A SPOTLIGHT BASED INTEREST MANAGEMENT APPROACH FOR DISTRIBUTED VIRTUAL ENVIRONMENTS

being a thesis submitted for the degree of

Doctor of Philosophy

in the University of Hull

by

Ian Dunwell

BSc. (Hons.) Imperial College London

MSc. (Dist.) University of Hull

December 2006

ABSTRACT

A rapidly emerging need within virtual environments is the ability to accommodate a large, distributed user-base. In particular, Internet-based distributed virtual environments represent a challenge in providing a heterogeneous environment accessible by many thousands of users. Interest Management represents the process of filtering network data based on relevance to users, in order to make the most efficient use possible of available network capacity. In this thesis, the potential for a visual-attention based refinement to interest management is explored, which utilises a spotlight-based model of human attention. By considering the relationship between attention and interest management, this thesis illustrates a tendency within existing systems to use a simplistic or object-based approach to measure relevance, typically based on proximity to the user or visibility. By considering the current state of the art alongside common theories of visual attention, it is hypothesised that a refinement aiming to accommodate a spotlight model may offer an approach capable of more efficient use of bandwidth when compared to existing techniques. In particular, it is suggested a spotlight model would be capable of accommodating key aspects of human visual attention in a processor-efficient and composable manner. Bandwidth conserved by the use of such an approach may be subsequently re-utilised to provide a higher degree of interactivity within a distributed virtual environment, or to support greater numbers of simultaneous users.

In order to evaluate the effectiveness of a spotlight-based approach, a test environment is created which allows users to participate in a number of activities alongside a large volume of simulated clients. Throughout these activities, interest management is changed seamlessly between variations of a spotlight approach, extremes, and an approximation of the current state of the art. Results obtained from a total of 15 users indicate a preference for such an approach when compared to the current state of the art in 80% of subjects, and suggest the capacity to reduce available bandwidth by up to 60%, dependent on task, without perceivable impact.

ACKNOWLEDGEMENTS

I would firstly like to thank my supervisor, Dr. John Whelan, for his tireless support and advice throughout all stages of this research. I am particularly grateful for his rigorous proofreading and comments on draft chapters, and the many constructive criticisms that have benefited this thesis, and my skills as a researcher, tremendously.

I would additionally like to extend my thanks to all the staff within the University who have supported this research, through their insight and advice. In particular I would like to thank Dr. Craig Gaskell and Dr. Tanko Ishaya for their help in securing funding and ongoing support for the research. Special thanks must go to all the staff and students at the Scarborough Campus, who have contributed towards my time here being as enjoyable as it has been. Furthermore, I would like to thank my fellow research students for their assistance and patience in testing and debugging the system developed during the research, and for their many helpful suggestions.

Finally, I would like to extend my personal thanks to my friends and family for their constant support throughout my time at the University of Hull.

CONTENTS

FIGURES

1 Introduction

In 1989, Lanier coined the term 'Virtual Reality' [LMF89], to describe the emerging phenomenon of the use of three-dimensional computer graphics to convey scenes mirroring real-world spaces. Rather than simple static images, the processing power of advancing technology was being exploited to allow users to traverse their viewpoint through a virtual space. Whilst early systems were primitive, they benefited enormously from the exponential increases in processing power as hardware evolved, and by the late 1990s the term 'Virtual Reality' was frequently expanded to 'Virtual Environment' (VE) or 'Virtual World', emphasising the increase in scope of systems. Perhaps the most striking aspect of VEs is the rapidity with which graphical technology has advanced, from primitive text-based displays to detailed images of real-world environments, over the space of less than a few decades. Hardware capable of creating detailed virtual environments is available for as little as a few hundred pounds, and well within the reach of many consumers. As technology continues to advance, it is easy to envision the 'Cyberspace' predicted by Gibson [Gib94], wherein many users co-exist within the same virtual space. Indeed, the simultaneous emergence of the Internet and VE technology has led to an inevitable convergence into the notion of the 'distributed' virtual environment (DVE).

To consider the application of such technology for a moment, research has covered a wide range of potential uses. As documented later in this thesis, the first such application of distributed VE technology was to create military simulations [ZS00]. In a commercial context, property ownership and land sales take place in Second Life [Lin06], with participants being offered virtual islands with real-world prices from $1,000. A cursory inspection of online auction sites [Kel04] shows virtual items from a host of online games retailing for real-world money. A theme common to all these applications is the co-existence of large numbers of users within a small virtual space. Whether the area in question is a virtual shopping mall, battlefield, construction site or stadium, the desire to provide interactivity on a grand scale is compelling. This presents one of the problems at the core of VE design; that the number of clients capable of co-existing within a virtual space is limited by the network bandwidth available. *Area of Interest Management* (commonly termed simply as "interest management") is introduced as a means to minimise this bandwidth requirement, and create an architecture capable of accommodating an infinite number of clients by partitioning and scaling network load accordingly. In essence, an interest manager seeks to determine what is relevant to clients within an environment and provide them with information accordingly. Interest management may thus be interpreted as a filter, removing redundant information from the network data stream and thus allowing more capacity for relevant data. Area of interest management stems from the need for scalability, in terms of having several thousand potential users within the same virtual space whilst maintaining an optimum level of service. In this scenario, interest management is needed to identify the information most relevant to each user, and to ensure it is transmitted as and when is necessary. Thus, in any large-scale distributed virtual environment, interest management is a central concern.

As a filtering process, it is easy to draw parallels between interest management and the processes of human visual attention. The work of Broadbent [Bro59], discussed in more depth in Chapter 2, argues human attention serves to filter redundant stimuli, and present us only with the stimuli that are most relevant. Similarly, interest management seeks to filter redundant network data and present each client with only the information most relevant.

1.1 Hypothesis and Objectives

This thesis considers the visual aspects of interest management. It presents a hypothesis stating that the direct application of a 'spotlight' model of visual attention, in place of the current proximity-based state of the art, offers a reduction in bandwidth requirements for large-scale environments with no perceivable impact for users. Such an enhancement is proposed by considering the process by which existing approaches determine relevance, and equating them to theories of human visual attention. In doing so, it becomes evident that existing systems use a simplistic measure of relevance between clients, when contrasted against existing models of human attention.

The primary objectives of this thesis are:

- To provide justification for the consideration of a spotlight-based approach to interest management, by considering the approaches of existing systems alongside models of visual attention, and in light of the current need for improvement.

- To describe as generic as possible a model for applying such a spotlight, and provide subsequent detailing of its implementation.

- To evaluate such an approach, and provide comparison with the current state of the art.

1.2 Thesis Structure

Following this introduction, Chapter 2 presents research into visual attention, illustrating the theories most relevant when considering virtual environments. Chapter 3 discusses more direct examples of the application of these theories in graphical rendering and virtual environment-based scenarios, covering a range of topics from perceptually optimised rendering to the concept of 'presence' in virtual environments. Throughout Chapters 2 and 3, a foundation of theory is presented by which the existing DVE systems and their respective interest management paradigms as presented may be critiqued. A key emergence from this critique is the lack of consideration of visual attention within existing approaches, and the subsequent potential for the application of a spotlight-based approach, whose potential implementation is described in Chapter 4.

Such an approach is suggested to offer realistic gains in the perceived quality of the environment when considering scenarios wherein users seek to interact and observe both distant and nearby clients. Chapter 5 outlines the development of an implementation to meet criteria in employing the perceptual model whilst remaining computationally efficient enough to function within existing distributed virtual environments. Evaluating the proposed system implies the need to simulate an environment with a large number of users, in order to examine fully the effects of various interest management approaches. In Chapter 6, a series of experiments are described which seek to utilise the environment to evaluate the proposal against the current state of the art. Such evaluation offers results indicating the success of the proposed method within the context of a practical implementation, and leads to discussion regarding conclusions and the potential for further perceptual refinements, which is the emphasis of Chapter 7.

2 Human Visual Attention

This chapter provides a focussed taxonomy of research into human visual attention, taking into account the specific relationship between attention and gaze, and the various models that attempt to explain the phenomenon of visual perception. The aim of this chapter is to provide discussion regarding the concepts of gaze and attention, and to present the most relevant models of human attention when considering virtual environments. Such a foundation is used in Chapter 3 to draw parallels between visual attention and interest management as filtering processes. Hence this chapter provides a background by which the applications of perceptual models discussed in Chapter 3 may be more effectively critiqued, and provides a foundation upon which to consider these models of attention and their potential application to interest management paradigms.

Section 2.1 introduces the broad concept of human attention as a research field, and summarises the fundamental work within this area. Attention and gaze are defined more explicitly in Section 2.2, which considers both the physiological and psychological aspects of vision most pertinent to the interactions between humans and virtual environments. Change blindness is used as an illustration of the limitations of perception that the methods in Chapter 3 seek to exploit.

There follows, in Section 2.4, a summary of several models of human attention. It rapidly becomes apparent there is a lack of general consensus in modelling human attention; however, the wide variety of models provides a large number of potential avenues for exploration when considering applications within virtual environments. This potential arises not only from the differences between existing models, but also their varying suitability to computational applications. The latter part of this section concerns itself primarily with computational models of attention, or theories of attention particularly suited to computational modelling. This leads to discussion regarding the methods by which such models have been exploited within virtual environments, which is the focus of Chapter 3.

2.1 Visual Perception

The study of human perception can claim to be one of the most ancient fields within psychology. Origins of established study within the field can easily be traced back as far as Aristotle in works such as *De Sensu et Sensibilibus* ("On Sense and what is Sensed"), written around 350BC as part of his *Parva Naturala*. Even such early works still maintain modern relevance, as the qualitiative and subjective nature of human perception eludes attempts to produce a definitive empirical model. Indeed, the value of accurate quantisation of perceptual stimuli in order to weight their relative magnitudes was identified as of interest even within these early works:

> *"That every sensible object is a magnitude, and that nothing which it is possible to perceive is indivisible, may be thus shown. The distance whence an object could not be seen is indeterminate, but that whence it is visible is determinate." – Aristotle [Bea04]*

Within this chapter, attempts to define and measure these magnitudes of "sensible objects" (and, indeed, what defines a "sensible object") are presented. In the context of this thesis, the ultimate goal of such attempts may be interpreted as being to form a basis upon which to construct a model capable of ubiquitously predicting the perceptual relevance of any given object to an observer. At the core of this notion lies the implied hypothesis that sense is indeed measurable, and the results of the volume of research presented within this chapter add weight to this theory in many different respects. Given its feasibility at least to some degree, the challenge in attaining such measurement lies in the sheer biological complexity of the senses, precluding physical attempts to assertain magnitudes of sense, and the individuality of the human being, which constrains attempts to qualitatively predict magnitudes in a form applicable to every individual.

A common phenomenon which illustrates several key characteristics of perception, particularly the effects of such individuality, is the concept of the ***multistable*** image. In essense, this is a static image which the brain cannot adequately resolve to a single interpretation and hence is percieved as an oscillation between two or more possibilities. The *Necker Cube* shown in Figure 1 is one simple and frequently-used example of multistable perception. Similarly, the adjacent image, *Rubin's Vase* may be perceived either as two faces or a chalice. An individual's perception will tend to shift between the two interpretations in both cases, the brain unable to settle on a 'correct' interpretation of the image. Interestingly, however, since people have a tendency to view objects from above more frequently than from below, Kleine-Horst argues there is a distinct tendency to interpret the cube as being viewed from above [Kle01]. This is perhaps one of the most simple yet compelling examples showing a relationship between past experiences and perception – the way in which humans perceive the world is influenced by learning. Such concepts form the basis for top-down and bottom-up models of perception discussed later in this chapter.

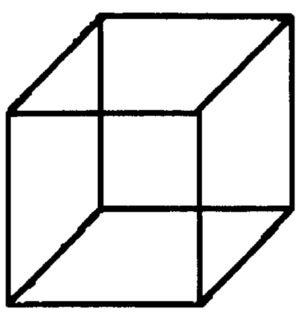
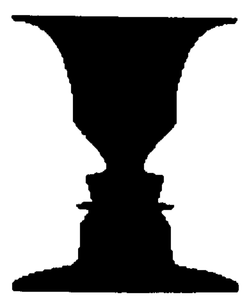



Figure 1: Multistable Perception. A Necker cube (left) and Rubin's Vase (right), both images commonly used to illustrate the phenomenon. Taken from [Leh03a].

Given such differences between humans in percieving situations, and in light of this overwhelming challenge in measuring magnitudes of sense, it may seem that attaining the goal of a ubiquitous model for sensory relevance is an impossibility. Indeed, providing a flawless model capable of providing an accurate and universal measure of perceptual relevance is indeed of debatable feasibility, and certainly only exists within the realms of biological science. However, by qualitative methods and the extreme nature of many stimuli in comparison to their background (a flashing light in a darkened room, for example), it is often possible – if not trivial – to classify the order in which items are likely to evoke sensory priority. The case, then, is to look for ways in which approximations of sensory importance may be made, rather than provide absolutes. This can be evidenced in the many experiments discussed in this chapter that rely on a more qualitiative approach as opposed to strictly quantitative methods.

A commonly described mechanism by which additional detail is constructed by the mind in the absence of physical stimuli is described by Gestalt theory . Gestalt theory can loosely be described as the nature of the mind to create wholes out of incomplete elements, a phenomenon frequently summarised by the phrase *"the whole is greater than the sum of the parts"*.

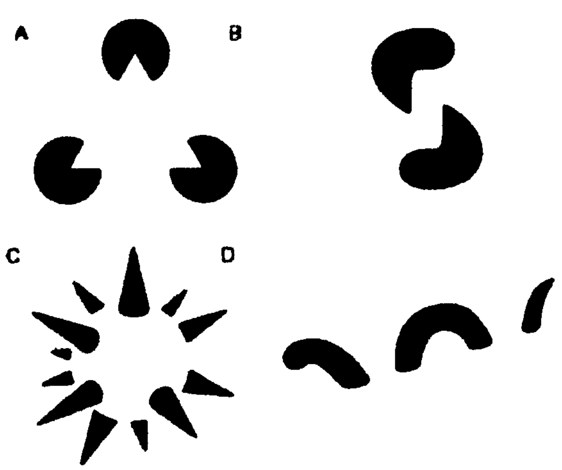


Figure 2: Reification. 4 separate examples of reification. Taken from [Leh03b]

Figure 2 illustrates some examples of images reinforcing this theory, commonly referred to as *reification*. In all four examples, there is a tendency to perceive shapes not present, due to the unusual surrounding shapes. In the first image (A), for example, a triangle is generally observed between the three shapes. This lends itself to a theory of perception as a *constructive* process – the brain will attempt to create a satisfactory interpretation of a scenario, even if such an interpretation requires the formation of elements absent from the actual image. Gestalt theory may be presented as forming a part of the basis behind phenomena discussed later within this chapter, such as change blindness. Rather than risk reiterating the contents of many textbooks serving to explain the origins and detailed theories behind Gestalt psychology, it would instead seem more relevant to provide a number of key references and thus provide more emphasis on the relationship between these fundamental theories and the remainder of this chapter. Thus the reader may be directed to the initial and fundamental works by Wertheimer [Wer12], [Wer20], [Wer22] that led to his accreditation as the founder of Gestalt theory, or any of the excellent textbooks on the subject [DW05]. To return to the earlier quote regarding the whole as the sum of its parts, an address by Wertheimer reiterates this statement more eloquently:

> *"There are wholes, the behaviour of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole. It is the hope of Gestalt theory to determine the nature of such wholes."* – *Max Wertheimer [Wer22b]*

Exploiting this phenomenon is key to any methods seeking to optimise images in terms of minimising content, whilst relying on the brain to subconsciously add detail. This can be seen in the works aiming to perceptually optimise computer graphics described within the next chapter, and forms a partial basis for the concept of *change blindness*, covered later in this chapter. The proposal within this thesis itself relies on such notions in order to manipulate what is perceived as efficiently as possible. Gestalt theory can be regarded as a primary source of questions in perceptual research, particularly with regards to vision as the illustrated examples such as those in Figure 2 demonstrate. Although, whilst Gestalt theory provides a great deal of insight into visual perception from a theoretical standpoint, it can be seen as partly distanced from research into the biological system at the core of visual perception due to the contradictory nature of the whole system as being indefinable in terms of sub-components. Thus in order to view the field more comprehensively, it is necessary to view other works in parallel. Whilst in 1910 Wertheimer was forming the basis for Gestalt theory, Helmholtz was establishing scientific research into the nature of vision in his *Treatise on Physiological Optics* [Hel10], again making crucial observations still studied in current research:

> *"The truth is we seldom use indirect vision anyhow to find out about the form, size and arrangement of objects visible in that way. Its main service consists in supplying us with a sort of rough sketch of the vicinity of the point of fixation where our attention is directed, and at the same time to divert our attention to any new or extraordinary phenomenon that may arise out toward the periphery of the field." – Hermann von Helmholtz [Hel10b]*

The phenomena not only of direct and indirect attention, but also the significant response to objects in the periphery of the field of view, are both concepts recognized in perceptual research and topics of ongoing study. Fundamentally, a recurring theme throughout early research was an appreciation of the complexity of human attention, and the concept of attention as a process incorporating many factors. Having presented such overriding theories behind research into human perception, emphasis is now shifted to more recent and specific works. In the subsequent sections, attempts to analyse images on a perceptual level are described, alongside more recent and detailed models for human attention. Additionally, a brief study of attention and gaze is included, partly in order to provide definition of terms, but also to give some weight towards the biological systems closely related to perception.

