to distinguish between exposures, we may only compare conditions $IVE_{Dim \rightarrow RWE}$ with $IVE_{Dim \rightarrow Rad}$. To test for differences between the two conditions we use Fisher's exact test. We find that any difference is not significant between the counts of the reported recognised-objects ($N_{Dim \rightarrow RWE} = 70$, $N_{Dim \rightarrow Rad} = 57$, two-sided, $d.o.f. = 15$, $p = 0.070$).

In $IVE_{Dim \rightarrow RWE}$ 5/11 subjects reported seeing objects that they could not recognise, and in $IVE_{Dim \rightarrow Rad}$ this number was 7/11. However, the only commonly reported unrecognised 'object' was the entirety of the IVE_{Dim} environment. The IVE_{Dim} environment was reported as an unrecognised object by one (1) subject in the $IVE_{Dim \rightarrow RWE}$ condition, and by four (4) subjects in the $IVE_{Dim \rightarrow Rad}$ condition. Just two subjects mentioned seeing the structural brace that spans the left hand side of the room as an unrecognised object, one subject in each condition.

Not only is it difficult to make any conclusions from these findings due to their exploratory nature, but there is little to be said except that subjects largely noticed and remembered the same objects regardless of whether they saw IVE_{RWE} or IVE_{Rad}.

6.11 Conclusions

The standard IVE and the Real World Environment

In this experiment, we were able to obtain evidence that an IVE model (IVE_{Full}) of our real world environment (RWE) leads to scanpatterns that are similar to those arising when viewing the RWE itself. This was achieved by comparing the similarity of these scanpatterns relative to those of participants whom viewed a severely diminished version of the IVE model (IVE_{Dim}). It was found that the scanpatterns of the diminished IVE model were significantly less similar to those of the real world environment. This was true despite the fact that the level-of-detail algorithm (QSlim) left what would appear to be salient polygons in many of the same locations that contained major elements of the IVE_{Full} environment (the monitor, desk, door, lighting units, and of course the walls and the corners of the room.)

Three measures of scanpattern similarity were tested. The first, based on computing log-probabilities was successful in finding significant differences in scanpatterns where expected. The remaining two measures were based on ROI transition counts, and failed to varying degrees. This failure may be solely attributed to the fact that there were fewer observations available when analysing transition counts (11 per treatment) versus log-probabilities (22 per treatment).

Of the transition count measures, one counted the number of new region-region transitions that occurred after the environment being presented was switched for another. This measure shows some promise for discerning between environments, given a visual inspection of the data (Figure 6.11), as the treatments that begin with the IVE_{Dim} environment appear to have a slightly higher number of transitions on the average. The variant of this measure also counted transitions that *had* occurred previously from the prior environment, but gave a yet weaker indication that it may have the potential to discern between environments.

In considering the log-probability measure, the experiment results imply that the underlying saccade and fixation theories (Buswell, 1935; Yarbus, 1967; Choi et al., 1995) are indeed useful for us to create predictive models concerning gaze over an IVE. That the IVE and RWE scanpatterns appear to be similar when the IVE contained sufficient visual cues is confirmation of an expected result, but one that is nevertheless important to have tested empirically.

The results are also evidence that the employed methodology can successfully exploit the scanpath and scanpattern theories by detecting differences in the perception of immersive environments, measuring the effectiveness of an immersive virtual environment in replicating a real world environment. Thus the method and the scanpattern models on which they are founded could be used to further investigate presence, and the related fields of computer graphics rendering and level-of-detail.

The standard IVE and the Radiosity IVE

The virtual environment model that included lighting detail based on radiance computations (IVE_{Rad}) led to scanpatterns that are indiscernible from the more basic but geometrically identical IVE model (IVE_{Full}). This result supports the minimal cues theory, because once the minimal cues threshold is exceeded, which we assume to be exceeded when viewing the IVE_{Full} model, we should not thereafter detect a significant increase in similarity between: scanpatterns elicited by the virtual environment model and those from the real world environment.

Differences in the transition counts between conditions were also examined. In general, the counts are not significantly greater for the cases in which a participant viewed the diminished environment (IVE_{Dim}) and then a complete model (IVE_{Full} or IVE_{Rad}), than when there was a switch between the complete models (IVE_{Full} and IVE_{Rad}.) This result is true for both methods used, all-transitions and new-transitions-only. However, again there remains scope for these methods to prove themselves given larger sets of observations.

The Presence Questionnaire

In the regression procedure on the questionnaire presence scores, few of the explanatory variables were left as elements in the resulting model. The fit of the model was so low as to render the model inconsequential.

The presence scores may be seen as being doubly subjective. Firstly, because subjects' opinions are collected rather than using a more objective measure. Secondly, because their responses are proffered relative to the subject's response to an alternate environment, either the first or second condition depending upon which condition their response concerns. Despite this, the presence scores as shown in Figure 6.12 appear to be surprisingly stable. The change in presence scores reflect both hypotheses (A and B), as described by the statistical test results (Table 6.8), and as made plain in the box-plotted data itself (Figure 6.13.)

Overall Conclusions

Overall, the experiment and its results suggest to us that the null hypotheses should be rejected in favour of the alternative hypotheses as set out at the start of this chapter. They also give us confidence in the underlying thesis concerning the usefulness of gaze for the investigation of presence, with the caveat that measuring the change-in-transitions would require further investigation to be shown useful.

Thus, it indeed seems that IVEs are not only designed to look and react realistically, but that we as humans do in fact respond in the same way to IVE content as we do if the displayed content is real. It is a fascinating phenomenon that more realistic renderings of an IVE may not improve this response.

Chapter 7

Thesis Conclusions

All of the investigatory work described in this thesis is designed to target the set of research questions initially set out in Chapter 1. Therefore, in this chapter we first discuss each of the research questions in the light of the work carried out, particularly the results, along with the related research of others. This accounts for Section 7.1 through 7.5. We then conclude with an overall summary of the work in Section 7.6, a critical assessment, or critique, in Section 7.7, and finally we discuss potential future work that could be carried out to extend this research in Section 7.8.

7.1 Research Question 1: Can the gaze scanpattern be used to detect RAIR in IVE users?

In Experiment I, our method was shown empirically to detect presence when defined as a behavioural response, in an environment that is designed to prompt a specific stress-inducing top-down perception. From the experiment we must remember that our approach to detecting presence, from the trends in the entropy of scanpatterns, does not provide us with the assurance that any attained state of presence occurs *for the perception that the investigator intended*. However, this should hold true in the case of our experiment because we confirm via a secondary device (EDA) that the participants experience a significant level of stress, a characteristic of our experimental condition. Put more simply, in general we may be able to detect whether participants felt present, but we cannot be certain that the experience was that which the environment's designer intended.

To obtain evidence that the participant experienced the intended environment, we must employ a secondary method, such as that applied in Experiment II. The methodology used in Experiment II allowed the comparison between environments based on scanpatterns, through the comparison of the specific transition probabilities between regions of interest. The greatest weakness of this method for empirical comparison is that a 'target' or 'reference' environment is required. Such reference environments often will not exist, or will not be available. Yet, using this methodology they are required in order to obtain data that reflects the 'ideal' behaviour expected, when a sense of presence is induced in a subject. Once a frame of reference is obtained though (either through the use of a 'real' version of the environment, or perhaps by some other method whereby an experimenter is satisfied of the immersive ability of their IVE), the method appears to be a useful one for detecting convergence or divergence from the reference environment when testing modifications to an IVE presentation. It is conjectured that the method might also be used not only for investigating content in terms of form and rendering quality, but perhaps also the presence-inducing abilities of hardware also. Further tests would however be needed to confirm the range of content between which this method can distinguish.

A secondary and unexpected result from Experiment I was due to the control condition for the scanpattern analysis, the ‘Random Environment’. It was evident from individuals’ questionnaires as well as from post-hoc comments that some notion of a sense of presence was experienced in the ‘Random Environment’; however, this was not reflected in our behavioural indicator which was based on scanpatterns. Therefore, if a wider definition of presence is used which includes meaningless environments, then this ad-hoc evidence indicates that our widely applicable presence indicator utilising scanpatterns is susceptible to Type II (‘false negative’) errors as far as the presence questionnaire is concerned. These could occur when an environment is not intended to be perceived in a top-down fashion, for instance in the case of a modern art exhibit, wherein abstract forms might have no clear relation to identifiable objects and/or environments. In hindsight, this point constrains our application domain a little further than originally anticipated, however, it is an extremely important point.

If by definition we preclude meaningless ‘environments’ from being capable of inducing presence, then such ad-hoc comments form Type I (‘false positive’) errors, and this issue has been raised before such as by Slater (2004). Here, the use of ‘presence questionnaires’ research has itself been questioned — particularly if they are used alone (see Section 4.10.) One of the solutions to this problem is provided through the application of much more objective measures, and as such it is hoped that the presence research community will find the results of this work beneficial as another available objective method for the detection of presence.

The apparent success of the method used here relies greatly upon the thesis of Gregory (1977). In particular he explains that percepts are in fact hypotheses formed using some visual stimulus, and that “[no perception] is allowed to stay when no one [hypothesis] is better than its rivals”(p221). The results of Experiment I support this theory, as it is only once the environment is perceived as meaningful that a stable perception is sustained, as reflected by the decreased entropy of the gaze scanpatterns. To use Gregory’s term, this reflects the arrival of a markedly ‘better’ hypothesis.

7.2 Research Question 2: Does a visual cues threshold for the inducement of presence exist in the context of IVEs?

One of the primary aims of Experiment I was to show the existence of a minimal cues presence threshold. The existence of minimal cues would be revealed, or at least evidence to support the theory would be obtained, if a discontinuity in a subject’s sense of presence could be demonstrated. The main difficulty in this, is that there is no known direct presence measure. However, given that presence is generally accepted as being measurable or detectable using surrogate devices, our behavioural and physiological measures allowed the design and execution of the experiment in which the onset of a ‘presence’ experience could be investigated.

Because of the variance that exists between the experiences of each participant in Experiment I and the low temporal resolution of the employed methods, it was not possible to locate a precise moment within the experiment when we could be certain that there were sufficient visual cues (i.e. minimal visual cues) to induce presence. However, when aggregating the data from subjects through the use of statistical tests and by averaging across subject data and clustering, there is evidence that presence was induced after minimal visual cues had been displayed. While the results support the existence of minimal visual cues, the act of locating them on a subject-by-subject basis is problematic given that the methodology relies upon eye/head scanpatterns, which must be collated over time.

This method puts into practice the previously untested theory that there are such things as minimal cues that are required to support the perception of one's immediate environment, as suggested by Slater (2002). By applying the process of the experiment, the required elements of a particular IVE may be found by trial and error. Such elements may be various types, such as primitives (e.g. edges, colours, or basic shapes), rendering parameters, or complex geometric items. The results also support the thesis of Stark (1995), that perception in an IVE is achievable because much of our perception of the real world is routinely based upon relatively small amounts of sensory information.