2.2 Attention and Gaze

Before analyzing the topics of attention and gaze, it is essential to first define precisely the meaning of these two terms, in their context within research into visual perception. Such definition can be related back to initial research on perception mentioned within the previous section, since the concepts are closely linked with perception itself.

Attention has been a noted subject of research since the 19^{th} Century, and whilst it can be considered as a focus of researchers such as the aforementioned Helmholtz, only recently has technology allowed attention to be studied in a more quantitative fashion. A common expression of attention is as a cognitive process of focusing on an object at the expense of the surroundings. James, often considered a founder of theories regarding attention, described the process in his "Principles of Psychology" [Jam90] thusly:

> *"It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought...It implies withdrawal from some things in order to deal effectively with others." – William James [Jam90]*

One model for attention is described by Broadbent, who seeks to define it in terms of a filter; removing redundant information [Bro59]. In effect, Broadbent suggests attention is a mechanism in place to prevent a form of 'information overload'; suggesting perceived sensory information has the potential to exceed the processing capabilities of the brain. Thus the task of such a filter could be described as working within the constraints of the brain to ensure the most salient information is given priority. In an experiment intending to verify this, Cherry [Che53] equipped volunteers with headphones, providing two different recordings into each ear, and requested that they listen only to one ear and report the information given to them. The results showed that the subject was almost universally unable to report the information given to the other ear, and were even incapable of recalling the language. In effect, it has been proven by such experiments that in the case of aural stimuli, the effectiveness of this filter can be extreme, almost completely negating the stimulus when required. Hence there exists distinct and substantial evidence that humans are capable of focussing the senses, both on a physical and cognitive level.

Gaze, by comparison, is rather simpler to define. Whilst the precise origin of the term is hard to identify, a 1975 taxonomy of eye-tracking technology by Young and Sheena identifies it simply as the physical property of the direction of the eye at a given point in time [You75]. Such a definition has subsequently been adopted by the emerging field of research into tracking gaze [HNL06] and thus it proves reasonable to use the term to indicate the direction of the eye independent of consideration of the cognitive, attention-based process which results in this direction. The metric used to measure gaze varies widely due to the different nature of gaze-tracking technology; many methods seek to quantify it explicitly, for example in radial co-ordinates with origin at the centre of the eye. Eye-tracking hardware has been capable of such measurements for several decades [Cor75]. Furthermore, qualitative models also exist such as that of Pappu and Beardsley [PB98], which seek to represent gaze as a series of states rather than precise metrics. Such an approach exhibits advantages when considering the difficulty of relating a measurement of the angle of the eye to the object being gazed upon – doing so requires a knowledge of the surrounding world and the ability to compute the path between eye and object. By adopting a qualitative method whereby the number of potential targets for gaze is considered finite and well defined, it becomes simpler to predict the current target. Many differences exist between methods for measuring gaze, yet all share a common interpretation of its meaning as the orientation of the eye relative to the remainder of the head. In the remainder of this section, the physiological nature of attention, and hence more specifically gaze is analyzed, in an attempt to provide a more solid understanding of the biological mechanics behind visual attention.

Consider first the physiological aspects of the visual system. At the risk of detracting from the focus of this chapter on more cognitive theories of visual attention, it is worth noting a few important characteristics that are relevant when considering the relationship between physiology and gaze, and hence subsequently between gaze and attention and thus perception. Firstly, there is a marked horizontal and vertical asymmetry within the human eye [SLG06], which offers a potential explanation for the preference towards objects in the lower half of the visual field, as described by Kleine-Horst [Kle01]. Additionally, human field of view offers a notably broader range in the horizontal compared to the vertical, a biological trait also arising from the nature of our surroundings and the relative infrequency of dealing with objects significantly above us, as mentioned in the previous section in the discussion of the Necker Cube. Combining the qualitative results of experiments such as those involving the Necker Cube with a biological analysis of the visual system demonstrate evolved bias in favor of focusing attention upon objects on a lateral plane, with emphasis towards objects below, rather then above the subject. Such a phenomenon is documented by Campbell et al. [CK66]. Another factor of consideration is that the eye's peripheral sensitivity is asymmetric, as shown by Regan and Beverley [RB83]. In essence, by taking a single eye and analyzing responses on each side of the visual field, it can be shown that a bias between left and right sides occurs with increasing eccentricity from the centre of the field of view. This asymmetry exhibits itself most dramatically between nasal and temporal retina (shown in Figure 3) beyond 20 degrees as shown by Sutter and Tran [ST91].

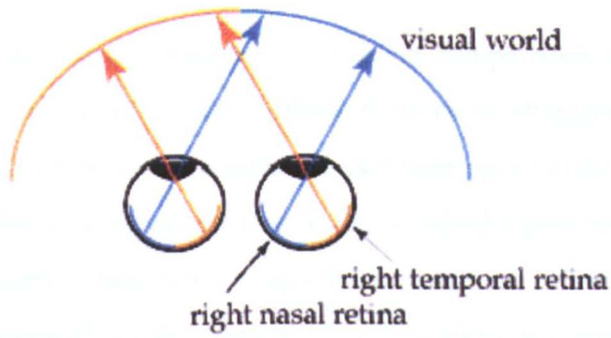


Figure 3: Illustration of the relationship between retina and visual world. Taken from [WUS06]

Increasing eccentricity from the centre of the field of view has been shown to have a number of other effects. One characteristic of such increase is a corresponding decrease in contrast sensitivity. The visual response to contrast at any degree of eccentricity can be predicted by scaling the response with an eccentricity-based factor, as illustrated by Regan and Beverley [RB83]. Attempting to provide a comprehensive model relating physical to perceived magnitudes of stimuli is the focus of the *Weber-Fechner Law* [Web34]. The law states that sensory perception functions in a logarithmic manner, with the ability to sense a change in stimuli being directly affected by the current magnitude of stimuli experienced. This applies to all of the 5 senses - the initial experiments by Weber dealt with handing a blindfolded subject weights and noting a relationship between the amount of weight held and the sensitivity to increases in weight. With regards to vision, one of the principal Weber-Fechner characteristics is a logarithmic response to brightness at all eccentricities. This may trivially be demonstrated by the limited ability of a subject exposed to continual bright light to detect a change in luminosity of a source, in comparison to subject in darkness. More recent research suggests the inflexibility of the Weber-Fechner law to be a weakness; although the widely used Stevens' Formula [Ste55] still implies logarithmic response. Whilst such laws serve to provide some insight into the laws governing human vision, the relationship between light entering the eye and the image perceived by a viewer is understandably complex. Many different characteristics exist, which are, to apply this to the context of this thesis, of note when considering a relationship between both rendered and perceived scenes within a virtual environment.

Though this section aims at this stage to avoid direct comparisons to applications within virtual environments, in order to better present theories of attention, it is worth relating these characteristics of vision to the context of user viewing a virtual environment. Firstly, the horizontal and vertical asymmetries lead to the standard aspect ratio of most monitors, can be seen as considered directly or indirectly by designers of both hardware and user interfaces. More significantly in the context of optimisation, the tendency for humans to give less emphasis to areas above or below them may be exploited in order to allocate more resources to objects on a similar horizontal level. The asymmetry between nasal and temporal retina whilst naturally balanced by the stereoscopic nature of vision, may be exploited in scenarios where hardware may provide differing images to each eye. Finally, contrast sensitivity provides potential for exploitation where pixel colours may be directly manipulated to affect this response. Research into rendering has led to methods such as *high dynamic range (HDR)* lighting within rendered scenes, wherein contrast between pixels is deliberately manipulated to provide a more realistic visual experience [Deb98].

Biological research, whilst potentially capable of ultimately providing a comprehensive understanding of attention, is limited by current technology and the inherent complexity of the visual and neurological system. Hence research into attention has frequently been limited to qualitative study and attempts to provide models that offer an ability to approximate attention using simpler traits. The next section describes a typical phenomena which proves extremely complex when considered purely as a biological process – that of Change Blindness.

2.3 Change Blindness

Change blindness is a noted event whereby large changes in a visual scene go unnoticed by an observer. McConkie et al. are frequently credited with coining the term around 1979 [MZ79], although research directly into the phenomenon can easily be traced back to French in 1953 [Fre53]. However, only within the last few decades has visual display technology advanced to enable experiments on change blindness within scenes to be performed simply and effectively.

Rather than consider the notion of change blindness directly in an attempt to explain the phenomena, many researchers such as Rensink et al. [RRC97] look instead at ***change detection***, effectively viewing the perception of change as the process, with change blindness a failure of the senses to perceive such change. Such research suggests that deliberate attention by subjects is essential for change detection, though not necessarily sufficient. An initial body of research such as that by Grimes [Gri96] suggested that a dramatic change induced during a saccade frequently went unnoticed; resulting in suggestions that there was a relationship purely between saccades and change blindness. However, Simons [Sim96] demonstrates that a central object in a film may be replaced whilst accompanied with a brief pan away from the object, and subsequently go unnoticed by all observers independent of eye movement (whilst no attempt was made to constrain eye movement, results were consistent for a large sample of viewers). Thus more recent work has shown that any interruption between the eye and viewed scene has a significant impact on change detection. Evidence such as this suggests an extremely high degree of importance between continued fixation on an object and the ability to perceive change; Blackmore et al. [BBN95] further reinforce this notion with results showing significantly reduced ability to detect change between two images when a blank screen is briefly inserted between them, alongside the techniques pioneered by Rensink mentioned previously.

Simons [SR05] notes on the subject that the most unexpected characteristic of the phenomena is how drastic change blindness can be, with users unable to detect highly significant scene changes to central objects with only brief obfuscation. Levin et al.

[LMD00] offers an explanation of change blindness as arising from an overestimation of our own ability to process visual information; when faced with experiments presented by Levin, subjects had a tendency to predict a superior ability to detect change than they exhibited. Thus Levin argues our assumption that all visual information we perceive is accurate contributes to our inability to detect change. Only as recently as 2005 was a physiological explanation of change blindness offered [SLW05]; suggesting the parietal cortex has a responsibility in vision separate from the visual cortex. By temporarily disabling the function of the parietal lobe with small electrical currents, subjects became unable to detect changes in a scene even as dramatic as a person's face changing. The importance of the parietal cortex indicates a deeper relationship between scene perception and change involving reasoning and concentration, rather than being a simple visual process.

Whilst not strictly a model of attention, in that it only encompasses a single aspect of the more general field, the concept of change blindness is particularly relevant when considering perceptual optimisation of scenes via the removal of non-relevant content, as discussed in Chapter 3. The importance of change blindness in this context stems from the phenomenon it describes where seemingly drastic changes in a visual scene go undetected by a viewer. This is advantageous when performing such optimisation, since a reliable method of predicting both the triggers of change blindness alongside which details within a scene are most likely to be obscured by the phenomenon, has the potential to allow substantial detail to be removed undetected. In order to accommodate phenomena such as change blindness, which arise from highly complex cognitive processes, attention is frequently modelled via empirical results, rather than as a series of controlled biological processes. The next section goes on to distil this volume of empirical research into visual attention, and identify the most significant models arising from observation.

2.4 Models of Visual Attention

In this section, a number of concepts central to theories of visual attention are presented. Primarily, these concepts centre on attempts to create a conceptual model of attention, capable of predicting the perceptual relevance of objects within an image or environment. Consequently, these concepts form a basis for computational models of attention, which may be used to analyse scenes to detect areas of most likely attentional focus (a process discussed in the next chapter). Whilst study of the eye alone provides some insight into the nature of attention, the visual mechanism is highly dependant on neural processing and the behaviour of the visual cortex, and hence simple mechanical study of the eye only provides a limited understanding of human perception. One key characteristic of the neural aspects of vision is that of *covert* attention. Whilst *overt* attention involves moving the eye to focus on an object (a *saccade*), covert attention may be described as mentally focussing on an object. Covert attention serves as the mechanism whereby humans may detect change within a scene without explicitly focussing the eye upon it. In this case, there is no physical movement by the body to respond to the object, although the user is visually aware of it. This deployment of attention occurs much faster than an overt response, and is preferred for situations such as detection of bright or fast moving objects where overt attention is unnecessary. In comparison, covert attention is incapable of shape detection, which requires a focal response and thus overt attention. However, it was originally assumed covert attention was deployed in order to establish the need for an overt response. These characteristics were identified in a series of experiments by Posner, Snyder, and Davidson [PSD80].

The assumption that covert attention preceeds an overt response is disputed by Findlay et al. [GHF99]. A number of subjects were studied in situations requiring overt attention. The common theme of these tasks was shape recognition, with activities such as reading being used to force subjects to deploy overt attention. Whilst doing so, the rate at which saccades occurred was studied. A conclusion was reached that the rate of saccadic movement, particularly in the case of reading, occurred so frequently covert attention could not, according to the response times obtained by Posner et al., have been deployed sufficiently rapidly prior to each saccade in order to determine the new location on which to focus.

However, further analysis suggested covert attention was still used in order to obtain some information prior to each saccade. Thus Johansson et al. [JWB01] proposed a theory that deployment of overt and covert attention is task-dependant, following experiments showing some simple tasks can be performed without overt attention, yet efficiency increases through its use. Looking more directly at research into covert attention, a variety of models exist, which partly conflict due to a lack of theoretical consensus, as noted by Wood et al. [WCC06]. In the following sections, several models of attention are discussed in order to illustrate both the difficulties in creating a ubiquitous model and the current understanding of the phenomenon. The first two models presented in the subsequent four sections include Broadbent's early notion of attention as a filter, which may be viewed as analogous to the interest management filter applied in virtual environments, and spotlight approaches, which represent attempts to provide an approximation of attention as a spotlight operating within the field of view. The final approaches described include top-down and bottom-up models, and Feature Integration Theory, both of which are used extensively throughout the techniques described in Chapter 3.

2.4.1 Attention as a Filter

Some of the earliest works seeking to quantify attentional responses to stimuli, such as the dichotic listening test of Cherry mentioned in section 2.2, alongside Broadbent's research [Bro59], define attention as a filter. As previously noted, such models are based on the underlying assumption the brain has a finite ability to process stimuli, and thus a filter is required to determine the most relevant information within this. Models centred around this notion seek to determine the characteristics of this filter, with much empirical work placing the emphasis on attempting to measure the limitations of perceived versus actual stimuli in an attempt to quantify the filter.

Considering the complex visual case, it can be simply demonstrated that a finite capacity for visual processing exists. Referring back to the previously cited works of Helmholtz, a variety of simple experiments can illustrate that the ability to interpret visual information is limited - not all visual information reaching the retina is consciously observed. Helmholtz illustrated this phenomena by simultaneously generating sparks within a subject's field of view, demonstrating that limitations exist on the number of events observable based on characteristics such as the duration and distance from the centre of the field of view (the *foveal angle*). The results are also biased by a user's expectations and experience within the experiment, indicating higher cognitive processes being fundamental to perception. Whilst the filter-based model represents the overriding view of attention within the early half of the 20th Century and the majority of research, it ultimately proves inadequate to describe the entire process of attention. The large volume of contributory factors in response to a stimuli, coupled with the biological complexity of the brain, result in many difficulties expressing the filtering mechanism definitively. Of particular note when considering this model is the difficulty of integrating task-driven behaviour alongside a model seeking to determine stimuli in terms of absolute effects on perception – to give an example, a viewer actively searching a scene for a given object will perceive it differently to a subject simply shown the scene. Additionally, the difference in subjects makes a ubiquitous model hard to define – such discrepancies are shown to effect the success of a model in the work of Wedell [Wed94]. Ludwig and Lachnit [LL04] additionally suggest perceptual characteristics may be modified by training and experience, further complicating the situation. The Stroop

Effect [Str35] shown in Figure 4 demonstrates the strength of the impact learning and experience can have on initial perception. An observer instructed not to read the words but simply list the colours they are printed in typically experiences a first inclination to list the words and thus answers incorrectly. Stroop argues this indicates a conflict between the basic property of an object, and its learnt meaning.

red

yellow

green

blue

red

blue

yellow

green

blue

red

Figure 4: The Stroop Effect (Adapted from [Str35b]).

This effect of learning, experience, and task on attention is often described as a "top-down" effect (described in more detail in Section 2.4.3). It suggests an approach seeking to provide a ubiquitous model of attention as a filter of basic stimuli fails to encapsulate the many high level variables contributing towards the perception of a scene. However, the paradigm of attention as a filter may be extended to incorporate a wide variety of models that aim to represent sub-sections of the attentive process, including top-down aspects. With specific regards to vision, one of the more prolific models is as attention as a spotlight, discussed in the next section. In general, whilst the filter model has been repeatedly shown to be an oversimplification by a variety of examples such as the Stroop effect, it remains an overriding theme with considerable validity as a means of expressing the nature and role of attention.

2.4.2 Attention as a Spotlight

A frequent metaphor for covert attention is as a 'spotlight' of attention, illuminating a section of the view sphere. This concept was first proposed by Eriksen and Hoffman [EH72], [EH73], who attempted to identify the size of the spotlight in a number of experiments. The outcome of these experiments was a conclusion the spotlight could vary in size, but encompasses at least a single degree in the field of vision. Response time for covert attention, again drawing on the results of Posner et al., is dependant on the task and decreases if the person is given prior awareness of an object's location (such as seeing it move behind another object), but is typically in the order of several hundred milliseconds. Eriksen and St. James go on to refine the spotlight model to a 'zoom lens' model [ES86], in an attempt to better quantify the varying size of the covert attention spotlight; effectively stating attention can be focused in high detail on a small area, or on a larger area with coarser resolution. In this case, the spotlight is considered behaving exactly as a mechanical zoom lens (Figure 5). This model is consistent with a number of experimental results, which show a significant increase in the ability to perceive change within a scene if a user has no specific focus, such of those of Most et al. [MSS00]. This particular study also seeks to analyse the relationship between distance from the spotlight 'beam' and level of change blindness; finding, however, that the relationship appears to relate directly (and more simply) to the distance from the point of fixation.

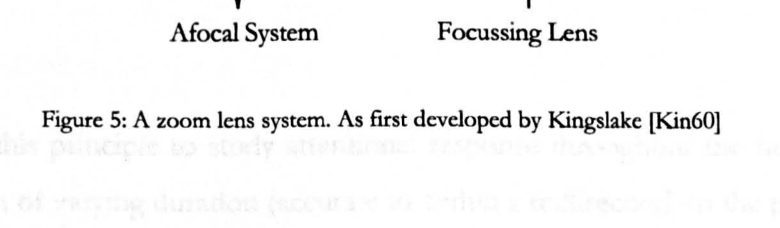


Figure 5: A zoom lens system. As first developed by Kingslake [Kin60]

In a zoom lens, motion of a central, concave lens and convex lens (shown leftmost in the above illustration), results in variable levels of magnification, dependant on the refractive indices and curvatures of the lenses, and their relative distances. Eriksen and St. James apply this to covert attention by stating a similar, psychological mechanism varies the size of the spotlight. If attention may thus be treated as a spotlight projected from the viewer, with a fixed diameter of focus, it would relate well to the rendering model common to many virtual environments, which renders each frame from a given user viewpoint.

The aforementioned research of Posner et al [PSD80] differs from Eriksen and St. James also in its interpretation of the movement of this spotlight of attention. Whilst Eriksen and St. James propose an analogue model, Posner et al. suggest quantised behavior. Driver and Baylis [DB89] consider more fully other aspects of the visual field, such as shape and colour, and argue the spotlight model breaks down since it views the visual field as an image, rather than a mental approximation of the world which takes into account other, top-down attentive features. Similarly, Duncan argues movement is of significance, adding further weight to the notion of the perceived field of view as a spatiotemporal surface [Dun84], which may not be adequately approximated as a static image. Tse et al. [TSL03] seek to resolve these arguments by performing a number of experiments that exploit the notion of *change blindness*. The notion follows from Rensink's [Ren02] introduction of the concept of attention as being focused only on a single object at a time, building upon a substantial body of existing research around the notion of change blindness [RRC97]. Effectively, Rensink argues human vision consists of a stored, mental approximation of any given scene, with individual objects updated only when they become the focus of attention. Since this focus is restricted to one item at a time, it follows a human is more likely to be oblivious to change when focusing intently on a given object. Tasks seemingly requiring multiple targets of attention (multitasking, as described by Gillie et al. [GB89]) such as juggling can be argued to be simply a sequence of rapid changes in attention [Ren00].

Tse et al. use this principle to study attentional response throughout the field of view by inducing a flash of varying duration (accurate to within a millisecond) in the peripheral field of view, and measuring detection rates between observers [TSL03]. The experiment is based around the notion inherent to change blindness that attention has to be focussed upon a change in order for a change to be detected. Users were instructed to observe an array of red and green cells. A white spot (flash) was then shown in one of four positions (25 degrees up, down, left or right of the centre of the field of view). A 47ms blank image then induced change blindness, and was followed by a final image with a random cell shaded white. The observer was tasked with maintaining fixation throughout the 47ms blank and determining whether the cell shaded white was previously red or green. These results are illustrated in Figure 6, and provide an illustration of the field of covert attention.

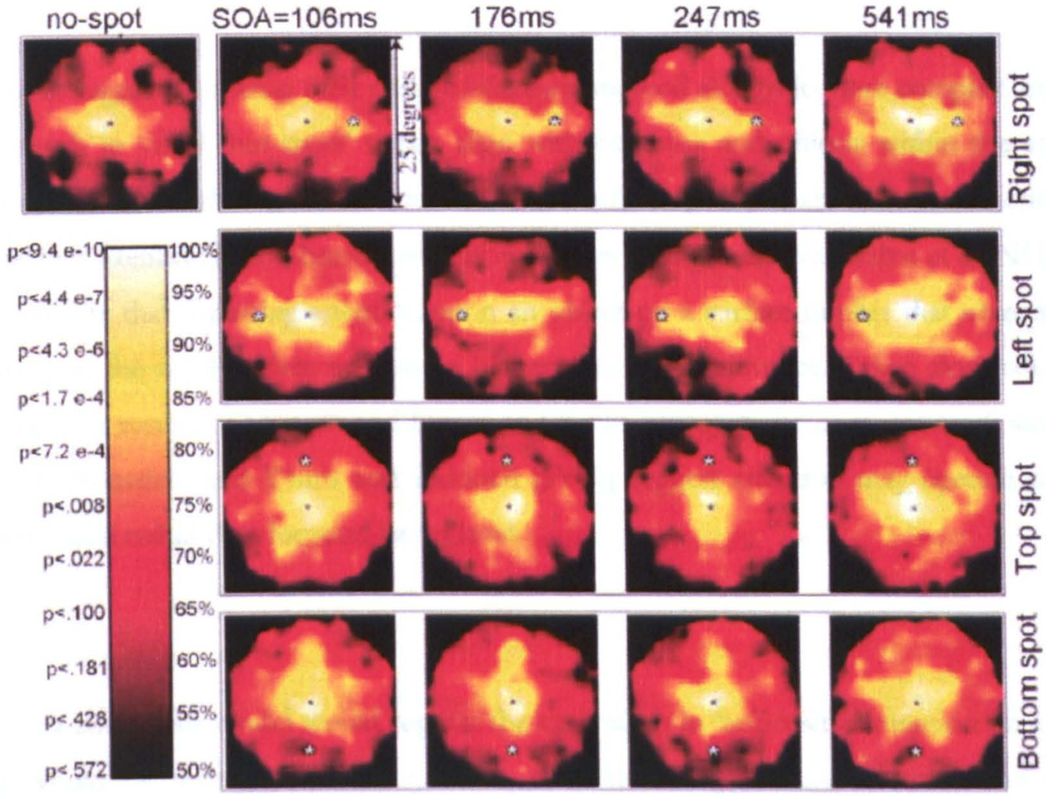


Figure 6: Attentional Mapping. Attentional maps for three observers, taken from Tse et al. [TSL03].

The initial no-spot image shown in the top left of Figure 6 may be seen as the 'default' distribution of attention for the task in the absence of stimulus. Flash durations are illustrated against response success rates for spots appearing in 4 points within the field of view. The bar on the left shows the probability of response occurring by chance (p) alongside the indication of success rate corresponding to the colour.

Tse notes that "After the cue flashed, the hot spot elongated significantly." ([TSL03] p. 7) indicating that a stimulus causes an abrupt redistribution of attention to accommodate the stimulus. Of further note was an inclination for attention to focus upon an axis – if drawn to the left dot, for example, the region elongates not only towards the dot on the left but also towards the right hand side. Thus a degree of symmetry is suggested with user's focusing on either a vertical or horizontal region. Accuracy was also shown to be lower on the vertical than horizontal axis, again reinforcing the theory that human perception has evolved to prioritise objects on the same vertical level suggested at the beginning of this chapter.

Tse et al. suggest that using their change-blindness method alongside more complex images may result in visual 'hot-spots' which align with contours and objects within the scene. This contradicts earlier work such as that of Broadbent [Bro82], and the filter model by suggesting attention is not distributed uniformly over a wide area. Nakayama et al [NHS95] demonstrate that focussing attention upon an object on a surface implies that objects co-existing on the same surface will have a higher attentional relevance. Taken alongside one another, the results of Duncan and Driver and Baylis, alongside other recent work such as that of Shomstein and Yantis and Riddoch [SY02] suggest attention may operate on an object-based rather than location-based level.

Whilst a zoom lens model would suggest the time taken in 'refocusing' attention on a new area would be related to distance, Kwak et al [KDE91] indicate attention shifts are instantaneous (to the extent of all measuring apparatus), regardless of the distance between regions of attention. Eysenck et al. [EK90] also succeed in demonstrating an inadequacy in the zoom lens metaphor by means of an experiment providing stimuli in a series of concentric rings. With attention deliberately focussed on a middle ring, a spotlight model would imply a bias towards the detection of stimuli within an inner ring (see Figure 7) over the outer region. However, with the participant focussed on the middle (grey) ring, they were less aware of stimuli in the central ring than a circular spotlight would imply. Hence, a circular spotlight was shown not to be the case in this situation, leading to the conclusion that the shape of the spotlight may potentially by dynamic. Despite such results, the method shows the capacity to approximate attention effectively, as illustrated by the initial work of Eriksen and Hoffman [EH73].

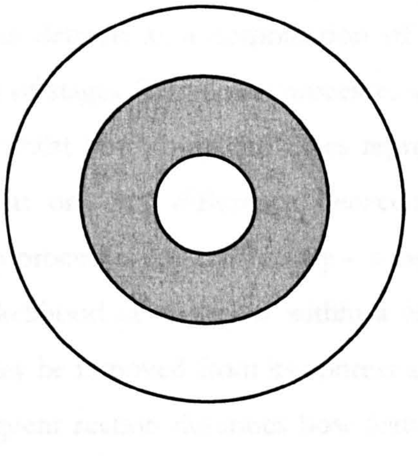


Figure 7: Limitations of the Spotlight Model.

In summary, the spotlight model, as proposed by Eriksen and Hoffman, proves an approximation of attention, with certain situational weaknesses. The experiment by Tse et al. reinforces the notion of a predominately circular spotlight, with certain minor deviances such as the tendency towards a slightly elliptical form due to a predisposition towards objects offset horizontally from the centre of the field of view, compared to those offset vertically. Whilst this model can be shown to break down when a subject is instructed to deliberately focus their attention on a given area (and, thus, may be equally prone to naturally breaking down when an individual specifically engages their attention upon a visual search task). However, in the majority of experiments such as those by Eriksen and Hoffman or Tse et al., approximately circular behaviour is observed, with the phenomena in Figure 7 being a deviation from the norm. The basis for this deviation is the specific relationship between task and attention, from which emerges the notion of top-down attention, described in the next section.

2.4.3 Multimodal Attention

A more complex view of attention is as a multimodal process, well defined by Pashler et al. [Pas98], in which attention is defined as a combination of both *top-down* and *bottom-up* processes, invoked in a series of stages. Top-down processes are driven by a subject's task, knowledge, and experience, whilst bottom-up processes represent more primal responses to stimuli. Pashler notes that one key difference between bottom-up and top-down processes are that bottom-up processes are involuntary – a fact of considerable use in any attempts to determine the likelihood of an object within a virtual environment being the focus of attention, since it may be removed from its context and viewed on the basis of its properties alone. The subsequent section describes how feature integration theory utilises this separation in order to identify a series of criteria with which to establish a "saliency map" – in essence a map of values representing perceptual relevance. Within a multimodal approach, bottom-up components comprise attributes such as colour and orientation, whilst facilities also exist for top-down analysis specified directly to meet the needs of a given application. *Bottom-up* processes are hence commonly defined as natural reflexes; if, for example, something moves quickly towards a human, it becomes the focus of attention regardless of context. *Top-down* processes represent a more cognitive level of processing, and allow the use of experience and knowledge to determine the focus of attention – to give another example, an air traffic controller may focus very intently on a screen which may have little attentional relevance to an untrained passer-by. Whilst models of attention as top-down or bottom-up processes are in many respects complementary, their separate nature allows for individual discussion of each notion.

Research into bottom-up attention focuses on natural, biological responses to stimuli. Innate responses to stimuli are evident throughout the natural world, from the reaction of pulling back a hand that brushes against a hot surface to the path of moths to a flame. Restricting ourselves again to the visual case, characteristics likely to evoke a bottom-up response as taxonomised by Thewes [The93] include:

- *Colour*, and more specifically the contrast of an object with surrounding colour;

- *Relative motion*; with objects near the periphery of the field of view being particularly emphasised, as shown by Poggel et al. [PSM05]. Rapidly moving objects in the periphery of the field of view evoke a particularly strong attentional response; in the case of the work of Poggel et al. overriding a central cue in many situations. This is likely an evolutionary response to perceived danger from fast approaching predators;

- *Size*; where width and height (or any dimension) may be considered independently;

- *Edges*; being defined as continuous regions with a high contrast to their surroundings;

- *Brightness;* With a logarithmic scale of response to intensity as described by the aforementioned Weber-Fechner law (Section 2.2).

The earlier research of Yarbus [Yar67] and more recently Hillstrom and Yantis [HY94] illustrates a particular bias towards changes in colour and sudden movements when evoking attentional response. Bottom-up models have the advantage of expressing attention in terms of external properties of objects, and thus being simple to model. This is of particular interest with respect to virtual environments, since such environments are typically represented in a way which easily maps to bottom-up models, as a series of objects with defined motion, colours and geometries (from which may be inferred size and edges). Subsequently bottom-up models have been applied to a variety of virtual environment applications – Peters and Sullivan utilise a bottom-up approach in determining the gaze direction of "virtual humans" [PS03]. In this case, a virtual environment is populated with computer-controlled occupants whose direction of gaze is controlled by a bottom-up analysis of their surroundings. The aim is to provide a more natural and realistic movement from these virtual humans, and has potential applications when creating believable virtual agents. Work such as that of Le Meur et al. [LLB06] has sought to unify and develop bottom-up models of attention into a comprehensive computational model. A comprehensive model would offer many benefits, yet it remains important to note that bottom-up models alone fail to accommodate the full process of attention – this requires integration with top-down techniques.

Whilst bottom-up attention seeks to extract primitive features and thus provide a data-driven means for predicting saliency, primitive features alone cannot represent the sole criteria by which objects draw attention. Consider, for example, the case of an artist reaching for a coloured pencil in a jar full of differently coloured pencils when creating a sketch. Whilst certain colours of pencils in the jar may visually appear more striking, the attentional search is ultimately focussed on selecting a specific pencil to suit the current goal. Research such as that of Corbetta and Shulman [CS02] and Wolfe et al [WHK04] reinforces the notion that a goal-driven aspect to attention exists. This aspect of attention forms the basis for top-down models of attention.

Top-down models are frequently expressed as "driving" the bottom-up processes [CEY04]. Walther et al. suggest top-down processing can be shown to take in the region of 20ms longer to invoke [WFK06]; furthermore an experiment series suggests the time to deploy top-down attention remains constant for an individual, regardless of task. This indication of top-down attention as a more costly neural process corresponds to greater difficulties and complexities in creating adequate models, compared to bottom-up attention. In particular, since top-down processes are driven by goals and experience, these goals and experience must be known, measured or approximated by any model before it can attempt top-down processing. This is subsequently reflected in the relatively limited number of attempts to implement top-down models of attention in a computational context.

This said, providing the facility exists to determine the goals of individuals prior to the task, it is possible to attempt to model the effects of top-down attention. As research suggests top-down processes are used to drive bottom-up responses – for example the learnt fear of a predator increases bottom-up response to sudden movement or certain colours [CEY04] – multimodal attention is frequently modelled as a series of layers, with top-down processes occupying the higher levels. Applications include robotics, where such models are applied to provide more realistic behaviour and increased functionality. [HHF05].

2.4.4 Feature Integration Theory

Feature integration theory, as defined by Treisman and Gelade in 1980 [TG80], has proven one of the most compelling and influential theories of perception. Feature integration theory suggests the capability of humans to identify objects as a collection of stimuli – an apple, for example, as an integration of physical object, visual object, taste, and smell. Valdes-Sosa et al. [VBR98] describe the ability to perceive objects in such a fashion to imply the ability to compound multiple characteristics into a whole. The theory itself is complex, and beyond the scope of this chapter, although of interest are the specific implications for visual attention.

In the case of a model focussed purely on visual processing, Treisman suggests the visual system builds separate feature maps when viewing a scene. The nature of these feature maps may be related back to the innate characteristics of scenes naturally afforded attentional relevance (as defined by bottom up attention) and higher cognitive recognition of features (from top-down models). Feature maps are then combined into a saliency map. The concept of saliency is taken literally to mean a map representing the relevance of a visual scene. Treisman goes on to suggest a relationship between these maps and cognitive search processes; in effect theorising that the brain holds a saliency map of the visual scene and then performs search tasks upon it. Two differing types of visual search are defined within the model, feature search, and conjunction search. Feature search is a rapid, covert search based on primitive features of objects within the visual field, and lies at the core of out ability to recognise and track objects moving within the field of view. The process is summarised by Treisman as follows:

"We assume that the visual scene is initially coded along a number of separable dimensions, such as color, orientation, spatial frequency, brightness, direction of movement. In order to recombine these separate representations and to ensure the correct synthesis of features for each object in a complex display, stimulus locations are processed serially with focal attention. Any features which are present in the same central "fixation" of attention are combined to form a single object. Thus focal attention provides the "glue" which integrates the initially separable features into unitary objects." – Treisman [TG80]

Conjunction search by comparison represents the ability to identify a substantially more complex object, and is much more time-consuming – an example would be searching through a crowd of faces within a stadium to identify an individual. This forms a basis for the notion of multimodal attention, as discussed in the previous section.