Both this research question and the previous one have lead to evidence supporting the already widely accepted understanding that fixations and saccades are idiosyncratic and repetitive over some stimulus (Buswell, 1935; Yarbus, 1967; Stark and Choi, 1996; Noton and Stark, 1971). Perhaps the most related research in this vein is found in the paper of Choi et al. (1995), where the character of saccades and fixations are described as 'searchpaths' and illustrated over a three-dimensional image.

7.3 Research Question 3: Could our gaze-scanpattern methodology be useful, given only an approximation of the line-of-sight?

When investigating the previous two research questions, the difficulties we experienced in carrying out the experiments lead us to analyse the methodology itself. Problems often arise when using eye-tracking systems, especially with those systems developed to allow a greater freedom of movement of the user. A number of these problems are documented in Schnipke and Todd (2000), wherein the authors not only lament the situation, but also go so far as to suggest that eye-tracking is not a technology ready for practical use, even in their relatively clinical research environment, saying "we believe [eye-tracking for usability testing] is still not a practical tool for most usability laboratories." (Schnipke and Todd, 2000). Although eye-tracking technology is improving and the number of published papers that include successful accounts using eye-trackers is increasing, the problems involved have not been fully overcome as yet (Morimoto and Mimica, 2005; Poole, 2005; Weigle and Banks, 2008). Such difficulties were indeed experienced by the author-experimenter, and given that the device used (the "ASL 501") was not originally intended for attaching to an HMD we experienced additional problems due to restricted physical space between the HMD and the eye, and an inadequately engineered equipment design. Not only did the difficulties lead to extremely low success rates in executing each experimental trial, but the repeated 'wear and tear' on the equipment lead to the need to dispatch the HMD for repair several times.

As with Schnipke and Todd, we had acquired years of experience using the eye-tracking equipment by this time, and had also had the national business representative for the device visit ‘in-situ’ to look at the problems, for which they had no answers. As well as the these potential problems when using an eye-tracker, there is also the fact that they expensive and require additional setting up, for their general use, and in many instances on a subject-by-subject case as well (Johansen and Hansen, 2006). Therefore once eye-tracking technology overcomes the problems associated with user-calibration, there remains good reason to evaluate alternatives to explicit eye-tracking. For us, the use of the eye-tracker raised very complex practical problems, so much so that it was decided that an investigation of scanpatterns using only head-tracking information would be performed. What needed to be established was whether head-tracking was a valid avenue by which the research could be completed. In chapter 5 we explored the relationship between eye-tracked (and head-tracked) and head-tracked only data, and found that not only was there the expected overlap in the information, but also that they had a similar level of entropy. This means that some entropy measures could be estimated from head-tracking data alone, and this was evidenced by the re-analysis of Experiment I with head-tracked data alone (i.e. without eye-tracking data). In terms of accuracy, it was found that the mutual information, information common to both Eye-Head and HeadOnly data, was significant but had a low value. It seems that the accuracy can be traded for ROI resolution; using 20 rather than 80 segments for classifying the line-of-sight increases the mutual information between the EyeHead and HeadOnly data. Due to this finding and our experience with eye-tracking, Experiment II was executed without eye-tracking equipment. The outcome of Experiment II was satisfactory and also lead to valuable results.

7.4 Research Question 4: Does the gaze scanpattern in an IVE correlate with that of the real world, when present?

From the evidence obtained from Experiment II, it appears that: virtual environments having visual cues that are modelled more closely on those from a specific real world environment will lead to similar gaze scanpattern behaviour to that when viewing the real world environment itself. This seems the case, at least up to the point of minimal cues (see Research Question 5.) This was demonstrated through the comparison of scanpatterns between a real world environment and a virtual representation of it with three levels of content and detail. We believe that this increase in similarity of behaviour is due to the the virtual environment exceeding the minimal cues threshold with respect to the real world environment it was based upon. An important point to make is that the ‘real world’, which provided our reference (i.e. experimental control) environment and thus resulted in a reference for gaze scanpattern behaviour, was actually mediated through the use of our IVE equipment and video cameras. This fact meant that the experiences were similar across both experimental conditions. This should not affect the findings of our investigation for which it is the content and level of visual realism that are of import, and so we are satisfied that we are comparing like for like. However, an unmediated experience would be the ultimate way in which to capture the ideal reference behaviour for a real world environment, and provide us with a yet more realistic measure of behaviour.

The comparison of gaze between the real and virtual worlds is somewhat limited due to the large regions of interest (i.e. segment sizes) used. It would be of value to further investigate this relationship, particularly when eye-tracking *is* viable. There is little doubt that differences will occur when restrictive head-mounted displays are used, but there appears to be little, if any, information regarding how this affects an IVE observer in terms of their scanpattern. While this does not preclude excellent research being performed between the virtual and real worlds such as that of Hayhoe et al. (2002), it would undoubtedly aid researchers to further understand the limitations that current IVE systems incur. The methodology used in Experiment II might easily be adapted for the purpose of such investigations, perhaps with the only modification necessary being a decrease in the size of the regions-of-interest.

7.5 Research Question 5: If visual cues are provided over and above the minimal visual cues will they affect the gaze scanpattern, and if so would this be indicative of a greater (or perhaps lesser?) presence response?

The motivation behind this line of enquiry is to question the rationale often assumed when desiring to produce more effective IVEs. One cannot be surprised that the layman often believes that improving visual fidelity is the simple and sole route to achieve this; contrary to this, experts have noticed that simple low fidelity renderings can be successful in inducing a sense of presence (Stark, 1995; Zimmons and Panter, 2003). This research question was devised to prompt an investigation that would demonstrate experimentally whether an improvement in fidelity might not lead to a more realistic response.

Experiment II not only included a comparison of an IVE with the Real World Environment that it represented, but it compared a ‘deficient’ IVE with the RWE as well (the experimental control). It was however a third IVE that is of relevance to this research question, wherein an ‘improved’ version of the IVE was viewed by experimental participants. Whilst the experiment showed that the ‘complete’ version of the IVE invoked behaviours that were significantly closer to the RWE than did the ‘deficient’ version of the IVE, the ‘improved’ version was indiscernible from the lesser ‘complete’ version. That is, the improved version of the IVE did not appear to invoke behaviours that were any closer to the RWE behaviours. In fact, the response-variable of the improved IVE condition had a median value that made it more distant from the RWE than the ‘complete’ IVE, and had wider variation — but this is only of secondary interest and one can only speculate as to the cause. This result suggests that the improved IVE, which to be specific had radiosity computations applied to vertices within the model, did not lead to an increased sense of presence with respect to our behavioural definition and our indicator of presence. This interesting result appears to further confirm the suspicions of some experts in the presence research field (Slater, 2002; Sanchez-Vives and Slater, 2005; Slater et al., 2009b) who believe that high-visual-fidelity is not necessary for a sense of presence. It is also evidence that supports the related investigation of Zimmons and Panter (2003) where the potential link between increased presence and high-fidelity environments has also been questioned though using other presence metrics and definitions (see also Section 2.4.3.) However, there is evidence that presence can be *increased*, such as in the work of Slater et al. (2009a), where additional cues were introduced that were not visually incremental improvements but rather of a new modality, namely those of dynamic shadows and reflections.

Having considered the research questions and addressed them in each of the investigations, we now summarise the overall contributions of this work, make a critical assessment of it, and point out some potential avenues for future work in the vein of this thesis.

7.6 Summary

One of the greatest problems to face technologists designing for presence inducing IVEs is that of content and fidelity. The assumption that improving presence by adding more content and increasing fidelity is often driven by advances in hardware technology which allows this type of exploitation, and improvements in software borne of the same objective. These are often perceived as the unique solution to inducing presence, even if it is not acknowledged as presence per se, but perhaps 'realism'[1]. The author believes that there is a danger here; to blindly progress under such assumptions without consideration of which techniques and factors are most beneficial for the inducement of presence. To determine this, evaluation is required, and thus so too are measures.

The foremost important goal of the research was to demonstrate the ability of gaze to be used as a presence-response indicator. This was thought worthy of investigation because the direction-of-gaze is part of the feedback loop when viewing an IVE, and thus is an obvious point in this perceptual loop for monitoring this process.

Indeed, through Experiments I and II the results of gaze analysis have been shown to reflect the perception of an observer's environment. Not only that, but importantly, the perception of an IVE as far as gaze is concerned appears to be indistinguishable from that which takes place when viewing real world content. This implies that gaze may be utilised as a presence-response indicator, as we can obtain evidence as to whether an IVE observer's gaze is realistic. That is, whether their gaze is directed in a realistic fashion.

In terms of presence research, as well as aiding the development of future tools for investigations, the positive results found in this work serve as further evidence to support a particular view of presence-response measures. The proponents of this view prefer an objective, behavioural/physiological-response based approach to presence research. There are many potential types of RAIR, ranging from subconscious and autonomic responses to highly cognitive responses. Although the direction of gaze may be controlled in a volitional manner, it normally operates subconsciously, being largely influenced by visual perception. This suggests that, for visually based IVEs, gaze is a good RAIR candidate for presence measurement. Gaze, and even more so, gaze direction, is just one way of measuring RAIR that ideally should be used as one element in a battery of presence indicators.

This work has also endeavoured to substantiate the concept of a presence threshold, and has produced compelling evidence that it exists for the inducement of a sense of presence in an intended envi-

---

[1] The relationship between presence and realism was discussed in Section 2.4.1.

ronment. This concept has been termed a ‘Minimal cues’ threshold, and has been mentioned indirectly or fleetingly in the presence literature, such as in Slater (2002), IJsselsteijn (2002) and Mania et al. (2005), but had not hitherto been investigated. It is because IVE presence experiments represent reality so poorly, intuitively it seems that there ought to be some threshold. In Experiment I, a link was shown by way of a common threshold in the domains of gaze and physiology, the latter already being established as a valid surrogate for measuring presence (via electro-dermal activity). This link served to strengthen the evidence for a perceptual-presence Gestalt occurring at the point at which minimal cues are presented. This novel method should afford a widely applicable presence indicator, not being reliant upon artificial stimuli devices that often have to be shoehorned into IVEs to provoke a measurable response, when they are contextually inappropriate. Its weakness however is its lack in temporal resolution, and thus the inability to use it in a situation that requires real-time detection of presence. The results of the investigation support the suggestion that there is a minimal-cues threshold, which is an important question to aid the progress of presence research as it is surely imperative that as many foundational beliefs as possible have scientific evidence supporting them. Without such assurance, future theories and investigations could fail after great expense.

Although we have already mentioned the way in which a gaze response reflects perception of an IVE, we have also *in general* demonstrated that it is possible to determine when alternative presentations of an IVE are closer to some ‘ideal’ reference environment. This could provide a valuable tool in its own right, and although imperfect (as is any model) it is expected that it would be viable in many situations. We demonstrated this specifically in Experiment II, with ‘better’ and ‘worse’ IVEs to evaluate how they compare with a reference model.

Having found that gaze in an IVE reflects gaze when viewing real world content, we also have evidence that *visual cues* are not isolated in the context of IVEs either, but that the visual cues in IVEs appear to relate to those of the real world (Experiment II). This is an important finding because it is something that we find intuitive but nevertheless worth testing from a scientific perspective. It is evidence that we are indeed right to look to the real world to inspire us when we desire to produce increasingly compelling presence inducing IVEs. Conversely however, we also found that we should be discerning when we do this, as some efforts to increase ‘realism’ may have little value in terms of improving a presence-response, as was demonstrated in Experiment II.

Having been able to achieve the results herein by utilising head tracking only is also notable. It is of value to have demonstrated that perceptually indicative data can be gleaned from this low-resolution source of information, without the aid of an eye-tracking device. This knowledge ought to be kept in mind to encourage others that it may be worth testing such robust and modest methods when eye-tracking may not be available or feasible in similar circumstances.

In sum, the author believes this work to provide several humble but valuable contributions to the field of presence research.

7.7 Critique

This work is an initial step in a new research direction, however from few experiments it is not possible to derive well-founded theories. Because of this, future investigations ought to be carried out to further substantiate the findings of this thesis, particularly to determine the constraints and limits of the results, that is, the domain of their applicability. Through the repetition of the methodology and the reproducibility of the results found in this thesis, these should also be further defined.

This work only looks at one particular way of measuring presence, and it must be remembered that other work has also been carried out. Alternative measures that have been proposed tend to have a greater set of constraints over the method proposed in this work, but it must be borne in mind that our methodology comes with its own limitations. This is particularly true of the wide temporal resolution imposed when gathering data (demonstrated by the need to use temporal 'sliding windows' in Experiment I or otherwise aggregating data over time as in Experiment II.) The capability to obtain an immediate (instantaneous) point measure of presence would surely be an attribute of an ideal presence measure, but was not attained. Because of this, we do not have the ability to detect breaks-in-presence, which is an important issue as regards presence (Section 2.4.4).

It should be noted that the environments described in this thesis are always indoor, and of a fixed size. Further work would be beneficial to discover how changing these parameters might affect these results, including how the range of content from minimalist to cluttered environments might affect gaze scanpatterns.

Whilst eye-tracking was problematic, utilising head tracking alone provided an answer for the experiments we undertook, but to what extent this methodology could be extrapolated is not (yet) known. Although eye-tracking was in our case circumnavigable, if and when it becomes a more viable tool and can provide in high resolution the exact fixation points in an IVE, then more detailed investigations could be carried out.

The method of 'saccade contingent updating' could have been utilised to streamline the experiments when using the eye-tracker. This would enable the 'in-vivo', dynamic, changes in the virtual environments in Experiment I to be made, while reducing any distraction that suddenly-appearing visual cues could incur. This ought to be seriously considered in future experiments of a similar nature, and would be one additional way that eye-tracking devices could prove their utility.

7.8 Further work

There are many ways in which this work can be extended in the future, and these broadly fall into two categories. Firstly, work should be carried out to further substantiate and research the limitations of the methods demonstrated. Secondly, some kind of standard framework for the application of the techniques could be developed, so that they may be used in practice for IVE design and for presence research. These are very general points. The remainder of this section however, points out some specific areas for further investigation.

For the experiments herein, the assumption was made that once an IVE presentation had sufficient visual cues to invoke a sense of presence, a transition would be made to a presence state. However, as pointed out in Section 2.4.4 there are such things as ‘breaks in presence’. While we have assumed that such breaks in presence would last for a negligibly short duration so that a state of presence would be generally retained, breaks in presence are an important part of presence research as well. It would therefore be of great value to attempt to discern the occurrence of these breaks in presence from gaze if this is indeed possible.

To investigate whether gaze can detect states of presence we required IVEs containing various levels of visual cues. Determining the nature of the minimally required visual cues (for inducing a state of presence in an IVE) would be a difficult task, and is beyond the scope of this work, as it would require some kind of systematic search of the myriad of conceivable and potential cues. Instead, this research relies upon detecting changes in perception that occur to establish when there are sufficient visual cues displayed, that is, when a threshold is exceeded and so evokes a presence-response. Knowledge of what the actual minimal cues might be remains largely unexplored, but would be of great value for constructing IVEs.

It is expected that this relative CER scanpattern technique could be fine-tuned; for instance the optimal size of the segments requires investigation, and ideally a clustering method could be utilised to instead determine the exact locations and boundaries of Regions of Interest. This would have more than one benefit, including obviating the need to estimate the optimal segment size(s), as well as lowering the required number of Regions of Interest which may allow the construction of a comprehensive scanpattern model of an environment.

It would also be of value to further investigate the levels of CER expected in scanpatterns, for instance to determine whether there are approximate levels of CER or characteristics of it that could provide probabilistic indicators of cognitive states. For instance, knowing whether a subject is ‘confused’ versus ‘performing a visual search’, or ‘present’ versus experiencing a ‘break in presence’ (Slater et al., 2003). Effects on the scanpattern, such as how it changes with time, task, and how it is affected by prior exposure to an environment (memory), also are of great importance and should be investigated (Henderson, 2003; Mania et al., 2005).

The relationship between presence as investigated under our adopted operational definition (as arising from an expected behavioural/psychophysiological response) and the definition implied by the questionnaire, is another avenue for future research. Whilst the objective-measure results of the experiments provided evidence supporting the hypotheses, the questionnaire was of little value in this respect, and as such the investigation of the relationship between the subjective and objective measures of such studies would be of value not just in the context of eye scanpattern presence research, but in a wider context that considers the various forms of presence itself.

The use of ‘saccade contingent updating’ could be of benefit for streamlining the methodology as used in Experiment I. Doing so could reduce the degree of distraction caused by the dynamic introduction of visual cues. An investigation of this sort would best be carried out given more robust eye-tracking technologies, particularly as high-speed sampling of the eye position is required for this purpose.

Finally, another potential avenue for future research includes microsaccades as possible indicators of behavioural/physiological presence, as they have proven to be indicative of attention, and might therefore provide an interesting source of information regarding the presence-state (Hafed and Clark, 2002). High-speed equipment is necessary to detect such subtle tremors of the eye. In a similar vein, because it is generally measurable by standard eye-tracking equipment, pupil dilation is another source of information that may provide some insight into the sense of presence in IVEs.

Appendices

Appendix A

Questionnaires

Demographic Questionnaire - Experiments I & II

General Questionnaire

1.

Gender	**Please tick against your answer**
Male	1
Female	2

2.

My status is as follows:	**Please tick against your answer**
Undergraduate student	1
Masters student	2
PhD student	3
Research Assistant/Research Fellow	4
Systems/technical Staff	5
Academic staff	6
Administrative staff	7
Other (please write in)……….	8

3.
Have you experienced “virtual reality” before?

I have experienced virtual reality	**Please tick against your answer**
1. Never before	1
2. …	2
3. …	3
4. …	4
5. …	5
6. …	6
7. A great deal	7

4.

To what extent do you use a computer in your daily activities?	**Please tick against your answer**
1. Not at all	1
2. …	2
3. …	3
4. …	4
5. …	5
6. …	6
7. Very much so	7

5.
How many hours per week on the average do you spend playing computer or video games (if any)?

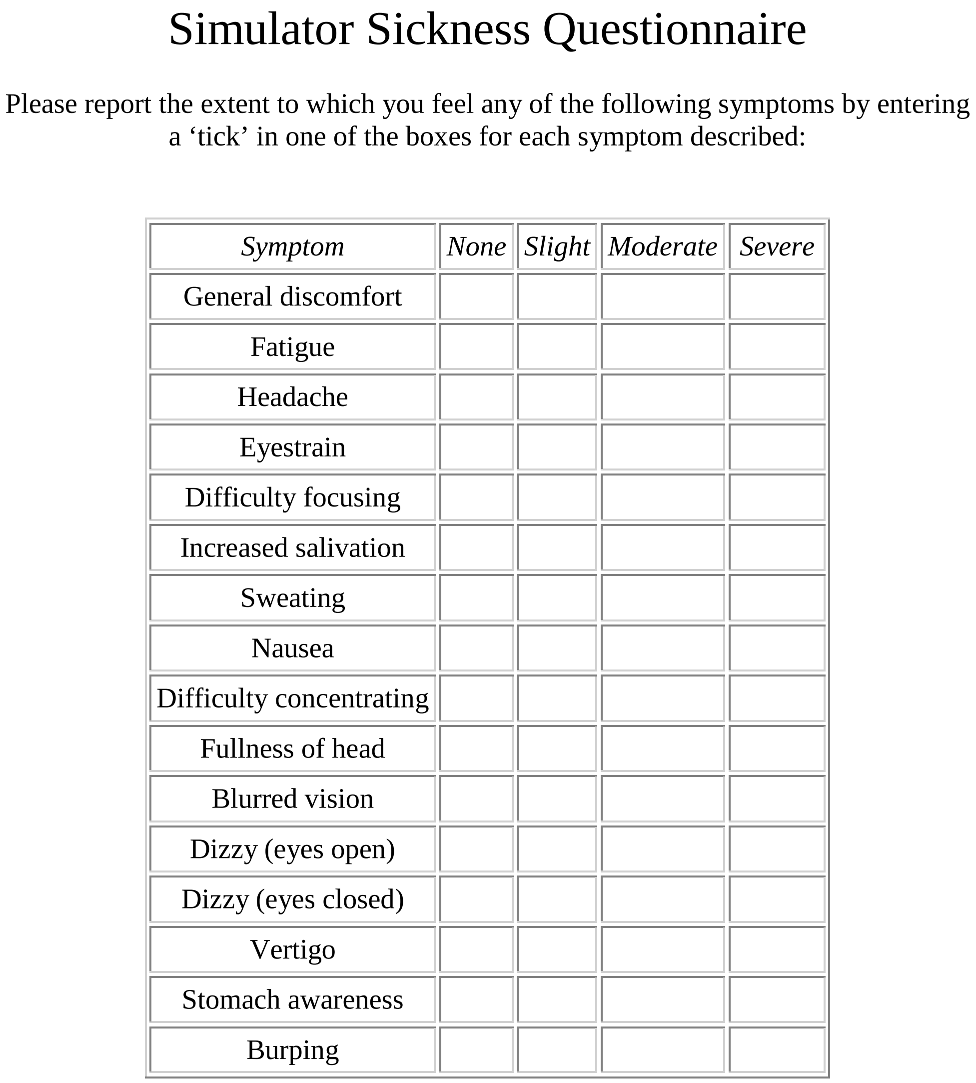
The number of hours per week I play computer or video games is:_____

6.
On the whole, how do you rate your expertise with respect to computer or video games?

My expertise with computer or video games is:	**Please tick against your answer**
1. Complete novice	1
2. …	2
3. …	3
4. …	4
5. …	5
6. …	6
7. Expert	7

Subject ID:____________

Simulator Sickness Questionnaire



Simulator Sickness Questionnaire

Please report the extent to which you feel any of the following symptoms by entering a ‘tick’ in one of the boxes for each symptom described:

Symptom	*None*	*Slight*	*Moderate*	*Severe*
General discomfort				
Fatigue				
Headache				
Eyestrain				
Difficulty focusing				
Increased salivation				
Sweating				
Nausea				
Difficulty concentrating				
Fullness of head				
Blurred vision				
Dizzy (eyes open)				
Dizzy (eyes closed)				
Vertigo				
Stomach awareness				
Burping				

Presence Questionnaire - Experiment I

Subject ID:_______

Questionnaire

The following questions relate to your experiences..