A particularly noteworthy aspect of Treisman's approach is the suitability to computational modelling and image processing. Introducing the notion of *saliency* values – numerical measures of the likelihood of a region of an image to evoke an attentional response - the concept of the *saliency map* is formed. Saliency maps are commonly rendered as greyscale images, with pixel colours corresponding to saliency values (illustrated in the next chapter). By comparing the map with the original image, the most salient regions may be identified. Walther et al. [WIR02] demonstrate such an approach, analysing regions of images based upon their contrast to the surroundings, and deviation from the horizontal centre of the image. The results of such an approach for a single image are shown in Figure 8. The saliency map illustrates the regions given highest sensory priority (and hence those in which change detection is most likely to occur) as lighter shades of grey, with the most salient areas shaded white.

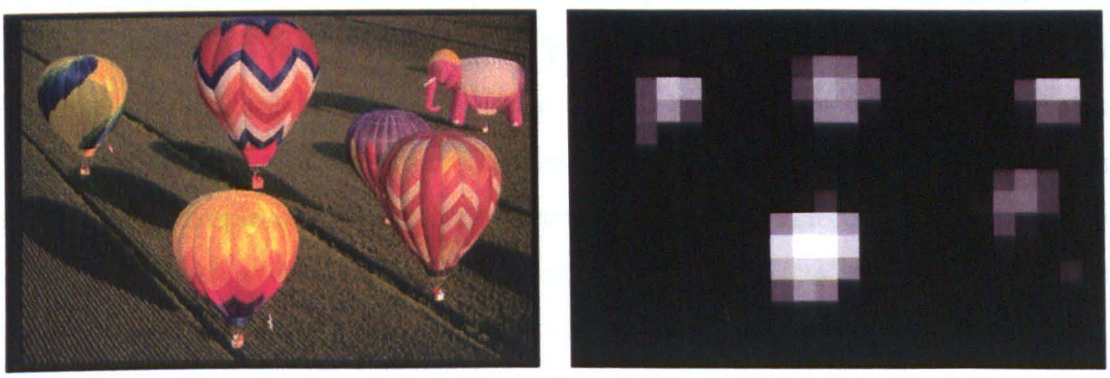


Figure 8: Original image (left) and saliency map (right). Taken from [WIR02].

A possible interpretation of Feature Integration Theory is in providing a link between focus (and thus gaze) and attention. In Treisman's description of focal attention as the "glue" providing the information used for feature integration, it can be seen that models such as the aforementioned spotlight model, may work in tandem with bottom-up interpretations of object properties. Itti [Itt05] notes that feature integration theory itself is the foundation for all computation models of attention. The potential for practical applications of computational models is demonstrated by Itti and Koch [IK99], in developing methods to analyse photographic images to autonomously detect military targets. Computational models of attention primarily offer applications in the ability to autonomously process large (or large numbers of) images. Whilst having an initially militaristic use in target detection, this can be applied directly to a render of a frame within a virtual environment, and provides the potential for identifying likely objects of focus.

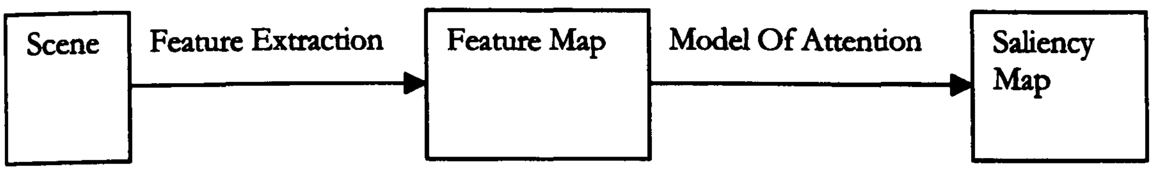


Figure 9: A Computational Model of Visual Attention

Figure 9 illustrates the most common approach to computationally modelling attention, with the clear parallel to feature integration theory made evident by the central use of feature mapping. A source image is first processed by a feature extraction algorithm. Feature extraction itself involves converting the raw pixel data of the image to create an abstract representation of the scene. The method used may vary significantly; either being based around simple image-processing filters such as edge detection [Can86], or more complex approaches which consider a number of linked effects in an attempt to more accurately model the phenomena inherent to the human visual system. The second stage in this model involves the direct application of a model of attention to the feature map, in order to create a resultant saliency map. The saliency values within the map may be interpreted as being proportional to the likelihood that an observer will fixate upon a point in the scene.

The process is spatiotemporal in nature; in human vision motion is a crucial aspect of attentional focus, and hence the incorporation of methods that consider the evolution of an image over time offer more accurate results. Many attempts to generate saliency maps focus on the bottom-up characteristics of objects in order to determine which areas within a scene are likely to attract involuntary attention. Yee et al. [YPG01] propose such an approach which analyses rendered frames, taking into account earlier frames in order to consider temporal effects and generate a saliency map. Figure 10 illustrates the results of this approach for an individual frame. Shown are the original three-dimensional scene, rendered from a given viewpoint, and defined saliency map The object on the left of the scene, which moves, generates a higher resulting saliency as can be evidenced from its prominence on the saliency map. Furthermore, the background image, through its complexity, evokes a higher priority than the surrounding wall.

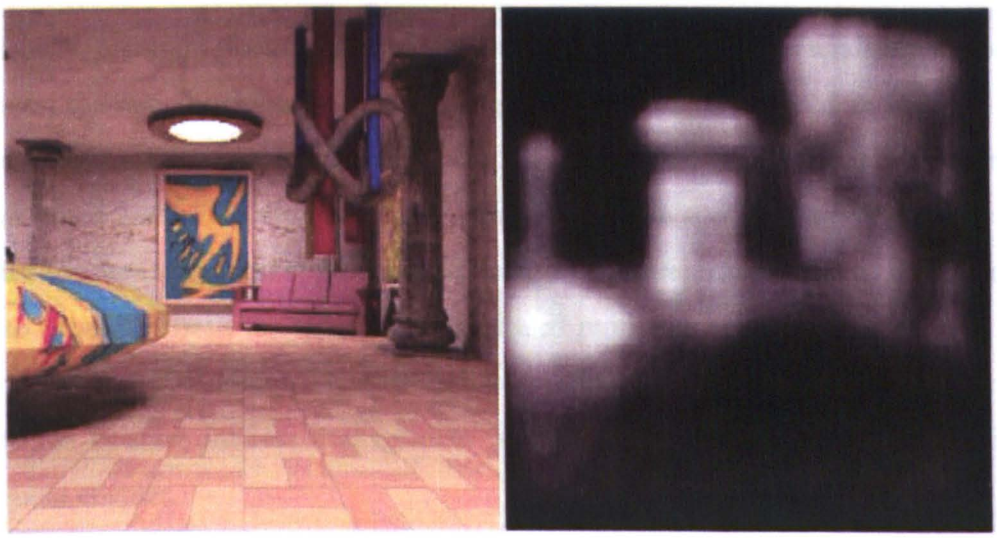


Figure 10: Saliency Mapping using the method applied by Yee et al. [YPG01].

Saliency mapping has several main applications. A primary role is within the automation of image processing, reducing the need for humans to analyse large volumes of data by hand. The previously mentioned military application falls into this category. Ho et al. [HCP03] describe a framework to detect a user's focus, and go on to apply it to digitally watermark media in the least salient areas, to minimize the impact the watermark has on the perceived quality. Chapter 3 focuses on the application most relevant to virtual environments – increasing performance by filtering out unnecessary details and better allocating resources.

2.5 Summary

In this chapter, a background summary of theories regarding human visual attention has been presented. Primarily, an emphasis has been given to demonstrating the ability to model attention as a filter, and theories arising from this assumption (and to address its limitations). One such model, the spotlight approach, seeks to approximate attention by comparison to a zoom-lens. Hence the notion of attention as being focussed around a point is established. Experimental evidence reinforcing attention as a focussed entity has been illustrated through examples such as the research of Tse et al. [TSL03].

Additionally, common phenomena such as change blindness, and the tendency to consider objects on the same horizontal plane with higher relevance, have been demonstrated. Such factors remain a consideration when developing approaches within virtual environments, and are noted within the next chapter. Alternatives to the spotlight model, such as more complex feature-integration approaches, have also been presented. Both of these methods serve to provide a basis for visual-attention based rendering approaches, discussed in the subsequent chapter. Additional factors such as top-down and bottom-up notions also form a basis for existing interest management approaches, and are again discussed in the subsequent chapter.

The next chapter hence utilises the models presented within this chapter in order to consider their application within virtual environments. In particular, an emphasis is placed on the relationship between attention as a filter, and interest management as a filter. Thus whilst this chapter has provided a background into general theories of visual attention, the next chapter includes case studies of applications within virtual environments in order to emphasise current limitations and assumptions.

3 Visual Attention and Virtual Environments

In this chapter, the relationship between the research fields of human attention and virtual environments is presented. Having presented in Chapter 2 a series of models of visual attention, this chapter emphasises the link, both explicit and implicit, between human attention and virtual environments. Section 3.1 seeks to reinforce the relationship between perception and virtual environments in a general sense, presenting research that demonstrates a clear relationship between theories of human perception and virtual environments, in defining such notions as *presence* and *believability*. Hence it rapidly becomes established that theories of perception based around real-world observations, such as those in Chapter 2, may be applied to virtual scenarios.

Such applications form the basis for much of the remainder of the chapter. Section 3.2 details the use of visual attention in rendering three-dimensional scenes (a task common to all virtual environments). Interest management is presented as a similar application of perception to virtual environments; as effectively a filtering task that serves to present a user the most believable environment possible. It aims to do so by presenting the most relevant network data with the highest quality available. Doing so allows the creation of a more believable simulation, wherein objects and avatars that are most relevant to the user exhibit smooth and uninterrupted motion. Thus, all methods must, to some extent, make assumptions regarding the perceptual traits of a user. Whilst many methods fail to recognise an explicit link to theories of attention, all methods can be hence related to human attention. Section 3.3.1 serves to express this notion more thoroughly, and emphasise the link to perceptual theories within Chapter 2.

Section 3.3.2 provides a more in-depth analysis of existing methods of interest management. Subsequently, the chapter concludes by discussing the limitations of such methods, and offering suggestions as to how a more thorough inclusion of perceptual models may offer benefits. Arising from this discussion is the unexplored potential of a spotlight-based approach, which forms the basis for the remainder of this thesis.

3.1 Perception and Believability

"Designers of all stripes must remember that perception and reality interact in complex ways, and cyberspace is an archetypal example of the phenomenon. " – Curtis Frye [Fry06]

Experimental psychologists have long been using computer technology to explore the nature of human perception – the technology offers the ability to rapidly present users with images and study response times with much greater ease than before its advent. In comparison, only within the last few decades have computer scientists exploited research into perception, seeking to gain performance by optimising systems to the perceptual characteristics of the user. This can be seen to be a consequence of the recent and rapid evolution of computer graphics from providing simplistic two-dimensional text displays to creating detailed three-dimensional worlds, with multimedia content. Applications of such environments are extensive, yet they are universally constrained by limits on rendering speed, and in the case of distributed virtual environments, network bandwidth.

With respect to both these limitations, the exploitation of human perception can be seen to provide an obvious route for optimisation. Whilst content may be highly detailed, in reality a user's visual attention will only be focussed on a small area of the display at any given time, and thus visual attention may be applied as a filter, removing or simplifying less relevant areas and improving performance. One such example discussed later in this chapter is the work of Reddy [Red97], optimising three-dimensional geometry by simplifying perceptually irrelevant meshes. The link between such methods, and interest management approaches, which are, in effect, a similar filtering mechanism with more clearly defined parameters, is emphasised. In both cases, the goal is to improve performance whilst ensuring the user's experience remains constant. Defining the criteria that make up this user experience is another psychological question, and works into the notion of 'believability' within a virtual environment.

A long-term goal in virtual environment development is the creation of environments sufficiently sophisticated to be comparable to the real-world. Towards this end, understanding what characteristics of a simulation affect its perceived realism is of obvious value. Kim et al. [KDE00] utilise the term 'believability' to define a two-layered model in an attempt to classify the various aspects of virtual environments of greatest concern when creating a real-world simulation. In this model, the first layer consists of three fundamental aspects of the system – immersion, presentation, and interaction. *Immersion* deals with the relationship between human sensory input and system output. The further human senses are exposed to external stimuli, such as background noise, the lower the level of immersion and hence the less believable the environment. This also applies heavily to visual and tactile stimuli; whilst a monitor-keyboard-mouse arrangement is certainly sufficient to allow interactivity within a virtual environment, it remains inadequate in terms of immersion. The next concept, *presentation,* refers to the content of the environment. A user is far more likely to find an environment believable that includes accurately sampled sounds, high-definition, photorealistic texturing and complex geometry. A simple out-of-place texture or inappropriate sound can have catastrophic effects for believability, as metrics of immersion such as the performance indications and cognitive surveys applied by Pausch et al. [PPW97] demonstrate.

The final notion in the first layer of the model proposed by Kim et al., *interaction*, applies to the level of interactivity permitted both on a hardware level, but also with respect to the capabilities of the environment. Environments frequently implement real-world physics such as those provided by the HAVOK engine [Hav06], since the increased immersion provided by a Newtonian model of action-reaction is substantial. Typically, users expect objects to collide and behave in a realistic fashion, and the static geometry of early environments is inadequate in this respect. Of further note with regards to multi-user environments is the extent of user-to-user interaction; typically such interactions are both the most important and complex for an environment to accommodate, due to the degree of chaos in their nature (e.g. users may attempt to simultaneously interact with the same object, or converge on a location), and also their complexity. Human avatars themselves provide a challenge in creating believable geometry and motion, as Reitsma and Pollard [RP03] highlight in developing metrics for believable motion.

This leads to the second layer of the model proposed by Kim et al., which emphasises the importance of emotion and personality in providing believability. Accommodating such aspects of human behaviour requires a fine degree of control over the animation and behaviour of avatars within the virtual world, both in terms of their visual representation and the means in which the avatar interprets the behaviour of its human controller to mimic their emotive state correctly. Existing systems adopt a number of novel approaches in order to achieve this, with a limited degree of success, including analysing text typed in by the user [OPI05], whereby certain words evoke a facial response, to allowing the user to explicitly define their mood and configuring their avatar's animations accordingly [DW97].

An additional recurring factor that has a significant impact on believability is *consistency*. In a study on immersion Robertson et al. [RCD97] looked specifically at the relationship between user and standard desktop PC as an interface for virtual reality, and compared it to head-tracked systems. Robertson claims "immersion should not be equated with the use of head-mounted displays: mental and emotional immersion does take place, independent of visual or perceptual immersion", an opinion reinforced by Csikszentmihalyi and Kubey [CK81]. Thus a further discretisation of the concept of believability between psychological and perceptual levels is identified. The role consistency plays is one of 'drawing in' the user, such that they experience a perceptual shift between simply viewing the screen and existing within the environment. Breaks in consistency, such as those induced by low framerate, or discontinuities in world content, are shown to have a significant negative impact on immersion. As a final consideration, the work of Zeltzer et al. [Zel92] expands upon the concept of *presence*. This is often defined as a state wherein users *believe* the virtual world. Hence the previously mentioned factors all contribute towards achieving presence, which can easily be described as an ultimate goal for virtual environment designers – as evidenced by the virtual environment journal that shares its name with the term.

All of this work reinforces the assumption that real-world studies of perception and attention may be employed, to at least some degree, within a virtual context. In the previous chapter, the concept of feature integration theory and several models of attention were discussed. This next section now goes on to present further evidence for this relationship between perceptual theories and virtual environments, by describing the application of such models of attention and their potential benefits within three-dimensional rendering. In particular, theories that seek to define saliency maps via image processing offer a means to optimise the rendering process. Thus the emphasis of the following section is in detailing approaches intending to exploit visual attention to accelerate the rendering process.

3.2 Visual-Attention based Rendering

In the case of interactive three-dimensional environments, the classical approach to three-dimensional rendering forms the basis for optimisation methods seeking to exploit visual attention. In this approach, objects are represented as a composite of vertices and edges, which may be simply rendered as a *wireframe* image by drawing the edges between vertices following a transform of the vertex locations according to view position. Wireframes for a collection of simple objects are shown in Figure 11.

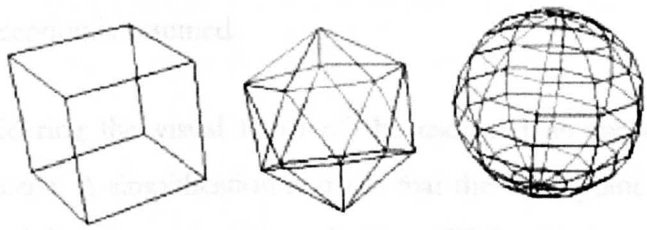


Figure 11: Wireframe renderings of (from left) a cube, icosahedron, and sphere approximation. Taken from [WF06].