1. Please rate the extent to which you were aware of background sounds in the real laboratory in which this experience was actually taking place. Rate this on the following scale from 1 to 7 (where for example 1 means that you were hardly aware at all of the background sounds):

While in the virtual reality I was aware of background sounds from the laboratory:	**Please tick against your answer**
1. not at all ...	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. very much ...	7

2. How dizzy, sick or nauseous did you feel resulting from the experience, if at all? Please answer on the following 1 to 7 scale.

I felt sick or dizzy or nauseous during or as a result of the experience...	**Please tick against your answer**
1. not at all	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. very much so	7

3. Gender:

1. Male	1.
2. Female	2

4. To what extent were there times during the experience when the virtual environment became the "reality" for you, and you almost forgot about the "real world" of the laboratory in which the whole experience was really taking place?

There were times during the experience when the virtual environment became more real for me compared to the "real world"...	**Please tick against your answer**
1. at no time	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. almost all of the time	7

5. When you think back about your experience, do you think of the virtual environment more as *images that you saw*, or more as *somewhere that you visited* ? Please answer on the following 1 to 7 scale.

The virtual environment seems to me to be more like...	**Please tick against your answer**
1. images that I saw	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. somewhere that I visited	7

6. Have you experienced "virtual reality" before?

I have experience virtual reality	**Please tick against your answer**
1. never before	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. a great deal	7

7. During the time of the experience, which was strongest on the whole, your sense of being in the virtual reality, or of being in the real world of the laboratory?

I had a stronger sense of being in...	**Please tick against your answer**
1. the real world of the laboratory	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. the virtual reality	7

8. Overall, how well do you think that you achieved your tasks?

I achieved my tasks...	**Please tick against your answer**
1. not very well at all	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. very well	7

9. Consider your memory of being in the virtual environment. How similar in terms of the *structure of the memory* is this to the structure of the memory of other *places* you have been today? By ‘structure of the memory’ consider things like the extent to which you have a visual memory of the environment, whether that memory is in colour, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such *structural* elements.

I think of the virtual environment as a place in a way similar to other places that I've been today....	**Please tick against your answer**
1. not at all	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. very much so	7

10. To what extent do you use a computer in your daily activities?

I use a computer...	**Please tick against your answer**
1. not at all	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. very much so	7

11. During the time of the experience, did you often think to yourself that you were actually just standing in an office wearing a helmet or did the virtual environment overwhelm you?

During the experience I often thought that I was really standing in the lab wearing a helmet....	**Please tick against your answer**
1. most of the time I realised I was in the lab	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. never because the virtual environment overwhelmed me	7

12. How many hours per week on the average do you spend playing computer or video games (if any)?

The number of hours per week I play computer or video games is:_____

13. On the whole, how do you rate your expertise with respect to computer or video games?

My expertise with computer or video games is:	**Please tick against your answer**
1. Complete novice	1
2. …	2
3. …	3
4. …	4
5. …	5
6. …	6
7. Expert	7

14. What objects do you remember being in the virtual environment?

15. Were there any objects you didn't recognise in the virtual environment?
Yes / No

If so, what did you think they might have been?

16. Further Comments

Please write down any further comments that you wish to make about your experience. In particular, what things helped to give you a sense of 'really being' in the virtual environment, and what things acted to 'pull you out' and make you more aware of 'reality'?

Subject ID: _____________

Reminder - all answers will be treated entirely confidentially.
Thank you once again for participating in this study, and helping with our research. Please do not discuss this with anyone for four days. This is because the study is continuing, and you may happen to speak to someone who may be taking part.

Presence Questionnaire - Experiment II

Subject ID:_______

Questionnaire

The following questions relate to your experiences..

1. Please rate the extent to which you were aware of background sounds in the real room in which this experience was actually taking place. Rate this on the following scale from 1 to 7 (where for example 1 means that you were hardly aware at all of the background sounds):

While in the virtual reality I was aware of background sounds from the real room:	**Please tick against your answer**
1. not at all ...	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. very much ...	7

2. How dizzy, sick or nauseous did you feel resulting from the experience, if at all? Please answer on the following 1 to 7 scale.

I felt sick or dizzy or nauseous during or as a result of the experience...	**Please tick against your answer**
1. not at all	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. very much so	7

3. To what extent were there times during the experience when the virtual environment became the "reality" for you, and you almost forgot about the "real world" of the room (room G01a) in which the whole experience was really taking place?

There were times during the experience when the virtual environment became more real for me compared to the "real world"...	**Please tick against your answer** **1st Environment**	**Please tick against your answer** **2nd Environment**
1. at no time	1	1
2.	2	2
3.	3	3
4.	4	4
5.	5	5
6.	6	6
7. almost all of the time	7	7

4. When you think back about your experience, do you think of the virtual environment more as *images that you saw*, or more as *somewhere that you visited* ? Please answer on the following 1 to 7 scale.

The virtual environment seems to me to be more like...	**Please tick against your answer** **1st Environment**	**Please tick against your answer** **2nd Environment**
1. images that I saw	1	1
2.	2	2
3.	3	3
4.	4	4
5.	5	5
6.	6	6
7. somewhere that I visited	7	7

5. During the time of the experience, which was strongest on the whole, your sense of being in the virtual environment, or of being in the real world (room G01a) ?

I had a stronger sense of being in...	**Please tick against your answer** **1st Environment**	**Please tick against your answer** **2nd Environment**
1. the real world (room G01a)	1	1
2.	2	2
3.	3	3
4.	4	4
5.	5	5
6.	6	6
7. the virtual environment presented on the HMD	7	7

6. Overall, how well do you think that you achieved your task?

I achieved my task...	**Please tick against your answer**
1. not very well at all	1
2.	2
3.	3
4.	4
5.	5
6.	6
7. very well	7

7. Consider your memory of being in the virtual environment. How similar in terms of the *structure of the memory* is this to the structure of the memory of other *places* you have been today? By ‘structure of the memory’ consider things like the extent to which you have a visual memory of the environment, whether that memory is in colour, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such *structural* elements.

I think of the virtual environment as a place in a way similar to other places that I've been today....	**Please tick against your answer** **1st Environment**	**Please tick against your answer** **2nd Environment**
1. not at all	1	1
2.	2	2
3.	3	3
4.	4	4
5.	5	5
6.	6	6
7. very much so	7	7

8. During the time of the experience, did you often think to yourself that you were actually just standing in a room wearing a helmet or did the virtual environment overwhelm you?

During the experience I often thought that I was really standing in the room wearing a helmet....	**Please tick against your answer** **1st Environment**	**Please tick against your answer** **2nd Environment**
1. most of the time I realised I was in the room	1	1
2.	2	2
3.	3	3
4.	4	4
5.	5	5
6.	6	6
7. never because the virtual environment overwhelmed me	7	7

9. Were there any specific moments in the experiment at which time the environment seemed overwhelming?
Yes / No

10. If Yes, then can you try to describe and pinpoint that time for us:

11. What objects do you remember being in the virtual environments?

12. Were there any objects you didn't recognise in the virtual environments?
Yes / No

13. If Yes, then what did you think they might have been?

14. Further Comments

Please write down any further comments that you wish to make about your experience. In particular, what things helped to give you a sense of ‘really being’ in the virtual environment, and what things acted to ‘pull you out’ and make you more aware of ‘reality’?

Subject ID: _____________

Reminder - all answers will be treated entirely confidentially.
Thank you once again for participating in this study, and helping with our research. Please do not discuss this with anyone for six months. This is because the study is continuing, and you may happen to speak to someone who may be taking part.

Appendix B

Further Data

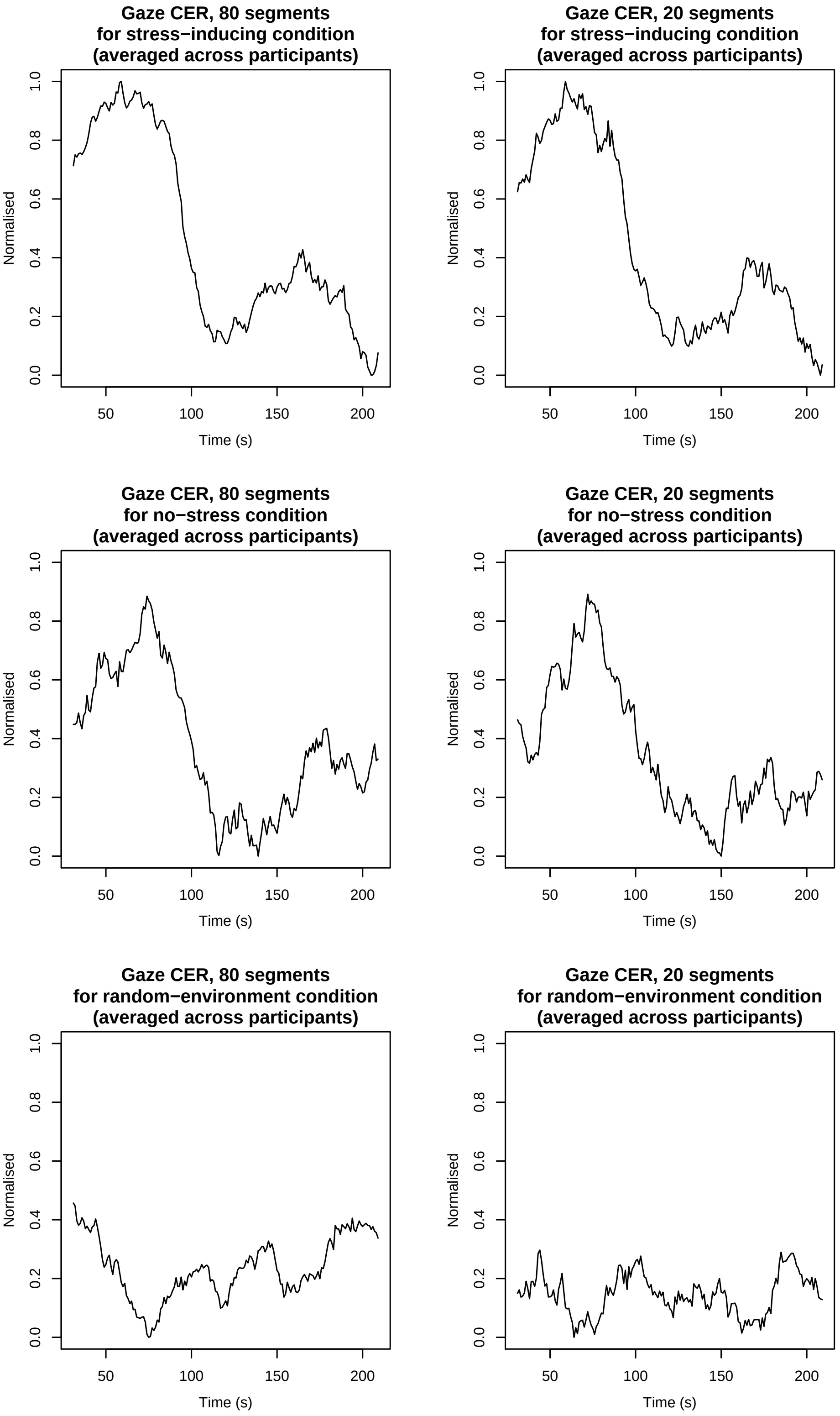


Figure B.1: Conditional Entropy Rate for HeadOnly data, averaged across participants ($n = 42$). Includes participants for whom eye-tracking was not calibarated, providing a slightly clearer picture than Figure 5.13.