Such an approach has limitations with respect to curved and continuous surfaces, which must be approximated as shown in the case of the sphere, but offers a standardized model which may easily be supported by hardware engineered to perform high-speed triangle rendering (since all meshes may be subdivided into triangle strips). Incorporating perceptual optimization within computer graphics implies firstly creating a perceptual mechanism capable of selectively discriminating between objects in the world, and secondly providing a means to translate such selection into polygon structure and level of detail.

Funkhouser and Séquin [FS93] illustrated one of the earliest attempts to apply such integration when considering a three-dimensional walkthrough of an architectural model (in this case Soda Hall, one of the principal buildings on-campus within Berkley University in California). The method adopted relies on objects being pre-specified with various levels of detail – in effect the environment must be discretised into a collection of "objects" and each one modeled at several levels of detail. A perceptual selection process subsequently determines which level of detail should be used on a per-object basis by weighting four criteria prior to rendering the frame:

- *Semantic*, representing the inherent weighting certain objects may have (Funkhouser et al. use the example of walls being more important than pencils to the subject of a building walkthrough). This may be seen as inherently accommodating the top-down model of attention described in Section 2.4.3, wherein an objective-driven view of perception is assumed.

- *Focus*, considering the visual focus of the user and its subsequent impact on a perceived scene. A simplification is made that the focal point of the user exists at the centre of the screen viewed at all times. Whilst this is an approximation, the method still gives good results suggesting the tradeoff (given the massive complexity of estimating focal position to a degree of accuracy) is viable. It relates directly to the spotlight approach described in Section 2.4.2.

- *Motion blur*, accommodating the spatiotemporal nature of perception to an extent, although rather than weighting moving objects more highly, Funkhouser et al. make the assumption motion blur will reduce the perceived detail, and subsequently *reduce* rather than increase the detail on fast moving objects.

- *Hysteresis*, to correct an implementation-level issue wherein rapid changes in level of detail for objects near the threshold result in a perceived effect. An object rapidly oscillating between levels of detail tends to draw a user's attention, and thus such oscillation is reduced by decreasing the perceptual weighting of objects which have recently shifted level of detail.

In order to accommodate these disparate concepts, an algorithm is created which uses a *cost versus benefit* approach to determine the level of detail. *Cost* is estimated by analyzing the impact of rendering objects on framerate, whereas *benefit* is a value generated by the application of the above perceptually-oriented weightings. Furthermore, the algorithm is engineered to attempt to maintain a constant framerate, and thus the allocation of objects to various levels of detail is dynamic dependant upon the cost of doing so. Figure 12 shows the level of detail weighting given to objects within the environment using the proposed approach. Darker shades of grey represent higher LODs. A frame using a static implementation of the method which seeks to simply allocate level of detail based on perceptual criteria is shown on the left. The optimized model, which attempts to maintain a constant framerate, is illustrated on the right. Note the higher relevance, more complex geometry towards the centre of the image. The higher resolution of this geometry is provided by issuing it with a higher inherent weighting, resulting in it being rendered with higher quality than the seats in the middle-ground, despite being more distant from the viewer.

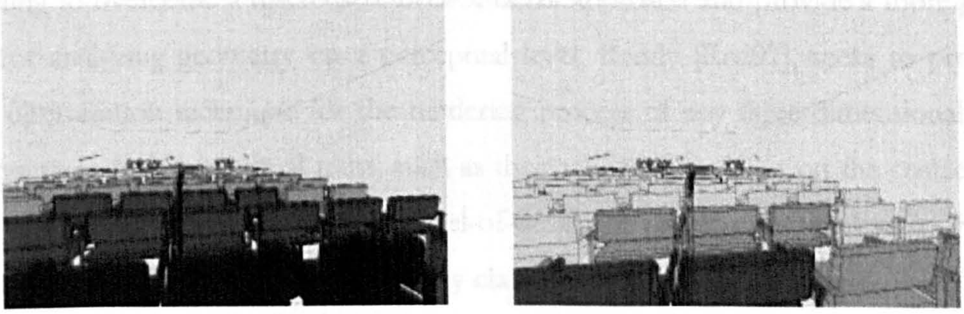


Figure 12: Adaptive Level-Of-Detail (LOD). Rendered scenes utilising the method proposed by Funkhouser and Sequin (taken from [FS93]).

Funkhouser et al. successfully provide an effective means for supplying a constant framerate within an environment ranging from simple to complex regions of geometric detail. The method has a variety of drawbacks; notably the need to specify the environment on an object-based level and at varying levels of detail. Hence the method requires considerable time to integrate into an environment and its compatibility and success is related to the ability of content designers to create and specify objects well (taking the above screenshot, what if all the chairs were specified as a single object?). The approach relies on discretised levels of detail being stored for all objects, and used as determined by the perceptual analysis of the scene. Discretised level of detail has been used extensively throughout previous decades (being first implemented by Clark as early as 1976 [Cla76]), although subsequent research has centreed around a *continuous* approach to level of detail, whereby the polygon mesh representing objects is modified continuously at runtime to define the total number of triangles rendered in a scene. This allows the level of detail to be determined dynamically, rather than relying on pre-generated models. The work of Hoppe [Hop96] into progressive meshes illustrates the feasibility of such approaches.

Attempting to overcome a discretised level-of-detail approach and provide a more generic means for analyzing geometry on a perceptual level, Reddy [Red97], seeks to provide a generic optimisation technique for the rendering process of any three dimensional scene. By analysing a user's perceptual traits, such as their tendency to focus on the centre of the screen, and subsequently reducing the level-of-detail for perceptually irrelevant objects a solution is provided which Reddy ultimately claims to offer a "four to five-fold increase in frame rates with no perceived loss in visual quality" ([Red97] p. 3). The approach itself uses a generic algorithm for triangle-reduction in geometry, alongside a saliency-map based approach that seeks to analyse each frame and thus identify less perceptually important areas of the scene. Once identified, the level-of-detail within these areas may be reduced with no detectable difference to user experience. Factors going into this analysis include the size and angular velocities of objects, and the degree to which they exist in the peripheral field of vision. Similar methods have been successfully applied to ray-traced rendering methods with further success [MRT00]

Sundstedt et al. [SDL05] propose a refinement to the rendering pipeline that accommodates visual attention. The goal of such refinement is to provide a means to selectively render animations with no perceived loss in quality. The ultimate goal of such an approach is to accelerate the high-quality, pre-rendered animations produced by the research such that they may be presented in real-time. Sundstedt et al. employ two mechanisms; *selective guidance*, which incorporates a saliency-based perceptual model, and *selective rendering*, which comprises the existing, non-attention oriented rendering process. Resources are allocated using the selective guidance model, then each frame rendered accordingly. Selective guidance incorporates a model of attention seeking to accommodate top-down and bottom-up attention by means of the creation of individual saliency maps which may be combined to produce an "image map" of perceptual relevance, as shown in Figure 13. From left in Figure 13, the original image is shown, followed by a map of objects specified as being task-oriented (and hence the focus of top-down attention), the resulting task map, the saliency map generated on a per-pixel basis, and combined "image map" of the scene.

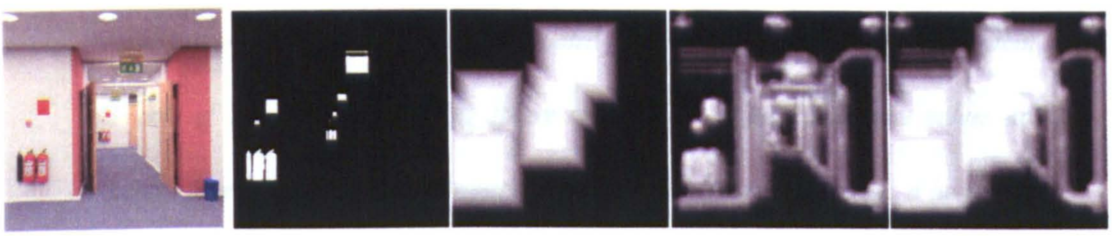


Figure 13: Modelling attention using the approach proposed by Sundstedt et al [SDL05].

The approach of Sundstedt et al. is aimed at a ray-traced approach to rendering, which particularly suits saliency mapping since the ray traced image is calculated on a per-pixel basis (by comparison, the less CPU intensive scan-conversion approach to commonly used in real-time interactive environments projects the scene based on viewer co-ordinates and geometric information). Ray-traced images provide a higher visual quality, serving to provide physically accurate simulation of the phenomena normal mapping simply seeks to approximate. Ultimately, a ray-traced approach that perfectly models all physical phenomena affecting each rendered pixel could be seen as providing a 'perfect' image; i.e. one which is identical to taking a real-world photograph of the scene. Hence optimisation approaches that seek to work within ray-traced rendering are of particular advantage when considering the potential for increasing processing power and need for photoreal imagery. Note, however, that even using this approach some degree of predefinition of environment content is required: in the illustrated example the fire extinguishers and exit signs are *defined* as top-down objects of interest. Such definition prohibits the creation of truly extensible virtual environments since it implies an arbitrary restriction on content definition; exactly how can a content author predict the top-down attentional needs of all users? Indeed, such prediction is made impossible simply by the situational and user-centric nature of top-down attention.

Brown et al. [BCP03] attempt to extend Reddy's approach by means of a model accommodating the four criteria of *Size*, *Position*, *Motion* and *Luminance*. The model first calculates importance as an absolute, Boolean characteristic with respect to these four criteria. For example, as Brown states:

> *"Luminance–the luminance feature is calculated from changes in luminance values within the area of the projected bounding sphere of the object. These luminance values are captured from a feedback buffer projection of low detail geometry. If there is a luminance contrast of greater than 1% within this area then the feature is important, otherwise it is not." – Brown [BCP03]*

This is subsequently combined with a *weighting* component, defined for each object, to provide an ultimate value indicating perceptual relevance. The approach may be seen as more comprehensive than Reddy's approach since it incorporates additional criteria when analysing importance; however it also illustrates the difficulties in producing a perceptually based system since the weighting values themselves are arbitrary. In effect they may be seen to represent top-down characteristics of attention; whilst importance may be defined by analysing easily measurable characteristics, a far more nebulous factor is incurred when considering the effects of goal-driven and learned attentional characteristics. Much as

Sundstedt et al. define objects as top-down objects of interest, so are Brown et al. forced to make assumptions about objects when creating their virtual scene.

In the long-term, integration of models of attention alongside methods used to optimise and reduce mesh detail offers an attractive potential to increase perceived quality of rendered images, with no impact on processing demand. As such, the research may be taken as an example highlighting both the potential and limitations of the use of theories of visual attention in computer-generated environments. More directly in the context of this thesis, a substantial advantage may be recognised when top-down objects of attention are specified and hence bottom-up characteristics may be used exclusively to determine interest. When considering multi-user networked environments, interaction amongst clients in some form is typically at the forefront of applications. Hence, the objects pertaining to network traffic may be correlated to top-down components, if a user's task involves them. Scene analysis is simplified in comparison to rendering by the fact only a small percentage of the environment content is typically shared via the network.

In this section a variety of attempts to optimise the rendering process, via the application of models of visual attention, have been described. Such rendering processes form the means by which a virtual environment is transformed from geometric details to a final rendered frame. In the context of distributed virtual environments, interest management is essential in providing this rendering system with the information regarding the positions and behaviours of networked entities for each frame. The next section considers more fully the role of the interest management process, and goes on to describe the current state of the art, both with regards to interest management itself, and the limited application of models of visual attention within the process.

3.3 Interest Management

Before discussing the relationship between interest management and attention, it is important to clarify, on both a conceptual and practical level, what interest management means, in the context of distributed virtual environments.

Consider first the case of a distributed virtual environment (DVE), which aims to mirror a real-world environment, to at least some degree. This includes the majority of practical applications of large scale, distributed environments, such as those detailed later in this chapter. In this case, each user is represented by an avatar, which has the capacity to interact with other users' avatars within the environment. World geometric and texture content is typically replicated across all users prior to run time, hence the network load is comprised entirely of updates regarding the positions and actions of avatars within the environment. Additionally, more interactive environments may incorporate other objects that may be interacted with at runtime (such as movable items or scenery). Each additional item also adds network overheads regarding updating its position and orientation, in order to maintain a consistent simulation amongst users. Collectively, all the objects and avatars within a DVE, which have attached network data streams, are commonly referred to as *entities*. Whilst environments (such as ActiveWorlds [AW06]) exist which seek to extend the networked content of the environment to include textures and geometry, this is commonly at the cost of visual quality and hence immersion. Ultimately, entirely dynamic and distributed content remains a goal of DVE design, although currently, providing a high quality of simulation for as few as several hundred entities within a static environment remains a challenge. Hence this thesis chooses to focus upon the current, rather than future state of the art, and seeks to consider interest management in the context of entities alone.

Interest management may be seen as an attempt to address two themes representing a recurring challenge when faced with environments containing more than several dozen entities: *latency* and *bandwidth*. Both arise from physical hardware limitations, and impact the quality of simulation within the virtual environment. Hence creating a believable environment as defined within the previous section implies taking steps to reduce or eliminate the perceived impact of these two factors.

Latency describes the delay between transmission and receipt of data packets between clients on a network. Ultimately this is limited by the transmission medium and physical distance between clients; in the case of fibre optic cable the absolute limitation is the speed of light within the medium (approximately 2 x 10^8ms^{-1}). In practice, far more significant delays are induced by routing and server hardware. Cheshire [Che96] notes a delay in practice between US West and East coasts of ~86ms, whilst the strict limitation in a perfect scenario utilising fibre optic cable is ~46ms. Additional delay is also induced by processing delays at end points on the network (packets must be assembled and deconstructed by clients). By comparison, in the context of networking, *bandwidth* refers to the absolute capacity of a network connection, typically measured in bytes per second. The lowest capacity node in the connection limits maximum bandwidth – in practice the majority of broadband Internet users within the UK experience connection speeds around 2 MB/s, although the capacity varies significantly between countries. Figure 14 shows the download speeds provided by a cross-section of Internet Service Providers over a 14-day period. This analysis suggests a modal downstream rate of 1.6 MB/s. Upload speeds are typically reduced in comparison; a factor advantageous for multi-user VEs since clients typically only need to upload information on their own current state, whilst downloading state information on multiple viewed clients simultaneously.

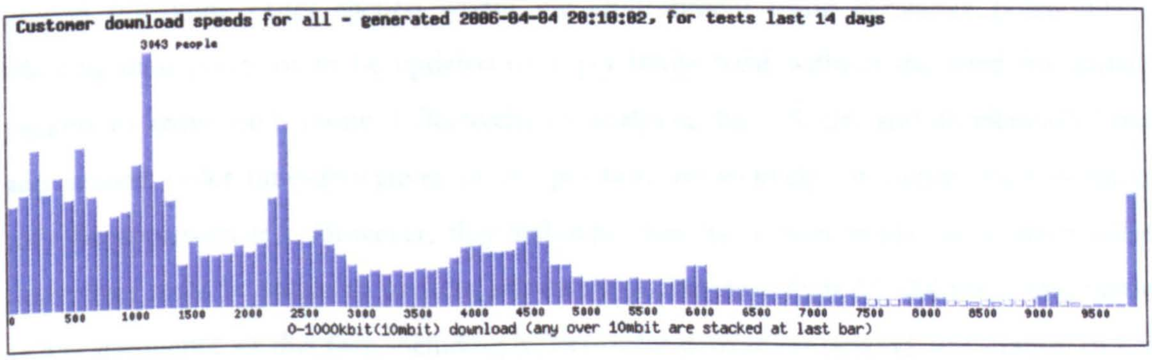


Figure 14: Average download speeds for 117 Internet Service Providers sampled (globally) over a 14-day period. Generated by [DSL06].

A relationship between bandwidth and latency may be assumed, since a server overloaded with packets will respond by queuing packets, increasing latency. An ideal virtual environment thus makes as efficient a use of bandwidth as possible, in order to minimise the impact on latency induced by queued data. However, since even in the best case scenario latency remains present and hence techniques to reduce its impact directly are frequently employed within virtual environments. A primary approach is the use of *dead reckoning*, first employed by Zyda et al. within the DIS Protocol [IEEE95]; a recognition of the fact providing multiple users with a shared simulation is a far more complex issue than it may initially appear. The DIS protocol was the first attempt to formalise techniques – primarily dead-reckoning - as a means to reduce perceived effects of latency. Objects within a DIS-defined environment have properties that can include acceleration and velocity, and also a flag stating whether to use these properties in 2nd or 3rd order dead-reckoning. It is important to note this close link between object properties and predictive algorithms – 2nd and 3rd order reckoning, as the names suggest, take the 2nd and 3rd order differentials over time of the position of an object into account, and hence the object is required to possess these properties (namely velocity and acceleration) to use these methods. More generally, it illustrates that to accurately simulate an object using predictive techniques, there is a need to provide (or extrapolate) not only set properties, but also their higher order differentials over time.