Bibliography

Andreassi, J. L. (2000), *Psychophysiology Human Behavior and Physiological Response*, 4th edition, L. Erlbaum.

Antes, J. R. (1974), The time course of picture viewing, *Journal of Experimental Psychology* **103**(1), 62–70.

Atkinson, J. and Braddick, O. (2003), Neurobiological models of normal and abnormal visual development, *in* M. de Haan and M. H. Johnson (eds.), *The Cognitive Neuroscience of Development*, Psychology Press, pp.43–71.

Attneave, F. (1954), Some Informational Aspects of Visual Perception, *Psychological Review* **61**(3), 183–193.

Attneave, F. (1959), *Applications of Information Theory to Psychology*, Holt, Rinehart and Winston.

Banos, R., Botella, C., Garcia-Palacios, A., Villa, H., Perpina, C. and Alcaniz, M. (2000), Presence and Reality Judgement in Virtual Environments: A Unitary Construct?, *CyberPsychology & Behavior* **3**(3).

Barfield, W., Baird, K. and Bjorneseth, O. (1998), Presence in virtual environments as a function of type of input device and display update rate, *Displays* **19**(2), 91–98.

Barfield, W. and Weghorst, S. (1993), The Sense of Presence Within Virtual Environments: A Conceptual Framework, *in* G. Salvendy and M. Smith. (eds.), *Human-Computer Interaction:Software and Hardware Interfaces*, Elsevier Publisher, pp.699–704.

Basdogan, C. and Srinivasan, M. (2002), Haptic Rendering in Virtual Environments, *in* K. M. Stanney (ed.), *Handbook of virtual environments: Design, implementation, and applications*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc, pp.117–134.

Biederman, I. (1987), Recognition-by-Components: A Theory of Human Image Understanding, *Psychological Review* **94**(2), 115–147.

Biguer, B., Prablanc, C. and Jeannerod, M. (1984), The Contribution of Coordinated Eye and Head Movements in Hand Pointing Accuracy, *Experimental Brain Research* **55**, 462–469.

Biocca, F. (1997), The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments, *Journal of Computer-Mediated Communication* **3**(2).

Blackmon, T. T., an Dimitri A. Chernyak, Y. F. H., Azzariti, M. and Stark, L. (1999), Dynamic Scanpaths: Eye Movement Analysis Methods, *in Proceedings of SPIE: Human Vision and Electronic Imaging IV*, Vol. 3644, pp.511–519.

Bles, W. and Wertheim, A. H. (2000), Appropriate Use of Virtual Environments to Minimise Motion Sickness, Technical Report RTO MP58, Human Factors Research Institute.

Bradburn, N. M., Sudman, S. and Wansink, B. (2004), *Asking Questions: The Definitive Guide to Questionnaire Design - For Market Research, Political Polls, and Social and Health Questionnaires (Research Methods for the Social Sciences)*, Jossey-Bass.

Brillouin, L. (1962), *Science and Information Theory*, 2nd edition, Academic Press.

Brogni, A., Slater, M. and Steed, A. (2003), More Breaks Less Presence, *in Proceedings of the 6th Annual International Workshop on Presence 2003*.

Bron, A. J., Tripathi, R. C. and Tripathi, B. J. (1997), *Wolff's Anatomy of the Eye and Orbit*, 8th edition, Hodder Arnold.

Buser, P. and Imbert, M. (1987), *Vision*, MIT Press.

Buswell, G. T. (1935), *How people look at pictures: A study of the psychology of perception in art*, University of Chicago Press.

Bystrom, K., Barfield, W. and Hendrix, C. (1999), A conceptual model of the sense of presence in virtual environments, *Presence-Teleoperators and Virtual Environments* **8**(2), 241–244.

Carpenter, R. (1977), *Movements of the Eyes*, Pion Ltd.

Cater, K., Chalmers, A. and Ward, G. (2003), Detail to Attention: Exploiting Visual Tasks for Selective Rendering, *in Eurographics Symposium on Rendering 2003*, pp.270–280.

Chatfield, C. and Collins, A. (1980), *Introduction to Multivariate Analysis*, Chapman & Hall, CRC Press.

Chertoff, D. B., Schatz, S. L., McDaniel, R. and Bowers, C. A. (2008), Improving Presence Theory Through Experiental Design, *Presence-Teleoperators and Virtual Environments* **17**(4), 405–413.

Cho, D., Park, J., Kim, G. J., Hong, S., Han, S. and Lee, S. (2003), The Dichotomy of Presence Elements: The Where and What, *in Proceedings of IEEE Virtual Reality*, IEEE Computer Society, p.273.

Choi, Y. S., Mosley, A. D. and Stark, L. W. (1995), String Editing Analysis of Human Visual Search, *Optometry and Vision Science* **72**(7), 439–451.

Communicore (1996), *Patient Awareness During Anaesthesia*, Communicore.

Cover, T. M. and Thomas, J. A. (2006), *Elements of Information Theory*, 2nd edition, Wiley-Interscience.

Cruz-Neira, C., D.J.S and DeFanti, T. (1993), Surround-screen projection-based virtual reality: the design and implementation of the CAVE, *in Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, ACM Press, pp.135–142.

Danforth, R., Duchowski, A., Geist, R. and McAliley, E. (2000), A Platform for Gaze-Contingent Virtual Environments, *in AAAI Symposium on Smart Graphics 2000*, American Association for Artificial Intelligence, pp.66–70.

Dawson, M. E., Schell, A. M. and Filion, D. L. (1990), The Electrodermal System, *in* J. T. Cacioppo and L. G. Tassinary (eds.), *Principles of psychophysiology: Physical, social, and inferential elements*, Cambridge University Press, pp.295–324.

Dawson, M. E., Schell, A. M. and Filion, D. L. (2000), The Electrodermal System, *in* J. T. Cacioppo and L. G. Tassinary (eds.), *Principles of psychophysiology: Physical, social, and inferential elements*, 2nd edition, Cambridge University Press, pp.206–208.

Dawson, M. E., Schell, A. M. and Filion, D. L. (2007), The Electrodermal System, *in* J. T. Cacioppo, L. G. Tassinary and G. G. Berntson (eds.), *Handbook of Psychophysiology*, 3rd edition, Cambridge University Press, pp.159–181.

Dillon, C., Keogh, E., Freeman, J. and Davidoff, J. (2000), Aroused and Immersed: The Psychophysiology of Presence, *in Proceedings of the 3rd International Workshop on Presence 2000*.

Draper, J., Kaber, D. and Usher, J. (1998), Telepresence, *Human Factors* **40**(3), 354–375.

Duchowski, A. T. (2003), *Eye Tracking Methodology: Theory and Practice*, Springer-Verlag.

Ehrsson, H. H. (2007), The experimental induction of out-of-body experiences, *Science* **317**(5841), 1048–1048.

Ellis, S. R. (1996), Presence of mind: A reaction to Thomas Sheridan's "further musings on the psychophysics of presence", *Presence-Teleoperators and Virtual Environments* **5**(2), 247–259.

Ellis, S. R. and Stark, L. (1978), Eye movements during the viewing of Necker cubes, *Perception* **7**, 575–581.

Ellis, S. R. and Stark, L. (1979), Reply to Piggins, *Perception* **8**, 721–722.

Ellis, S. R. and Stark, L. W. (1986), Statistical Dependency in Visual Scanning, *Human Factors* **28**(4), 421–438.

Emmelkamp, P., Krijn, M., Hulsbosch, A., de Vries, S., Schuemie, M. and van der Mast, C. (2002), Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia, *Behaviour Research and Therapy* **40**(5), 509–516.

Engbert, R. and Kliegl, R. (2004), Microsaccades Keep the Eyes' Balance During Fixation, *Psychological Science* **15**(6), 431–436.

Everitt, B. S., Landau, S. and Leese, M. (2001), *Cluster Analysis*, Arnold.

Fenz, W. D. and Epstein, S. (1967), Gradients of Physiological Arousal in Parachutists as a Function of an Approaching Jump, *Psychosomatic Medicine* **29**, 33–51.

Findlay, J. M. and Gilchrist, I. D. (2003), *active vision: The Psychology of Looking and Seeing*, Oxford University Press.

Floridi, L. (2005), Presence: From Epistemic Failure to Successful Observability", *Presence-Teleoperators And Virtual Environments* **14**(6), 656–667.

Freeman, D., Dunn, G., Garety, P., Bebbington, P., Slater, M., Kuipers, E., Fowler, D., Green, C., Jordan, J. and Ray, K. (2005a), The psychology of persecutory ideation I - A questionnaire survey, *Journal of Nervous and Mental Disease* **193**(5), 302–308.

Freeman, D., Garety, P., Bebbington, P., Slater, M., Kuipers, E., Fowler, D., Green, C., Jordan, J., Ray, K. and Dunn, G. (2005b), The psychology of persecutory ideation II - A virtual reality experimental study, *Journal of Nervous and Mental Disease* **193**(5), 309–315.

Freeman, D., Slater, M., Bebbington, P., Garety, P., Kuipers, E., Fowler, D., Read, C., Jordan, J. and Vinayagamoorthy, V. (2003), Can virtual reality be used to investigate persecutory ideation?, *Journal of Nervous and Mental Disease* **191**(8), 509–514.

Freeman, J., Avons, S., Meddis, R., Pearson, D. and IJsselsteijn, W. (2000), Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli, *Presence-Teleoperators and Virtual Environments* **9**(2), 149–164.

Freeman, J., Avons, S., Pearson, D. and IJsselsteijn, W. (1999), Effects of sensory information and prior experience on direct subjective ratings of presence, *Presence-Teleoperators and Virtual Environments* **8**(1), 1–13.

Freeman, J., Lessiter, J., Keogh, E., Bond, F. and Chapman, K. (2004), Relaxation Island: virtual and really relaxing, *in Proceedings of the 7th International Workshop on Presence 2004*.

Frey, W., Zyda, M., McGhee, R. and Cockayne, W. (1996), Off-the-Shelf, Real-Time, Human Body Motion Capture for Synthetic Environments, Technical Report NPSCS-96-003, Computer Science Dept., Naval Postgraduate School.

Fua, P., Gruen, A., D'Apuzzo and Plankers, R. (2002), Markerless Full Body Shape and Motion Capture from Video Sequences, *International Archives of Photogrammetry and Remote Sensing* **34**(5), 256–261.

Gamberini, L. and Spagnolli, A. (2003), On the Relationship between Presence and Usability: a Situated, Action-Based Approach to Virtual Environments, *in* G. Riva, F. Davide and W. IJsselsteijn (eds.), *Being There: Concepts, effects and measurement of user presence in synthetic environments*, Ios Press, 2003, Amsterdam, The Netherlands.