The implications of the use of dead-reckoning on interest management are notable. In providing a means to predict the position of an object, the need for frequent updates on its position diminishes. In dead reckoning, each client independently predicts the position of objects (including other clients) within the environment using derivative polynomials, allowing their positions to be updated on a per-frame basis without the need for update packets to arrive each frame. Effectively, by analysing the velocity and acceleration (first and second order time-derivatives of the position, respectively) of clients, their position may be approximated. However, this technique has bandwidth implications since these derivatives must be added to packets, increasing packet size. Zyda [ZS00] notes that partly as a consequence of this fact, including higher-order derivatives beyond acceleration offers diminishing returns. Additionally, the technique serves only to ameliorate, rather than eliminate the effects of latency, since erratic movement from clients which deviates from the predicted path results in visible inconsistencies as clients "warp" between predicted and correct positions as packets arrive. Additional algorithms, which seek to perform this correction with less perceptual impact or analyse behaviour to generate more accurate initial predictions, are a topic for ongoing research [CH04].

More efficient bandwidth consumption thus has advantages in allowing reduction in latency. It is in providing this efficiency that the notion of *Interest Management* arises. For an environment with thousands of users, simultaneous global state updates across all clients are impractical, since an individual client is, at any given time, unlikely to be interacting directly with more than a small handful of users. Establishing which users are of central concern to a single client is the role of an Interest Manager, and allows additional, less relevant clients to be ignored. Bandwidth reduction may be achieved by *inducing* latency on less or non-relevant clients; increasing the update interval on these clients in order to generate more available bandwidth with which to decrease latency on the more relevant clients. The following section considers the variety of current means by which this relevance is determined, alongside the mechanisms by which bandwidth is preserved. However, before such consideration, this section seeks to provide a more comprehensive definition of Interest Management. Drawing from the previous discussion regarding latency and bandwidth, interest management may be seen as the mechanism whereby bandwidth limitations are addressed.

Morse [MBD96] provides one of the earliest explicit definitions of Interest Management with reference to the DIS protocols developed by Zyda et al. [MPZ94] when implementing the first large scale distributed virtual environment, SIMNET. Noting that interest management is also often termed *"relevance filtering"* and *"data subscription"*, Morse attempts to clarify and present the concept by means of taxonomy. The DIS simulation itself may be regarded as the first successful attempt to provide a large-scale distributed environment, being central to the SIMNET (Simulator Networking) system developed by the US Department of Defense to provide large-scale battle simulation [MT95].

Figure 15 provides a generic illustration of the role of an interest manager within a networked environment. Residing on the host machine, the interest manager acts upon a local database to determine what is of interest, and then queries the network to retrieve relevant data. Simplified to this extent, the process may be viewed as a highly generic concept independent of application.

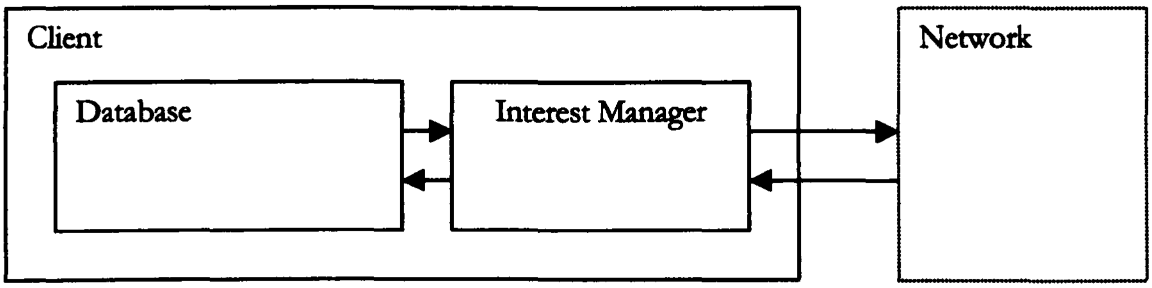


Figure 15: The role of interest management.

In such a model, the nature of a virtual environment, both in terms of architecture and application, are of concern within the interest management process – taking the above diagram into account, some understanding of the nature of the database must exist for relevance filtering to take place. Hence 'generic' interest management is limited to providing methods which seek to function within the most common application scenarios. These include military, collaborative and entertainment purposes; in the case of all three typically allocating users a virtual 'avatar' and rendering the world from their perspective.

3.3.1 Interest Management and Attention

Given the previous section's definition of interest management through discussing its role within a virtual environment, this section seeks to take this definition and stress the implicit link with perception present on a conceptual level.

Interest management seeks to provide filtering of network data. In short, the amount of data potentially present in a rendered frame of the virtual environment, outstrips the network capacity. A simple example is a large crowd of users – whilst a user may view several hundred entities simultaneously, the network load required to provide high-resolution updates regarding the position and activity of each user, per frame, restricts the ability of current virtual environments (as discussed in the next section) to handle large crowds. Hence interest management seeks to filter redundant data, in order to provide the best quality of simulation possible.

Referring back to Chapter 2, Broadbent draws parallels between human attention and filtering. A notion common to all theories of attention discussed in Chapter 2 is that the information entering our senses, such as the light falling on the retina, is considerably more voluminous than the information we perceive. Phenomena such as change blindness reinforce the assumption that our view of the world is a mental approximation, rather than perfect facsimile. Given this notion of both interest management and human attention as filtering processes, it becomes possible to identify clear links between the two. A "perfect" interest manager would be capable of analysing both user and virtual environment, so as to provide the user with only the data required to maintain an uninterrupted experience which accommodates the notions of believability and presence as noted at the start of this chapter. In doing so, the bandwidth requirements could be minimised, with no perceivable impact. In order to evaluate existing methods with regards to their effectiveness in this role, a key question is how best to extrapolate the determination of interest from the practical overheads associated with interest management, such as the network architecture.

Figure 16 illustrates the interest management process as a whole, from the conceptual level, to hardware. Central to the process is an analysis of the state of all entities within the environment (or more accurately *locale*, discussed later in this section), which leads to a determined relevance and hence relevance value (either explicit or implicit) for all entities. This is then translated by the interest manager, which performs resource allocation on a hardware level based on these relevance values. It can be observed that whilst substantial differences are needed in networking to accommodate the outcome of the interest management filter across architectures, higher-level conceptual aims regarding the establishment of relevance remain constant.

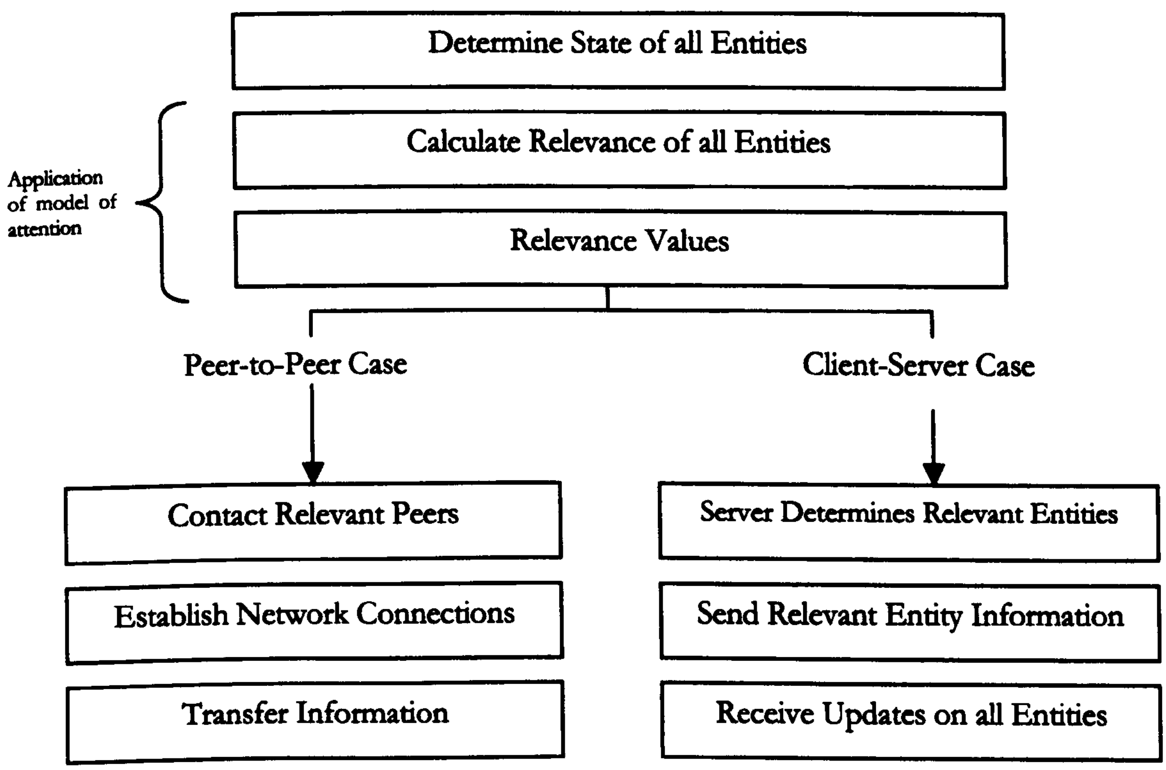


Figure 16: A Model of the Interest Management Process for both Peer-to-Peer and Client-Server environments.

The critical part of an interest management approach is in establishing the relevance of networked entities. Much research, such as that of Masa and Zara [MZ04], focuses on the lower levels of the model, wherein pre-established relevance values are used, typically based on a simple proximity filter, and filtering becomes a task for network structuring and protocols. However, in the context of this thesis, we study the process by which relevance is attained, rather than the hardware and software issues responsible for interpreting these values.

In many systems [GB95], [MZP94], [WZ98], [HPG02] the relevance determination procedure leading to the relevance values is simplistic, being based purely on distance or visibility, with relevance values being Boolean in nature. Scope therefore exists for closer analysis of how such relevance values are obtained and how the process may be related to visual attention. Hence so far within this chapter, the concept of interest management has been defined, and an abstraction from the hardware and networking level presented. We thus in the next section critique the current state of the art in terms of its ability to determine relevance precisely, rather than focussing on practical and implementation issues.

3.3.2 Methods of Interest Management

Early taxonomies of interest management approaches, such as that of Morse [MBD96], sought to classify approaches based on the network architecture used. Since systems utilise either unicast, broadcast, or multicast approaches, this proves a simple and effective means of classification. Since multicast-capable hardware was comparatively limited in the development of the first multi-user environments, with respect to both the number of multicast groups supported and the time taken to reconfigure them, broadcast and unicast approaches remained attractive, despite the potential for multicasting to reduce network load. Thus early multi-user virtual environments (in this case used for military simulation) such as JPSD and STOW-E used unicast and broadcast respectively [PMW96]. However, even in this early phase of large scale VE development, multicast was rapidly proving effective in systems such as ModSAF [SRS95], and NPSNET [MZP94], an architecture still in development today, both of which implemented multicasting successfully showing its significant advantages in reducing network load. PARADISE [SC94] can be credited as the first system to demonstrate the potential for multicast architecture – since it was unavailable during development in 1993 PARADISE emulates multicasting using software. As Macedonia notes [MZP94] the early technological drawbacks with multicasting were rapidly overcome with advances in hardware, and currently the Internet itself provides means for large-scale multicasting via technologies such as the MBONE.

Following initial classification by communication model, Morse further categorised techniques into *intrinsic* and *extrinsic* approaches. The key difference between the two approaches is whilst an intrinsic approach seeks to filter an entity based on its locally-stored attributes, an extrinsic approach filters based on an analysis of other entities and hence requires information regarding their attributes. To clarify, an example of extrinsic filtering would be to filter out all objects with a certain name, or colour. This may thus be related to both bottom-up and top-down models of perception. Conversely, an intrinsic filter would consider only the location of the object within the virtual universe or its multicast group subscription. Initially, intrinsic filtering was preferred in multicast environments due to the length of time taken to reconfigure groups for extrinsic filtering (which by nature is more fine-grained). As mentioned, however, advances in hardware coupled with a demand for larger-scale environments have made extrinsic, multicast-based interest management desirable. MASSIVE [GB95] and VELVET [OG02] can both be seen to implement such approaches successfully.

In order to categorise systems more effectively, the next sections adopt the common approach of refining the themes of intrinsic and extrinsic approaches into the more conceptual notions of *aura* and *locale* based filtering. Locale based interest management can be seen to draw on the notion of intrinsic filtering to filter objects by their location within the environment. By comparison, aura-based approaches seek to centre the region of interest upon the user and are thus classifiable as extrinsic approaches. These two approaches, often used in combination, can be accurately used to describe many current approaches to interest management.

3.3.2.1 Locale-Based Interest Management

Locale-based filtering (also commonly referred to as *region* or *grid*-based interest management), as the name suggests, is, at the simplest level, a method of filtering based on the location of the user relative to the environment. Such a method was first used by SIMNET and later NPSNET [MZP94]. SIMNET (**SIM**ulator **NET**working) was the first such attempt by the US Department of Defense (DoD) to link simulators together, to provide a networked environment for military training [MT95]. Whilst in many respects a primitive system in comparison to current environments (the project beginning in 1983, and being delivered by March 1990), many of the concepts first created for SIMNET remain at the forefront of current networked virtual environment design. Concepts such as heterogeneity, and the existence of multiple user-entities within networked virtual environments, which are central to the current NPSNET-V (an evolution of SIMNET technologies also developed by the US DoD), online games [MLS05], and many other frameworks, were first developed and explored during the SIMNET project. These and other concepts were further refined and condensed into the Distributed Interaction Simulation Network Software Architecture [MZP94], a framework with its foundations in the SIMNET project but which was intended to allow for a far more heterogeneous and scalable system than the limited, application-specific SIMNET system. Such refinement chiefly involved the separation of data modelling and transmission techniques into the aforementioned Distributed Interaction Simulation (DIS) protocol, which embodied early concepts within these two fields central to large-scale multi-user virtual environments.

Interest management within these systems initially centred on splitting the environment into a grid, as shown in Figure 17 overleaf. However, issues were observed when a user nears the boundary between grid sections. A user near the boundary between two regions must receive updates from both regions, since they are likely to be viewing entities present in both regions. Without receiving updates from the adjacent regions, the simulation breaks down as entities can be seen vanishing or appearing as a client moves between cells. Hence it is important to transmit not just the content of the current cell towards the boundary location, but also that of the adjacent cell. By changing from an orthogonal to hexagonal grid as shown in Figure 17, the maximum number of regions a client in the worst-case scenario (on the boundary of several cells) need receive data from simultaneously can be reduced.

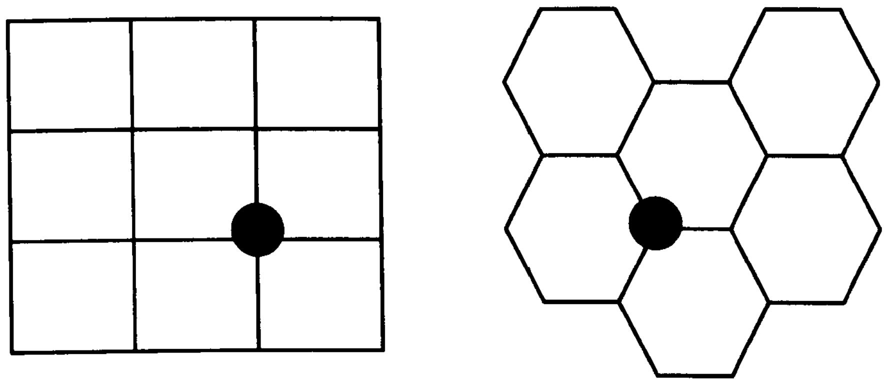


Figure 17: Orthogonal versus hexagonal area-of-interest management. A client on the boundary between areas of interest is marked with a black dot.

Such a method of interest management has several strengths. Firstly, it is simple to implement, and also has low (and distributed) processing overheads, since clients need simply subscribe to the relevant cells based on their co-ordinate location within the virtual world. It is particularly suited to large, open areas where a user's maximum visibility is defined by a graphical clip plane, and thus the size of a grid cell may simply be defined based on this value. It is primarily these two advantages that lead to its use within NPSNET, which deals with military simulation over open areas of terrain.

A fundamental problem with locale-based interest management emerges when dealing with entities that tend to group together. As the cells are defined statically in the case of NPSNET, yet the users by nature are able to move freely, the potential exists for users to gather within a single cell and overload the system. A potential workaround, which alleviates the problem to some degree, is a partition of the grid hierarchically and dynamically, such that the number of entities per grid section remains constant. The Bamboo architecture [WZ98] attempts such partitioning - being developed by the same research group as NPSNET, Bamboo shares several of the technologies and concepts driving later versions of NPSNET. Both support dynamic protocols, towards the end of supporting extensibility [WZ98b]. Bamboo goes further by adding support for persistent worlds, together with the ability to suspend and resume state of the world. Zyda describes both these features as essential "to implement highly scalable worlds" [ZS00].

The Bamboo system particularly focuses upon area of interest management. A three-tiered approach is taken [AWZ98]. The first tier filters based on partitioning of the world into a grid along Cartesian coordinates. This closely mirrors NPSNET, but is further refined by the ability for the grid to resize dynamically to suit user distribution. If a large number of users enter a single cell, it is partitioned into new cells, whilst empty cells are recombined into larger areas. A second tier goes further by adding a pass based on client requirements. This essentially involves defining an area of interest on the client, typically a spherical region, outside which entities will be ignored even if they co-exist within the same grid region. This level hence requires that entities be assigned their own multicast addresses, to enable them to be filtered separately. The final tier is described as 'protocol-dependant' filtering. Such a method of filtering consists of querying information (hence 'protocol-dependant') about objects within the area of interest calculated in the second layer, and eliminating those deemed unimportant by the client. This may be interpreted as a certain degree of accommodation of top-down models of attention; the rule set used to determine relevance may be dynamically altered based on the user's current task.

Since the data needs to have been transferred in order for this querying to take place, the final layer provides no benefit to networking, but elimination of unnecessary objects prior to rendering enables a simple split between the tasks of determining what to render, and the actual rendering process - hence facilitating the platform independence Bamboo is keen to stress. Figure 18 shows grids partitioned using such a method, which can be equated to the quadtree and octree division techniques used in a wide range of areas, such as graphical compression and terrain modelling. Whilst such methods can reduce the impact of grouping behaviour by redistributing network load, there remains the problem for potential overloading since grid cells may effectively be reduced to lower than the visibility range of clients. In this case the client may experience visible artefacts as entities leave or enter their region of interest whilst still within view.

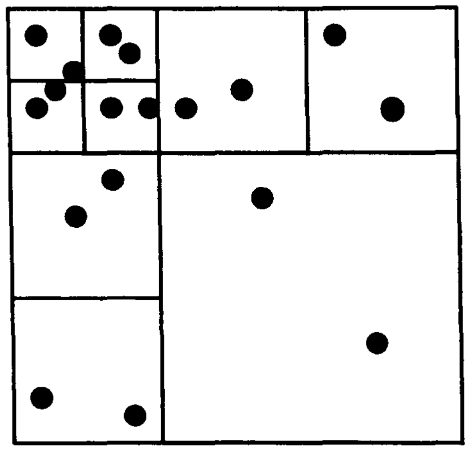


Figure 18: A quadtree-based grid partitions itself based on the dispersal of the clients contained. Clients are again marked as black dots. Diagram based on the Bamboo system [WZ98].

A further level of sophistication is to balance the partitioning around the environment itself. Environments may be broken down into what are commonly termed 'locales' – typically enclosed regions where the user would not be expected to 'see' outside of them. In a simple simulation, this could involve separating rooms within a building into individual locales – since a user would be unlikely to see, or interact with, the contents or occupants of another room, yet would expect to see everything within their own room. However, whilst many algorithms exist for determining geometric visibility within environments, they are heavily processor-intensive. Output is typically pre-calculated based on the notion that the environment will have largely static geometry and a handful of moving objects, making such methods unsuitable to highly interactive environments. Subsequently, this method is incompatible with the dynamic grid mentioned previously, since recalculating locale volumes in real-time is prohibitively expensive in processing terms. The work of Hu and Liao [HL04] takes advantage of a peer-to-peer network architecture to provide interest management based on a grid which is formed using a Voronoi-diagram based approach, shown in Figure 19. The right-hand diagram shows a client with area of interest (red circle), and corresponding peer-to-peer connection establishment with nearest neighbours (green), next to nearest (blue) and other cells (yellow).

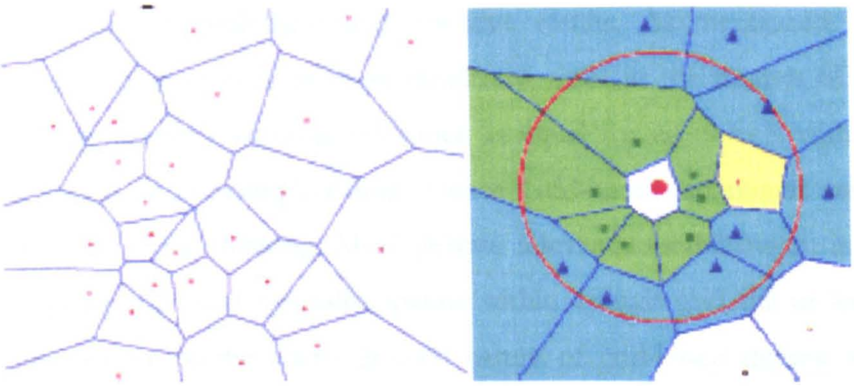


Figure 19: Voronoi-grid based filtering.

In this approach, the environment is partitioned dynamically based on the creation of non-overlapping areas that contain one client in each region. Each client maintains a peer-to-peer connection with clients corresponding to neighbouring regions. The diagram itself is based around the client's position in the virtual world, rather than their network proximity, and hence the method provides a form of grid-based interest management. However, since the grid consists of one client per cell, the filtering is applied by specifying a volume corresponding to the client's *area of interest* (AOI). Cells contained either wholly or partially within this volume correspond to peer-to-peer connections.

From a standpoint regarding theories of attention, grid or locale-based interest management may be seen as a very simplistic top-down measure of salience, as described in Section 2.4.3. Users are considered to be potentially interested in interacting with all other users within the same (pre-defined) area and hence provided with the relevant information. Note that, in the typical grid approach, direction of view or other properties possessed by the client are ignored.

Grid based interest management systems remain a mainstay of existing applications, due to both their simplicity to implement and effective results. As mentioned, however, it struggles to handle large crowds of users efficiently, since as the models of attention in Chapter 2 suggest, simple measuring relevance as equal for all objects within the same locale of the user is an oversimplification. Hence Grid-based techniques are commonly used as a means of coarse filtering. More precise relevance establishment is the goal of Aura-based approaches, which typically operate within a single grid cell or locale. It is an additional consequence of the coarse-grained nature of grid-based interest management that it accommodates theories of visual attention only loosely. Attempting to provide a high-resolution filter by a locale or grid-based approach implies extremely small grid cells, making it particularly prone to the aforementioned overcrowding issues. Thus in applications such as NPSNET, cells are typically approximately the same size as the user's maximum view distance. Furthermore, levels of resolution are difficult to support, with the existence of a client within a grid cell being Boolean in nature. Subsequently, there is very little scope for the application of models of attention, with grid based interest management being used on a high level to determine the clients coexisting within a single region. However, in the next section, aura-based approaches are discussed, which due to their greater precision, offer much greater scope for visual-attention based enhancement.

3.3.2.2 Aura-Based Interest Management

The first implementation of aura-based interest management was within the MASSIVE framework developed by Greenhalgh et al. [GB95]. The original MASSIVE (Model, Architecture and System for Spatial Interaction in Virtual Environments) project was intended as a teleconferencing system, with the concepts being developed as part of this specific application being continued to the more general-purpose and conceptual MASSIVE-II system. At the core of these concepts lies a 'spatial model' of interaction. Much of the research within the MASSIVE-II project was directed at user-to-user interaction, which can occur over a variety of mediums within a MASSIVE-II environment, including text messages, users writing within the environment, to voice communication. In enabling this high degree of interactivity, MASSIVE-II makes use of peer-to-peer communication alongside a client-server model. The most immediate benefit of this architecture is a reduction in server load, particularly in terms of bandwidth (for which voice communication is relatively costly).

Whereas locale-based approaches seek to filter based on the relationship between an entity and the environment, MASSIVE implements an aura-based approach, centring the process on the entities themselves. Greenhalgh adopts an approach based upon the notions of *focus* and *nimbus*. Focus defines the region of interest belonging to an entity, whereas nimbus defines the entities intrinsic level of relevance to other entities. When an entity's focus intersects with the nimbus of another entity, the observed entity is deemed relevant and network resources allocated to providing a high frequency of entity information updates. The aura-based process is illustrated in Figure 20.

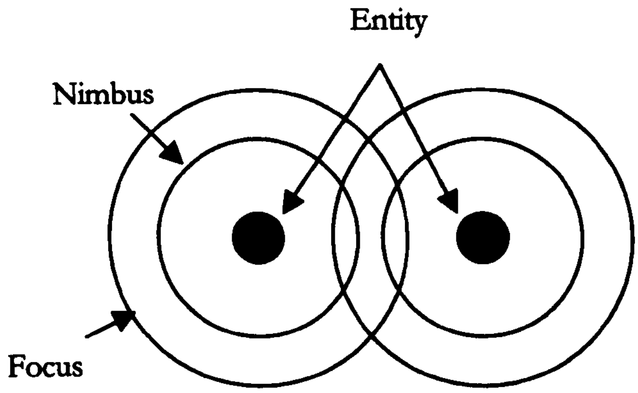


Figure 20: Illustration of Aura-based Interest Management. Two clients near to one another within the virtual world and focus / nimbus intersection.

One drawback of the aura-based approach is that in order for focus / nimbus intersections to be detected, all clients on the network must exchange data on some level in order to provide an approximation of position by which to perform the detection process. The frequency of these updates must be sufficiently rapid to allow intersections to be determined accurately; in practice this is affected by a number of factors such as the maximum speed with which clients may move within the VE. A common example is that of a fighter aircraft flying over a foot soldier. In order for the simulation to appear seamless, the interest manager must detect the aura collision and allocate resources accordingly. However, in practice, the high speed of the aircraft may mean aura intersection occurs so briefly the network may not allocate resources rapidly enough, as suggested by Morgan and Lu [MLS05]. They suggest:

"A solution to this would be to extend the fighter aircraft's aura to enable such interaction. However, extending the aura may result in the fighter aircraft potentially influencing many more objects than is necessary and may result in scalability problems as the node hosting the fighter aircraft would be required to participate in redundant message exchange with many nodes." – Morgan and Lu [MLS05]

Morgan and Lu go on to propose the use of dead reckoning to anticipate aura intersections before they occur, in an attempt to ameliorate this problem. Whilst such an approach may be shown to reduce the problem, issues with the focus and nimbus approach persist. The VELVET architecture considers manipulating the focus and nimbus of clients independently to provide scalability and heterogeneity, and in doing so identify a *degree-of-blindness* problem [OG02]. This occurs when the focus and nimbus of one client is scaled such that it encompasses another client, whilst the observed client's aura is scaled down and hence mutual visibility does not occur (Figure 21). Induced in a hide-and-seek game by varying focus and nimbus sizes, the 'hider' can observe the 'seeker', but due to the hider's smaller nimbus, the seeker cannot observe them. This is problematic when processing interactions between clients whilst maintaining consistency.

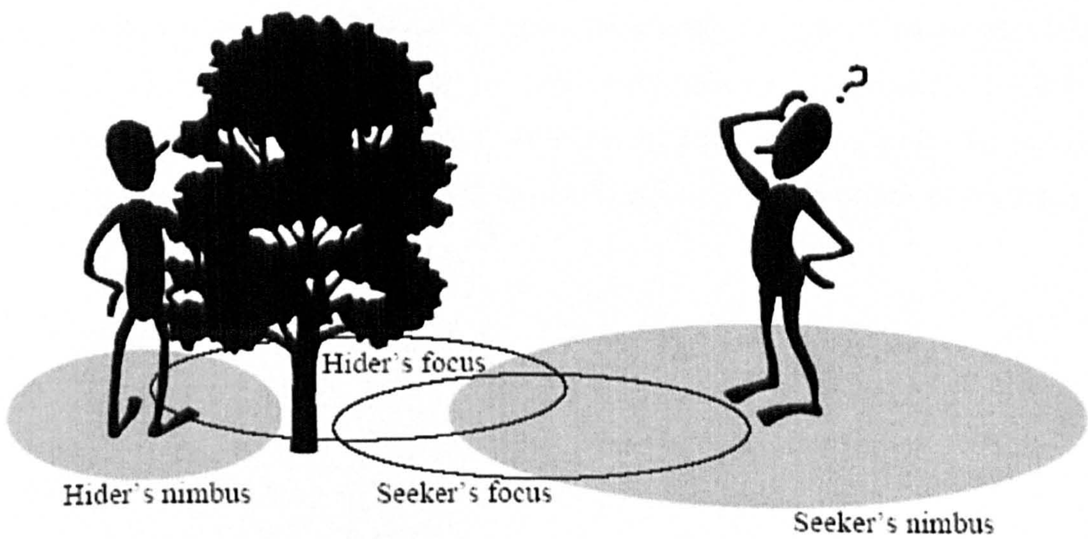


Figure 21: Degree-Of-Blindness. Taken from [SKH02].

Aura-based approaches accommodate notions of perception both in the form of the aura, and the way in which they interact. A spherical approach such as that of Greenhalgh implies an approximation of a linear relationship between distance and saliency. The dynamic approach explores this relationship further, by suggesting the relationship between saliency and distance is not constant; though both fail to draw parallels to research into visual attention in doing so. Similarly, offsetting the focus as a sphere in front of the client (as in Figure 21) resolves issues with non-visible entities being judged salient, but is again a simplistic approximation of visual attention.

An advantage of the aura-based approach is a far simpler capacity (in comparison to grid-based techniques) to introduce level-of-detail, by means of layering auras of increasing radius upon one another. In this context at least, it is possible to see the relationship between approaches which provide layers of detail such as that of Park et al. [PKK00], and the concept of saliency values described in Section 2.4.4. The advantages these methods offer are the capability to judge relevance as a quantity, rather than absolute value, and hence better allocate resources. Again, a basic relationship between distance and saliency is implied, though through a level-of-detail implementation a higher quality of simulation may be offered for nearby clients.

3.3.2.3 Summary

In this section, the paradigms driving the state of the art of interest management have been presented. Grid based interest management has been presented as being typically utilised as a coarse-grained filter, which consequently gives little consideration to perception. This lack of consideration is justified, however, when grid-based filtering is simply used as a means of providing a high-level filter within which a more precise aura-based paradigm functions. The integration of the two approaches has been demonstrated in the work of Watson et al. [WZ98]. Thus a typical model for a virtual environment accommodating several thousand users, such as NPSNET, revolves around grid-based interest management as a coarse filter, limiting data transfer to entities co-existing with the same general area (in the case of the aforementioned military simulation this is around 1km), whilst an aura-based approach handles more accurate interest management within each cell. This eliminates the aforementioned scalability issue of aura-based approaches (that all clients must be aware to others to some extent to detect aura intersection), by ensuring this awareness is limited to a locale and thus reducing global network traffic. Thus the focus of visual-attention based filtering may be considered to be the contents of a single grid cell, as a refinement to the aura-based process.

In the next section, we consider more fully the concept of interest management approaches that seek to exploit visual attention. Effectively, an approach accommodating perception can be seen as redefining the traditional, spherical aura-based approach as defined by Greenhalgh [GB95], either by affecting its form or the measures by which it determines relevance (colour, for example). Such is the case described in the next section.

3.3.3 Visual Attention-based Interest Management

Grid and aura-based approaches have, as described, typically made no explicit acknowledgement of their relationship to visual attention. However, Beeharee [BWH03] acknowledges such an explicit link, and successfully demonstrates the potential for the use of visual-attention based IM in a variety of simulations, ultimately concluding with a simulation of an 800-entity virtual city. The perceptual culling is based around the notion of change blindness; in effect aiming to filter out items of data in which the user is unlikely to identify a change. The approach adopted combines top-down analysis of attributes with a bottom-up component, effectively creating a means for interest management centred on the notion of multimodal attention presented in Chapter 2. In the process of developing an interest management paradigm, Beeharee performed a number of perceptual experiments within virtual environments to better determine the properties of objects that evoke a shift in a user's attention. It was consequently found that colour and motion are primary attributes in this context (conforming with existing bottom-up models of attention as described in Section 2.4.3), and this having been noted the interest management technique was developed and implemented within an existing VE, which comprised of a simple virtual city with coloured blocks representing avatars. Subsequently the technique was evaluated with a "small but non-trivial" ([BWH03] p. 214) sample of users who claimed to perceive no difference with the interest management approach in use despite a clear reduction in bandwidth consumed. By this means, evidence is provided that the technique is a success.

Figure 22: The "Virtual City" used by Beeharee et al. [BWH03]

The system is shown to provide a distinct reduction in bandwidth with users unable to detect change, and may thus be regarded as successful, though only within the context that processing overheads are deemed insignificant. In particular, saliency mapping is shown to lend itself well to any perceptually oriented interest management technique, since it offers a clear separation between establishing the perceptual relevance of discrete network entities and performing the actual filtering. Certainly, as processing capacity advances whilst bandwidth remains a commodity, the uptake of such an approach becomes increasingly attractive. However, additional overheads are involved in defining environment content such that it may be analysed via a bottom-up mechanism. Furthermore, as with any image-analysis based technique, the processing overheads are noted as being substantial. In providing both top-down and bottom-up components that provide extensive image-analysis, the demand in terms of system resources is significant.

In processing avatars based on properties such as colour, additional complexities introduce themselves as environment content becomes more sophisticated – whilst a simple three-dimensional box has clearly definable attributes, a textured and animated avatar offers a greater challenge in processing to determine saliency. Such complexities, coupled with processing overheads, present a notable drawback to attempts to provide comprehensive bottom-up modeling of attention (discussed in more detail in the next section). In attempting to apply a top-down component, some requirement is placed upon the user to predefine their context and goals within the environment. Since top-down attention is experience and goal-driven, even more restrictions are placed on the content and design of the environment such that these goals may be established by the interest manager. When considering the models of attention stated in Chapter 2, although Beeharee provides a comprehensive incorporation of top-down and bottom-up attributes, the question remains as to whether alternative means of modelling attention may offer additional – and compatible - gains.