Gardner, H. J. and Martin, M. A. (2007), Analyzing Ordinal Scales in Studies of Virtual Environments: Likert or Lump It!, *Presence: Teleoperators and Virtual Environments* **16**(4), 439–446.

Gibson, J. (1979), *The Ecological approach to visual perception*, Boston: Houghton Mifflin.

Goldstein, E. B. (2002), *Sensation and Perception*, 6th edition, Wadsworth.

Gregory, R. L. (1977), *Eye and Brain: The Psychology of Seeing*, 3rd edition, Weidenfeld and Nicolson.

Gregory, R. L. (1998), *Eye and Brain: The Psychology of Seeing*, 5th edition, Oxford University Press.

Grimes, J. (1996), On the failure to detect changes in scenes across saccades, *in* K. Akins (ed.), *Perception*, Vol. 2, New York: Oxford University Press, pp.89–110. Vancouver Studies in Cognitive Science.

Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A., Friedman, D. and Slater, M. (2004), Heart-Rate Variability and Event-Related ECG in Virtual Environments, *in Proceedings of the The 7th Annual International Workshop on Presence*, pp.240–245.

Haber, J., Myszkowski, K., Yamauchi, H. and Seidel, H.-P. (2001), Perceptually Guided Corrective Splatting, *Eurographics 2001* **20**(3), 270–280.

Hafed, Z. and Clark, J. (2002), Microsaccades as an overt measure of covert attention shifts, *Vision Research* **42**(22), 2533–2545.

Hand, C. (1994), Other Faces of Virtual Reality, *in Proceedings of the East-West International Conference on Multimedia, Hypermedia and Virtual Reality*, pp.69–74.

Harrison, S. and Dourish, P. (1996), Re-place-ing space: The roles of space and place in collaborative systems, *in Proceedings of the 1996 ACM conference on Computer Supported Cooperative Work*, New York, ACM, pp.67–76.

Harvey, M. and Sanchez-Vives, M. (2005), The Binding Problem in Presence Research, *Presence-Teleoperators and Virtual Environments* **15**(5), 616–621.

Hayhoe, M. and Ballard, D. (2005), Eye Movements in Natural Behavior, *TRENDS in Cognitive Sciences* **9**(4).

Hayhoe, M. M., Ballard, D. H., Triesch, J., Shinoda, H., Aivar, P. and Sullivan, B. (2002), Vision in natural and virtual environments, *in Proceedings of the 2002 symposium on Eye tracking research and applications*, pp.7–13.

Hayhoe, M. M., Shrivastava, A., Mruczek, R. and Pelz, J. B. (2003), Visual memory and motor planning in a natural task, *Journal of Vision* **3**, 49–63.

Hecht, D., Rainer, M. and Halevy, G. (2006), Multimodal Virtual Environments: Response Times, Attention, and Presence, *Presence-Teleoperators and Virtual Environments* **15**(5), 515–523.

Heeter, C. (1992), Being there: The Subjective Experience of Presence, *Presence-Teleoperators and Virtual Environments* **1**(2), 262–271.

Heeter, C. (2003), Reflections on Real Presence by a Virtual Person, *Presence-Teleoperators and Virtual Environments* **12**(4), 335–345.

Held, R. and Durlach, N. (1992), Telepresence, *Presence: Teleoperators and Virtual Environments, 1992* **1**(1), 109–112.

Helmholtz, V. (1925), Treatise on Physiological Optics.

Henderson, J. (2003), Human gaze control in real-world scene perception, *Trends in Cognitive Sciences* **7**(11), 498–504.

Henderson, R. (1997), *A Guide To Laser Safety*, Springer.

Hendrix, C. (1994), Exploratory Studies on the Sense of Presence in Virtual Environments as a Function of Visual and Auditory Display Parameters, Master's thesis, University of Washington.

Hendrix, C. and Barfield, W. (1996), Presence within Virtual Environments as a Function of Visual Display Parameters, *Presence-Teleoperators and Virtual Environments* **5**(3), 274–289.

Herrera, G., Jordan, R. and Vera, L. (2006), Agency and Presence: A Common Dependence on Subjectivity?, *Presence-Teleoperators and Virtual Environments* **15**(5), 539–552.

Hoffman, H., Richards, T., Coda, B., Richards, A. and Sharar, S. (2003), The illusion of presence in immersive virtual reality during an fMRI brain scan, *CyberPsychology & Behavior* **6**(2), 127–131.

Hooge, I. T., Over, E. A., van Wezel, R. J. and Frens, M. A. (2005), Inhibition of Return is not a Foraging Facilitator in Saccadic Search and Free Viewing, *Vision Research* **45**, 1901–1908.

Hu, H. H., Gooch, A. A., Creem-Regehr, S. H. and Thompson, W. B. (2002), Visual Cues for Perceiving Distances from Objects to Surfaces, *Presence: Teleoperators and Virtual Environments* **11**(6), 652–664.

Hwang, A. D., Higgins, E. C. and Pomplun, M. (2009), A model of top-down attentional control during visual search in complex scenes, *Journal of Vision* **9**(5), 1–18.

IJsselsteijn, W. (2002), Elements of a multi-level theory of presence: Phenomenology, mental processing and neural correlates, *in Proceedings of PRESENCE 2002*, pp.245–259.

IJsselsteijn, W., de Ridder, H., Freeman, J. and Avons, S. (2000), Presence: Concept, determinants and measurement, *in* B. E. Rogowitz (ed.), *Proceedings of SPIE: Human Vision and Electronic Imaging V*, pp.520–529.

IJsselsteijn, W., Freeman, J. and Ridder, H. (2001), Presence: Where are we?, *Cyberpsychology and Behavior* **4**(2), 179–182.

IJsselsteijn, W., Ridder, H., Hamberg, R., Bouwhuis, D. and Freeman, J. (1998), Perceived depth and the feeling of presence in 3DTV, *Displays* **18**(4), 207–214.

Itti, L. and Koch, C. (2000), A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Research* **40**, 1489–1506.

Itti, L., Koch, C. and Niebur, E. (1998), A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. PAMI* **20**(11), 1254.

Johansen, S. A. and Hansen, J. P. (2006), Do We Need Eye Trackers to Tell Where People Look?, *in CHI '06 Extended Abstracts on Human Factors in Computing Systems*, pp.923–928.

Jordan, J. and Slater, M. (2009), An Analysis of Eye Scan Path Entropy in a Progressively Forming Virtual Environment, *Presence-Teleoperators and Virtual Environments* **18**(3), 185–199.

Jovancevic, J., Sullivan, B. and Hayhoe, M. (2006), Control of attention and gaze in complex environments, *Journal of Vision* **6**, 1431–1450.

Kennedy, R. S. and Stanney, K. M. (1997), Aftereffects of Virtual Enviornment Exposure: Psychometric Issues, *in* M. Smith, G. Salvendy and R. Koubek (eds.), *Design of computing systems: Social and ergonomic considerations*, Vol. B, Amsterdam, Netherlands: Elsevier Science, pp.897–900.

Kennedy, R. S., Lane, N. E., Berbaum, K. S. and Lilienthal, M. G. (1993), Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness, *The International Journal of Aviation Psychology* **3**(3), 203–220.

Kern, J. C. (2006), Pig Data and Bayesian Inference on Multinomial Probabilities, *Journal of Statistics Education* **14**(3).

Klein, R. M. (1988), Inhibitory tagging system facilitates visual search, *Nature* **334**(4), 430–431.

Klein, R. M. and McInnes, W. J. (1999), Inhibition of Return is a Foraging Facilitator in Visual Search, *Psychological Science* **10**(4), 346–352.

Koch, C. and Ullman, S. (1985), Shifts in selective visual attention : Towards the underlying neuronal circuitry, *Human Neurobiology* **4**, 219–227.

Koffka, K. (1935), *Principles of Gestalt Psychology*, Lund Humphries.

Kootstra, G., Nederveen, A. and de Boer, B. (2006), On The Bottom-Up and Top-Down Influences of Eye Movements, *in Proceedings of CogSci 2006*, p.2538.

Krieger, G., Rentschler, I., Hauske, G., Schill, K. and Zetzsche, C. (2000), Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics, *Spatial Vision* **13**(2,3), 201–214.

Laarni, J., Ravaja, N. and Saari, T. (2003), Using eye tracking and psychophysiological methods to study spatial presence, *in Proceedings of The 6th Annual International Workshop on Presence*, p.38.

Land, M. (2004), The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations, *Experimental Brain Research* **159**, 151–160.

Land, M. F. and Hayhoe, M. (2001), In what ways do eye movements contribute to everyday activities?, *Vision Research* **41**, 3559–3565.

Lee, K. (2004), Why presence occurs: Evolutionary psychology, media equation, and presence, *Presence-Teleoperators and Virtual Environments* **13**(4), 494–505.

Lessiter, J., Freeman, J., Keogh, E. and Davidoff, J. (2001), A cross-media presence questionnaire: The ITC-Sense of Presence Inventory, *Presence-Teleoperators and Virtual Environments* **10**(3), 282–297.

Locher, P. J. and Nodine, C. F. (1974), The role of scanpaths in the recognition of random shapes, *Perception & Psychophysics* **15**(2), 308–314.

Loftus, E. (2003), Our changeable memories: legal and practical implications, *Nature* **4**, 231–234.

Lombard, M. and Ditton, T. (1997), At the Heart of It All: The Concept of Presence, *Journal of Computer-Mediated Communication* **3**(2).

Loomis, J. (1992), Distal attribution and presence, *Presence-Teleoperators and Virtual Environments* **1**, 113–119.

Mack, A. and Rock, I. (1982), *Innattentional blindness*, Cambridge, MA: MIT Press.

MacKay, D. J. (2003), *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

Mackworth, N. H. and Morandi, A. J. (1967), The gaze selects informative details within pictures, *Perception & Psychophysics* **2**(11), 547–552.

Mania, K. and Chalmers, A. (2001), The effects of levels of immersion on memory and presence in virtual environments: A reality centered approach, *CyberPsychology & Behavior* **4**(2), 247–264.

Mania, K. and Robinson, A. (2004), An experimental exploration of the relationship between subjective impressions of illumination and physical fidelity, *Computers and Graphics* **29**, 49–56.

Mania, K., Robinson, A. and Brandt, K. R. (2005), The Effect of Memory Schemas on Object Recognition in Virtual Environments, *Presence: Teleoperators and Virtual Environments* **14**(5), 606–615.

Mania, K., Troscianko, T., Hawkes, R. and Chalmers, A. (2003), Fidelity Metrics for Virtual Environment Simulations Based on Spatial Memory Awareness States, *Presence-Teleoperators and Virtual Environments* **12**(3), 296–310.

Marmitt, G. and Duchowski, A. T. (2004), Modeling Visual Attention in VR: Measuring the Accuracy of Predicted Scanpaths, *in Eurographics 2002*, pp.217–226.

Marr, D. (1982), *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York: Freeman.

Marsh, T. (2003), Presence as experience: Film informing ways of staying there, *Presence-Teleoperators And Virtual Environments* **12**(5), 538–549.

May, G. and Badcock, D. (2002), Vision and virtual environments, *in* K. M. Stanney (ed.), *Handbook of virtual environments: Design, implementation, and applications*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc, pp.29–63.

Meehan, M., Insko, B., Whitton, M. and Brooks, F. (2002), Physiological measures of presence in stressful virtual environments, *ACM Transactions on Graphics* **21**(3), 645–652.

Meehan, M. J. (2001), Physiological Reaction as an Objective Measure of Presence in Virtual Environments, PhD thesis, University of North Carolina.

Meehan, M., Razzaque, S., Whitton, M. and Jr, F. B. (2003), Effect of latency on presence in stressful virtual environments, *in Proceedings IEEE*, pp.141–148.

Mennie, N., Hayhoe, M. M. and Sullivan, B. T. (2007), Look-ahead fixations: anticipatory eye movements in natural tasks, *Experimental Brain Research* **179**(3), 427–442.

Minsky, M. (1980), Telepresence, *Omni (magazine)* pp.45–52.

Mohsin, S. M. (1984), *Research Methods in Behavioural Sciences*, Orient Longman Ltd.

Morimoto, C. H. and Mimica, M. R. (2005), Eye gaze tracking techniques for interactive applications, *Computer Vision and Image Understanding* **98**, 4–24.

Morningstar, C. and Farmer, F. R. (1991), The Lessons of Lucasfilm's Habitat, *in* M. Benedikt (ed.), *Cyberspace: First Steps*, MIT Press, pp.273–301.

Mulder, M., van Paassen, M. M. R. and Boer, E. R. (2004), Exploring the Roles of Information in the Manual Control of Vehicular Locomotion: From Kinematics and Dynamics to Cybernetics, *Presence-Teleoperators And Virtual Environments* **13**(5), 535–548.

Munn, P. and Drever, E. (2004), *Using Questionnaires in Small-scale Research: A Beginner's Guide*, The SCRE Centre, UK.

Navarro, G. (2001), A Guided Tour to Approximate String Matching, *ACM Computing Surveys* **33**(1), 31–88.

Noton, D. and Stark, L. (1971), Eye Movements and Visual Perception, *Scientific American* **224**(6), 34–43.

Nunez, D. (2004), Working memory and presence: Reconsidering the role of attention in presence., *in Proceedings of the Presence 2004 Workshop*, pp.44–48.

Ohshima, T., Yamamoto, H. and Tamura, H. (1996), Gaze-Directed Adaptive Rendering for Interacting with Virtual Space, *in Proceedings of VRAIS*, pp.103–110.

O'Keefe and Nadel, L. (1978), *The Hippocampus as a Cognitive Map*, Oxford University Press.

Oliva, A. and Torralba, A. (2006), Building the Gist of a Scene: The Role of Global Images Features in Recognition, *Progress in Brain Research* **155**, 23–36.

Oliva, A., Torralba, A., Castelhano, M. S. and Henderson, J. M. (2003), Top-Down Control of Visual Attention in Object Detection, *in Proceedings of international conference on Image Processing*, Barcelona, Catalonia: IEEE Press, pp.253–256.

Papillo, J. F. and Shapiro, D. (1990), The Cardiovascular System, *in* J. T. Cacioppo and L. G. Tassinary (eds.), *Principles of psychophysiology: Physical, social, and inferential elements*, Cambridge University Press, pp.456–512.

Parkhurst, D. J. and Niebur, E. (2005), Stimulus-Driven Guidance of Visual Attention in Natural Scenes, *in* L. Itti, G. Rees and J. Tsotsos (eds.), *Neurobiology of Attention*, Elsevier, pp.240–245.

Pelton, L. H. and Solley, C. M. (1968), Acceleration of Reversals of a Necker Cube, *American Journal of Psychology* **81**(4), 585–588.

Pelz, J. B., Canosa, R., Babcock, J. and Barber, J. (2001a), Visual Perception in Familiar, Complex Tasks, *in Proceedings of the International Conference on Image Processing*, pp.12–15.

Pelz, J. B., Canosa, R. L., Kucharczyk, D., Babcock, J., Silver, A. and Konno, D. (2000), Portable Eyetracking: A Study of Natural Eye Movements, *in* B. Rogowitz and T. Pappas (eds.), *Proceedings of SPIE: Human Vision and Electronic Imaging V*, p.3659.

Pelz, J. B., Hayhoe, M. M., Ballard, D. H., Shrivastava, A., Bayliss, J. D. and von der Heyde, M. (1999), Development of a virtual laboratory for the study of complex human behavior, *in Proceedings of SPIE, The Engineering Reality of Virtual Reality*, Vol. 3639B.

Pelz, J., Hayhoe, M. and Loeber, R. (2001b), The coordination of eye, head, and hand movements in a natural task, *Experimental Brain Research* **139**, 266–277.

Pertaub, D., Slater, M. and Barker, C. (2001), An experiment on fear of public speaking in virtual reality, *Stud Health Technol Inform* **81**, 372–378.

Petkova, V. I. and Ehrsson, H. H. (2008), If I Were You: Perceptual Illusion of Body Swapping, *PLoS ONE* **3**(12), 1–9.

Poole, A. (2005), Tips for Using Eyetrackers in HCI Experiments, http://www.alexpoole.info/academic/lecturenotes.html.

Posner, M. I. and Cohen, Y. (1984), Components of Visual Orienting, *in* H. Bouma and D. Bouwhuis (eds.), *Attention and Performance*, Vol. X, Lawrence Erlbaum, pp.531–556.

Posner, M. I., Snyder, C. R. and Davidson, B. J. (1980), Attention and the Detection of Signals, *Journal of Experimental Psychology: General* **109**(2), 160–174.

Potter, M. C. (1976), Short-Term Conceptual Memory for Pictures, *Journal of Experimental Psychology: Human Learning and Memory* **2**(5), 509–522.

Privitera, C. M. and Stark, L. W. (2000), Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(9), 970–982.

Prothero, J. and Hoffman, H. (1995), Widening the Field-of-View Increases the Sense of Presence in Immersive Virtual Environments, Technical Report TR-95-2, Human Interface Technology Laboratory.

Pugnetti, L., Meehan, M. and Mendozzi, L. (2001), Psychophysiological correlates of virtual reality: A review, *Presence-Teleoperators and Virtual Environments* **10**(4), 384–400.

Quastler, H. (1953), *Information Theory in Biology*, University of Illinois Press.

Reddy, M. (1997), Perceptually Modulated Level of Detail for Virtual Environments, PhD thesis, University of Edinburgh.

Regenbrecht, H. and Schubert, T. (2002), Real and illusory interactions enhance presence in virtual environments, *Presence-Teleoperators and Virtual Environments* **11**(4), 425–434.

Reinagel, P. and Zador, A. M. (1999), Natural scene statistics at the centre of gaze, *Network: Computation in Neural Systems* **10**, 341–350.

Renaud, P., Chartier, S., Albert, G., DeCarie, J., Cournoyer, L.-G. and Bouchard, S. (2007), Presence as Determined by Fractal Perceptual-Motor Dynamics, *CyberPsychology & Behavior* **10**(1).

Renaud, P., DeCarie, J., Gourd, S.-P., Paquin, L.-C. and Bouchard, S. (2003), Eye-Tracking in Immersive Environments: A General Methodology to Analyze Affordance-Based Interactions from Oculomotor Dynamics, *CyberPsychology & Behavior* **6**(5).

Rinalducci, E. J. (1996), Characteristics of Visual Fidelity in the Virtual Environment, *Presence-Teleoperators and Virtual Environments* **5**(3), 330–345.

Riva, G. (2006), Virtual Reality, *in Wiley Encyclopedia of Biomedical Engineering*, Vol. 4, John Wiley & Sons, p.117.

Riva, G., Waterworth, J. A. and Waterworth, E. L. (2004), The Layers of Presence: A Bio-cultural Approach to Understanding Presence in Natural and Mediated Environments, *CyberPsychology & Behavior* **7**(4), 405–419.

Ron, S. and Berthoz, A. (1991), Eye and head coupled and dissociated movements during orientation to a double step visual target displacement, *Experimental Brain Research* **85**, 196–207.

Salvucci, D. and Goldberg, J. (2000), Identifying fixations and saccades in eye-tracking protocols, *in Proceedings of the Eye Tracking Research and Applications Symposium*, ACM Press, pp.71–78.

Sanchez-Vives, M. and Slater, M. (2005), From Presence to Consciousness through Virtual Reality, *Nature Reviews Neuroscience* **6**(4), 332–339.

Schloerb, D. W. (1995), A Quantitative Measure of Telepresence, *Presence-Teleoperators and Virtual Environments* **4**(1), 64–80.

Schneider, D. and Godard, R. (1996), Virtual Environments for Education, Research and Life, *in Accepted Positional Papers for the WWW5 Workshop on Virtual Environments and WWW*.

Schnipke, S. K. and Todd, M. W. (2000), Trials and tribulations of using an eye-tracking system, *in CHI: extended abstracts on Human Factors in computing systems*, ACM, pp.273–274.

Schubert, T., Friedmann, F. and Regenbrecht, H. (1999), Decomposing the sense of presence: Factor analytic insights, *in Proceedings of the 2nd International Workshop on Presence*.

Schubert, T., Friedmann, F. and Regenbrecht, H. (2001), The experience of presence: Factor analytic insights, *Presence-Teleoperators and Virtual Environments* **10**(3), 266–281.

Schuemie, M. J., van der Straaten, P., Krijn, M. and van der Mast, C. A. (2001), Research on Presence in Virtual Reality: A Survey, *CyberPsychology & Behavior* **4**(2), 183–201.

Sheridan, T. (1992), Musings on Telepresence and Virtual Presence, *Presence: Teleoperators and Virtual Environments* **1**(1), 120–126.

Shilling, R. and Shinn-Cunningham, B. (2002), Virtual auditory displays, *in* K. M. Stanney (ed.), *Handbook of virtual environments: Design, implementation, and applications*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc, pp.65–92.

Slater, M. (1999), Measuring presence: A response to the Witmer and Singer presence questionnaire, *Presence-Teleoperators and Virtual Environments* **8**(5), 560–565.

Slater, M. (2002), Presence and the sixth sense, *Presence-Teleoperators and Virtual Environments* **11**(4), 435–439.

Slater, M. (2004), How colorful was your day? Why questionnaires cannot assess presence in virtual environments..., *Presence-Teleoperators and Virtual Environments* **13**(4), 484–493.

Slater, M. (2009), Place Illusion and Plausibility can lead to realistic behaviour in immersive virtual environments, *Philosophical Transactions of the Royal Society B* **364**, 3549–3557.

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N. and Sanchez-Vives, M. (2006a), A virtual reprise of the stanley milgram obedience experiments, *PLoS ONE* **1**(1).

Slater, M., Brogni, A. and Steed, A. (2003), Physiological Responses to Breaks in Presence: A Pilot Study, *in Proceedings of The 6th Annual International Workshop on Presence*, pp.3–23.

Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A. and Friedman, D. (2006b), Analysis of Physiological Responses to a Social Situation in an Immersive Virtual Environment, *Presence-Teleoperators and Virtual Environments* pp.553–569.

Slater, M. and Garau, M. (2007), The Use of Questionnaire Data in Presence Studies: Do Not Seriously Likert, *Presence: Teleoperators and Virtual Environments* **16**(4), 447–456.

Slater, M. and Steed, A. (2000), A virtual presence counter, *Presence-Teleoperators and Virtual Environments* **9**(5), 413–434.

Slater, M. and Usoh, M. (1993a), Presence in Immersive Virtual Environments, *in IEEE Virtual Reality Annual International Symposium (VRAIS)*, pp.90–96.

Slater, M. and Usoh, M. (1993b), Representation Systems, Perceptual Position, and Presence in Immersive Virtual Environments, *Presence-Teleoperators and Virtual Environments* **2**(3), 221–233.

Slater, M. and Usoh, M. (1994), Body Centred Interaction in Immersive Virtual Environments, in Artificial Life and Virtual Reality, N, *in* N. Thalmann and D. Thalmann (eds.), *Artificial Life and Virtual Reality*, John Wiley and Sons, pp.125–148.

Slater, M. and Wilbur, S. (1997), A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments, *Presence-Teleoperators and Virtual Environments* **6**(6), 603–616.

Slater, M., Khanna, P., Mortensen, J. and Yu, I. (2009a), Visual Realism Enhances Realistic Response in an Immersive Virtual Environment, *IEEE Computer Graphics Applications* **29**(3), 76–84.

Slater, M., Linakis, V., Usoh, M. and Kooper, R. (1996), Immersion, Presence, and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess, pp.163–172. ISBN: 89791-825-8.

Slater, M., Lotto, B., Arnold, M. M. and Sanchez-Vives, M. V. (2009b), How we experience immersive virtual environments: the concept of presence and its measurement, *Amario de Psicologia* **40**(2), 193–210.

Slater, M., Spanlang, B. and Corominas, D. (2010), Simulating Virtual Environments within Virtual Environments as the Basis for a Psychophysics of Presence, *ACM Transactions on Graphics* **29**(4).

Slater, M., Steed, A., McCarthy, J. and Maringelli, F. (1998), The influence of body movement on subjective presence in virtual environments, *Human Factors* **40**(3), 469–477.

Slater, M., Usoh, M. and Chrysanthou, Y. (1995), The influence of dynamic shadows on presence in immersive virtual environments, *in Selected papers of the Eurographics workshops on Virtual environments,* Springer-Verlag: Barcelona, Spain, pp.8–21.

Slater, M., Usoh, M. and Steed, A. (1994), Depth of Presence in Immersive Virtual Environments, *Presence-Teleoperators and Virtual Environments* **3**(2), 130–144.

Smith, S., Marsh, T., Duke, D. and Wright, P. (1998), Drowning in Immersion, *in Proceedings of UK VR-SIG '98*, pp.113–128.

Spagnolli, A. and Gamberini, L. (2002), Immersion/Emersion: Presence in hybrid environments, *in Proceedings of the fifth annual international workshop, Presence*, pp.421–434.

Stanney, K., Salvendy, G., Deisinger, J., DiZio, P., Ellis, S., Ellison, J., Fogleman, G., Gallimore, J., Singer, M., Hettinger, L., Kennedy, R., Lackner, J., Lawson, B., Maida, J., Mead, A., Mon-Williams, M., Newman, D., Piantanida, T., Reeves, L., Riedel, O., Stoffregen, T., Wann, J., Welch, R., Wilson, J. and Witmer, B. (1998), Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda, *International Journal of Human-Computer Interaction* **10**(2), 135–187.

Stark, L. (1995), How Virtual Reality Works!: Illusions of Vision in Real and Virtual Environments, *in Proceedings of SPIE: Symposium on Electronic Imaging: Science and Technology*, Vol. 2411, pp.5–10.

Stark, L., Ezumi, K., Nguyen, T., Paul, R., Tharp, G. and Yamashita, H. (1992), Visual Search in Virtual Environments, *in Proceedings of SPIE: Human Vision, Visual Processing and Digital Display III*, Vol. 1666, pp.577–589.

Stark, L. W. (1993), Top-Down Vision in Humans and Robots, *in* J. P. Allebach and B. E. Rogowitz (eds.), *Proceedings of SPIE: Human Vision, Visual Processing, and Digital Display IV*, Vol. 1913, pp.613–621.

Stark, L. W. and Choi, Y. S. (1996), Experimental Metaphysics: The Scanpath as an Epistemological Mechanism, *in* Zangemeister, Stiehl and Freksa (eds.), *Visual Attention and Cognition*, Elsevier, pp.3–69.

Stark, L. W., Privitera, C. M., Yang, H., Azzariti, M., Ho, Y. F., Blackmon, T. and Chernyak, D. (2001), Representation of human vision in teh brain: How does human perception recognize images?, *Journal of Electronic Imaging* **10**(1), 123–151.

Steuer, J. (1992), Defining Virtual Reality: Dimensions Determining Telepresence, *Journal of Communication* **42**, 73–93.

Strickland, D., Hodges, L., North, M. and Weghorst, S. (1997), Overcoming Phobias, *Communications of the ACM* **40**(8), 34–39.

Sutherland, I. (1965), The Ultimate Display, *in Proceedings of IFIP Congress*, pp.506–508.

Taylor, R., Hudson, T., Seeger, A., Weber, H., Juliano, J. and Helser, A. (2001), VRPN: A Device-Independent Network-Transparent VR Peripheral System, *in ACM Symposium on Virtual Reality Software and Technology*, ACM Press, pp.55–61.

Treisman, A. M. and Gelade, G. (1980), A Feature-Integration Theory of Attention, *Cognitive Psychology* **12**, 97–136.

Triesch, J., Ballard, D. H., Hayhoe, M. M. and Sullivan, B. T. (2003), What you see is what you need, *Journal of Vision* **3**, 86–94.

Triesch, J., Sullivan, B. T., Hayhoe, M. M. and Ballard, D. H. (2002), Saccade contingent updating in virtual reality, *in Proceedings of the 2002 symposium of Eye tracking research and applications*, pp.92–102.

Tsakiris, M. and Haggard, P. (2005), The Rubber Hand Illusion: Visuo-tactile Integration and Self-Attribution, *Journal of Experimental Psychology: Human Perception and Performance* **31**(1), 80–91.

Turner, P. and Turner, S. (2006), Place, Sense of Place, Presence, *Presence-Teleoperators and Virtual Environments* **15**(2), 204–217.

Ullman, S. (1996), *High-Level Vision*, MIT Press.

Usoh, M., Arthur, K., Whitton, M., Bastos, R., Steed, A., Slater, M. and Brooks, J. F. P. (1999), Walking $>$ walking-in-place $>$ flying, in virtual environments, *in Proceedings of the 26th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, pp.359–364.

Usoh, M., Catena, E., Arman, S. and Slater, M. (2000), Using presence questionnaires in reality, *Presence-Teleoperators and Virtual Environments* **9**(5), 497–503.

Viviani, P. (1990), Eye movements in visual search: cognitive, perceptual and motor control aspects, *in* E. Kowler (ed.), *Reviews of Oculomotor Research: Eye movements and their role in visual and cognitive processes*, Vol. 4, Elsevier Science, pp.353–393.

Vora, J., Nair, S., Gramopadhye, A. K., Melloy, B. J., Medlin, E., Duchowski, A. T. and Kanki, B. G. (2001), Using Virtual Reality Technology to Improve Aircraft Inspection Performance: Presence and Performance Measurement Studies, *Human Factors and Ergonomics Society Meeting* pp.1867–1871.

Waller, D. and Bachmann, E. R. (2006), The Borderline of Science: On the Value of Factor Analysis for Understanding Presence, *Presence-Teleoperators and Virtual Environments* **15**(2), 235–244.

Ware, C. (2008), *Visual Thinking for Design*, Morgan Kauffmann.

Ware, C. and Balakrishnan, R. (1994), Reaching for Objects in VR Displays: Lag and Frame Rate, *ACM Transactions on Computer-Human Interaction* **1**(4), 331–356.

Waterworth and Waterworth (2001), Focus, Locus, and Sensus: The Three Dimensions of Virtual Experience, *CyberPscyhology & Behavior* **4**(2), 203–213.

Watson, B., Walker, N. and Hodges, L. F. (1995), A User study evaluating level of detail degradation in the periphery of head-mounted displays, *in Proceedings of Framework for Interactive Virtual Environments (FIVE)*, pp.203–212.

Weigle, C. and Banks, D. C. (2008), Analysis of eye-tracking experiments performed on a Tobii T60, *in* K. Brner, M. T. Grhn, J. Park and J. C. Roberts (eds.), *Visualization and Data Analysis 2008, Proceedings of SPIE-IS&T Electronic Imaging*, Vol. 6809, pp.923–928.

Welch, R. (1997), The Presence of Aftereffects, *in Proceedings of the Seventh International Conference on Human-Computer Interaction*, Vol. 1, Elsevier Science Inc, pp.273–276.

Welch, R., Blackmon, T., Liu, A., Mellers, B. and Stark, L. (1996), The effects of pictorial realism, delay of visual feedback, and observer interactivity an the subjective sense of presence, *Presence-Teleoperators and Virtual Environments* **5**(3), 263–273.

Wiederhold, B. K., Jang, D. P., Kaneda, M., Cabral, I., Lurie, Y., May, T., Kim, I. Y., Wiederhold, M. D. and Kim, S. I. (2001), An Investigation into Physiological Responses in Virtual Environments: An Objective Measurement of Presence, *in* G. Riva and C. Galimberti (eds.), *Towards CyberPsychology: Mind, Cognitions and Society in the Internet Age*, Amsterdam, IOS Press, pp.175–183.

Witmer, B. and Singer, M. (1998), Measuring presence in virtual environments: A presence questionnaire, *Presence-Teleoperators and Virtual Environments* **7**(3), 225–240.

Yang, X. and Chalmers, A. (2005), Perceptual level of detail for efficient raytracing of complex scenes, *in UK Theory and Practice of Computer Graphics : Eurographics*, pp.1–6.

Yarbus, A. L. (1967), *Eye Movements and Vision*, Plenum Press. Originally published in Russian, 1962.

Youngson, R. M. (1994), *The Guinness Encyclopedia of the Human Being*, Guinness Publishing.

Zahorik, P. and Jenison, R. (1998), Presence as being-in-the-world, *Presence-Teleoperators And Virtual Environments* **7**(1), 78–89.

Zeltzer, D. (1992), Autonomy, Interaction, and Presence, *Presence-Teleoperators and Virtual Environments* **1**(1), 127–132.

Zimmons, P. and Panter, A. (2003), The Influence of Rendering Quality on Presence and Task Performance in a Virtual Environment, *in Proceedings of the IEEE Computer Society*, pp.293–294.

Zimmons, P. M. (2004), The Influence of Lighting Quality on Presence and Task Performance in Virtual Environments, PhD thesis, University of North Carolina.