The next section goes on to firstly outline the need for improvement, and then critique the current state of the art. A particular, recurrent theme throughout the previous sections has been the relationship between interest management and human attention as filtering processes; both exist as mechanisms intended to eliminate extraneous data. Hence the emphasis of the next section is upon the limited uptake of perceptual considerations within existing methods, and the limitations specific to existing methods that seek to exploit visual attention.

3.4 Limitations of Existing Methods

In reviewing the existing themes of interest management, alongside applications of theories of attention within rendering, and theories of attention, it becomes possible to consider the challenges faced by the current state of the art. Before considering the limitations of existing systems with regards to visual attention, it is worth giving some consideration to more practical concerns in distributed virtual environment design, and thus provide some evidence that bandwidth is indeed an ongoing limitation and obstruction in virtual environment design. In Section 3.4.1, the notion of bandwidth efficiency is considered, based upon the notion that more efficient use of bandwidth offers the potential for richer, more interactive environments It is noted that the evaluation of interest management is frequently application-dependent, and thus it proves difficult to directly compare existing approaches, though all share a common goal of more efficient utilisation of bandwidth. Thus it may be argued that bandwidth efficiency when faced with large number of entities is a challenge for all current interest management approaches.

Interest management approaches typically acknowledge the relationship with visual attention in only a simplistic, implicit fashion (e.g. by taking distant objects as being less salient). The work of Beeharee et al. [BWH03] is unique in directly accommodating visual attention as a means for interest management. Section 3.4.2 establishes potential avenues for improvement, by analysing the limitations of existing systems particularly with respect to accommodating visual attention. Subsequently, the tendency for existing methods to adopt rudimentary approaches to incorporating models of visual attention is re-emphasised. This leads to a more detailed discussion of the potential advantages offered by a spotlight model in Section 3.6.

3.4.1 Need for Improvement

Since, due to advances in technology, the bandwidth available for both Internet and LAN-based virtual environments has increased exponentially, the argument may be put forward that eventually hardware will outstrip the need for effective IM, as available bandwidth far exceeds the amount required to mutually update all users. Since IM places an inherent demand on CPU time, this point is worth consideration. Adopting such a paradigm, however, ultimately leads to less efficient virtual environments. The bandwidth interest management saves may be re-utilised to serve either additional clients, or enhance interactivity by adding further objects which may be manipulated by users (and which also require network data). Furthermore, interest management may actually offer improvements in performance, as the CPU is no longer required to handle redundant network data or update clients within the rendering engine that are of no significance. Thus a strong case may be put forward for the need for interest management within large-scale virtual environments in both the short and medium-term.

Given these facts, it becomes clear that an approach which offers a direct reduction in bandwidth is advantageous, and in the particular case of an Internet-based virtual environment which gives consideration to heterogeneity and scalability (criteria well-established by Zyda [ZS00]), the ability to gracefully degrade the quality of simulation in a situation that available bandwidth falls below the total required to mutually update all clients is also of paramount importance. Additionally, it should be noted virtually all existing systems utilise multicasting, since the benefits largely outweigh any drawbacks, as demonstrated by Macedonia et al. [MZP95].

In proposing the interest management within the NPSNET system Abrams [AWZ98] considers the data sent in terms of packets passed between network nodes over time. This has the advantage over raw bandwidth measurement (in terms of bytes / second) since it remains application-independent – for a more sophisticated simulation more bytes may be sent per packet, for instance to include higher orders of dead reckoning or animation states, but the update rate will remain constant. Other evaluations of interest management also tend towards this model for evaluation, such as those used within MASSIVE [GB99], [BWA96]. In both cases, this evaluation is done against a backdrop of constant quality of simulation as previously mentioned (i.e. the entities near the user receive the same quality of network service).

Despite the advantages of measuring the total packets passed between nodes over time, the measure of raw data sent is often advantageous when evaluating different interest management paradigms within the same system. A number of experiments within DIVE have adopted this approach [ZYM02], since the ubiquity of the environment allows for comparisons to be made between differing approaches. Hence two clear themes of bandwidth evaluation may be observed – packets / time to allow for general comparisons and bytes / time for comparison of techniques within a single system. Furthermore, a consideration of consumption on a per-node basis may be made to better understand the distribution of load within the environment, a factor critical when developing heterogeneous environments since a high load on a single node is undesirable.

Considering next the network architectures used by DVEs; a common trend is observable in commercial environments adopting client-server architectures, whilst research environments utilise peer-to-peer configurations. Research environments such as DIVE (the first significant peer-to-peer environment) identify many advantages to such an approach. Foremost, the distribution of load leads to the capacity to accommodate far more users without increasing the load on a single network node. Additionally, the decentralisation leads to greater fault tolerance and the capability to create environments with zero-downtime. Originally developed by the Swedish Institute of Computer Science in 1993, DIVE (**DI**stributed **V**irtual **E**nvironments) [Car93] is a peer-to-peer system, making it fairly unique compared to the other systems discussed within this section. The original DIVE system scaled poorly (to a maximum of 32 clients), as a direct result of the peer-to-peer implementation. However, the most recent release, DIVE 3 [FS98], uses multicasting to overcome many of the drawbacks with peer-to-peer frameworks and successfully provides a scalable platform. The world database for the DIVE 3 system is treated near-identically to a shared memory space over a network. The structure and content of the database itself is comparatively limited – object geometry is stored, together with a single flag that enables collision detection. Object motion and interaction is handled entirely by separate applications, which either modify the world database directly or provide scripts triggered by events in the world. Since the world database in DIVE 3 is partitioned between clients, its developers argue that a client-server architecture would require such a high degree of client–server communication (effectively occurring each time an object was modified) that "such an architecture would have negative implications on interaction time and would introduce lags."[FS98b]. The success of the DIVE 3 system (Pettifer [Pet99] notes that "DIVE has almost become a *de facto* standard platform for research in distributed VR applications"), illustrates the feasibility of a peer-to-peer approach. By manipulation of the architecture, the system has been used to explore spatial [HLS97] and subjective [SGB95] techniques for interest management.

However, commercial environments (such as World of Warcraft, which has approximately 7 million subscribers, leading to several hundred thousand simultaneous users at any given time [MMoG06]) are typically built around subscription systems. In such a situation, security is pivotal – not only in ensuring only authorised users can access then environment, but also in preventing malicious users disrupting the service (e.g. cheating in online games). Creating such security in peer-to-peer environments represents a substantial challenge; in the parallel case of the Internet, research into *trust* within a peer-to-peer scenario has yet to offer a comprehensive solution. Hence commercial environments commonly adopt client-server architectures out of necessity.

Ultimately, any interest management process has implicit links to the network technology underpinning the environment, and consequently several key limitations in existing systems are frequently described as arising from more practical concerns. However, as has been shown in section 3.3.1, it is possible to view interest management as separate from architecture, building upon the notion that saliency-measurement is independent of network structure. Thus, whilst research into developing networking hardware and software to support interest management remains significant, the next section chooses to build upon this separation in order to focus on the methods used by existing systems for the determination of salience, and limitations thereof.

3.4.2 Scope of Existing Solutions

Since the very first large scale multi-user virtual environments, such as NPSNET, the need for efficient interest management has rapidly emerged as an issue central to their development. However, whilst many different interest management approaches and their implementations have been well documented, very few directly acknowledge the relationship between interest management and human perception, despite an implicit link. Many systems such as MASSIVE employ spherical auras around clients as regions of interest as mentioned previously in this chapter, and whilst such methods have the advantage of being computationally simple, the assumption that distance is the only measure of perceptual relevance within a virtual environment is easily disproved by the simple example of a client stood behind the viewer. In this case, although the client may be near, they are not visible, and hence providing a high-resolution of updates regarding their position and behaviour is a waste of resources. Refinements that incorporate methods such as visibility culling such as the approach of Hosseini [HPG02] or partition the environment to overcome this problem only address part of this issue, since many studies into effects such as change blindness shown in Chapter 2 prove presence within field of view alone does not necessarily equate to perceptual importance.

As a result, the evaluation of the effects of interest management on a perceptual level is often performed superficially. Whilst certain techniques such as the environment partitioning used within NPSNET can lead to gains with no effect on the simulation perceived by the user, these methods alone remain inadequate for scenarios wherein the number of users within a locale exceeds the available bandwidth, and is consequently why these simulations remain limited in their total number of users. Fortunately, whilst existing interest management approaches generally give less regard to evaluation of perceptual effects within the environment, a considerable body of existing work in evaluating the success of interest management in terms of its impact on bandwidth consumption exists. Ultimately the goal of any interest management approach is to offer a reduction in bandwidth consumed by a distributed virtual environment, and many existing techniques have been evaluated within this context.

Furthermore, despite the aforementioned lack of perceptual consideration for many interest management techniques, a number of approaches such as that of Beeharee [BWH03] do make this direct analogy. Whilst the approach of Beeharee is valid, it attempts to achieve interest management perceived to be lossless, whereby the quality of the simulation to the user is unaffected by the interest management approach. This has two weaknesses; firstly that it leads to techniques which may be ineffective or untested in low-bandwidth situations, where as previously noted bandwidth may be insufficient to update all users, and secondly that the approach may result in network overloading in 'crowding' situations, where a large number of clients attempt to occupy a small virtual space. Using many techniques such as locales, this results in a large number of clients being flagged as relevant by the interest manager and can subsequently lead to network crashes – one example of such a problem, as noted by Minson [Min05], was demonstrated in Blizzard's online game "World Of Warcraft", when a group of players gathered in one place to protest a change in the game, and subsequently managed to deliberately induce a server crash.

Relating the work of Reddy to the theme of interest management presents a variety of issues for consideration. Whilst a direct parallel can be drawn between avatars and geometry (and thus, in theory, Reddy's approach could be applied directly), a number of key differences exist between interest management and geometry culling. Firstly, the amount of available bandwidth is finite in any distributed simulation, and in the context of a crowded virtual environment, perceptually lossless techniques are frequently inapplicable. Thus graceful degradation is of vital importance, and failure to provide the facilities for such degradation can result in an environment prone to crashes or instability. The previous example of players of the online game "World of Warcraft" can be taken as an example of both the weakness and need for solution. Secondly, since interest management works on typically the most perceptually important aspects of an environment – the avatars – criteria such as object size or angular velocity may be less significant than in the case of geometry depending on the nature of the simulation. Whereas Reddy effectively looks at a situation of information overload, and filters out a large amount of extraneous data, a typical interest management scenario involves attempting to make a more precise evaluation based upon a smaller set of objects.

Despite these weaknesses, such an approach has several notable characteristics, firstly in that it proves success of the interest management paradigm outside of these extreme circumstances. This may be sufficient within an environment wherein these situations may be controlled by simple rules and mechanics within the environment – only allowing a set number of people to enter a given area, for example. Additionally, it allows for a much more simple and definitive evaluation of the perceptual effects, simply by running two simulations both with and without the interest management applied and simply asking the sample whether they notice a difference. Thus whilst it is ideal to obtain a fuller set of results dealing with these extreme yet potentially common situations, this basic test is typically central to the evaluation of interest management.

To stress such limitations, consider an environment containing thousands of simultaneous users, operating alongside the bandwidth constraints implied by an Internet-based scenario. In this situation, grid or locale-based interest management offers the highest-level method for filtering. These methods are highly successful when the population of the virtual world is distributed and has minimal tendency to gather in small regions, as illustrated by NPSNET, alongside the other systems mentioned in Chapter 3 using grid-based interest management. The breakdown point occurs when clients gather within a single grid cell or locale, eventually viewing more clients simultaneously than the available bandwidth would allow. This is common behaviour in a variety of applications, such as troops gathering and engaging one another in a military simulation or game. In this case the only options within a grid or locale based method are the removal of clients which may potentially reside within the field of view (leading to degree-of-blindness problems as reported by VELVET, discussed previously) or a uniform decrease in update interval for each client to match available bandwidth. To provide an example of this bandwidth reduction concept, consider the case of 50 mutually visible clients (which represents a small crowd in real-world terms), each with a limited 50 kilobyte per second (kbps) downstream capacity. This restricts each client to obtaining a maximum of ~1kbps data on each visible client. Increasing the number of clients within the locale to 200 would decrease this maximum linearly to

0.25kbps, quadrupling the delay between updates to maintain the same bandwidth. The visible effects of this would be a steady degradation in performance, as avatars appear to jump from point to point between updates. More generally the update rate in seconds for each client may simply be expressed as:

$$U = \frac{B}{PN_c}$$

Where B represents the available bandwidth in kilobytes per second, P the packet size of each update in kilobytes and N_c the number of clients within the same locale. More sophisticated environments allowing for a variety of avatar states and levels of interaction necessitate larger values for P. Visual artefacts appear where this update rate U becomes sufficiently low that humans may perceive the motion as a series of individual movements rather than a single, smooth motion. This may in turn be related to the rendering and graphical capabilities of the environment; a value of U lower than the frame rate provided by the renderer will offer no visual improvement and effectively waste bandwidth. An optimum value may be considered as residing at the threshold for perceptual detection whilst being equal to or greater than the delay between rendered frames. Such a threshold is commonly given as being 35 frames per second (for a thorough analysis of the effects of framerate see [Red97b], or U ≈ 0.03 seconds / update.

Ignoring momentarily the fact that available bandwidth may fluctuate over time; B and P may be considered constant for any given client. Thus it would appear that purely the number of clients sharing the same grid cell as the viewer governs the performance of the environment. Yet considering the case of 200 users within this cell, and the capacity of human perception discussed in Chapter 2, it is impossible all these clients are ***perceptually relevant*** to the user simultaneously. Simply considering visibility alone, it is unlikely all the clients occupying a grid cell or locale are mutually visible to one another; taking a 90-degree field of view and random client distribution would imply only an average 25% are visible to any given client. This is exploited by many existing interest management systems to provide filtering by visibility culling, such as the method by Hosseini already discussed [HPG02]. Furthermore, clients will tend to interact most directly with the nearest visible clients. This assumption lies at the centre of the aura-based interest management system as first used by MASSIVE, which may operate within a locale, parallel to locale-based interest management (as exemplified by MASSIVE-3). Thus a general approximation of the current state of the art is an aura-based interest management paradigm operating within a single locale or grid cell, alongside visibility culling. In this case the aura-based interest management performs fine-grained filtering and as such may be seen as the component open to visual-attention based enhancement, since the locale-based component simply filters all users within the region, whilst the aura-based level identifies those most salient.

As has been discussed in this chapter, however, such a system continues to generate visual artefacts with distant clients being filtered out, and offers no resolution to degree-of-blindness. DeOliviera [Oli01] discusses the potential to resize the aura in response to changes in available bandwidth. Whilst this ameliorates the problem to some degree whilst providing scalability, it increases the probability of a degree-of-blindness situation (as described in Section 3.3.2.2) between two clients in a heterogeneous situation since a strong likelihood exists of different clients having different aura sizes. A more general solution is to provide the capacity for different update rates, and thus create a model with multiple auras, each corresponding to a different update rate. Available bandwidth may be distributed between auras such that more distant clients are assigned longer update intervals. This does however have the implication that upstream data must increase for all clients, since they need to support the provision of information on a variety of different levels of resolution simultaneously. Abrams et al. [AWZ98] illustrate the advantages of the provision of different levels of resolution in providing a high-fidelity to nearby clients whilst retaining the capacity to display more distant avatars. Adapting the previous equation for a situation where n levels of resolution exist:

$$B = \sum_0^n U_n P N_c$$

And thus the available bandwidth may be matched by considering the update rate U_n of each level of resolution in comparison to the number of available levels of resolution. Many basic techniques such as the aura approach provide an accommodation of perceptual considerations; the first implementation of an aura within the MASSIVE system bases itself upon the principle that all objects whose spherical auras intersect with that of the user are on relevance – essentially distance-based filtering. In the situation that a sufficiently low users / available bandwidth ratio exists, this method is entirely adequate. However, as environments aim to accommodate more users, with more sophisticated behaviours into smaller virtual areas, visibility and distance alone prove inadequate filters. The evidence of this are the subsequent attempts to refine approaches along a variety of avenues including multiple layers of relevance such as the approach of Abrams et al., the addition of locale-based filtering [PG00], and of course the explicit inclusion of perception [AWH03] in order to support larger numbers of users.

Fundamentally, however, the basis for perceptually oriented enhancement lies in the simple concept that aura-based management overlooks; what is *near* to a viewer is not necessarily what is *salient*, as shown in Chapter 2. A multiple-layer aura-based method such as that described above effectively implies salience decreases linearly with and in direct proportion to distance from the observer. Whilst in certain situations basing salience on proximity proves adequate, to suggest this is uniformly and consistently the case is easily disproved by the case of a user viewing a distant client. The key to deriving a means to enhance this aura-based system via perception lies in formulating a method by which to detect the most salient clients and hence provide the highest-fidelity possible to the viewer on these clients alone. Furthermore, this must be done efficiently and with minimal processing overhead, or the potential gains are limited by the impact on framerate. Additionally, since the IP address space and Internet architecture provides a finite number of available multicast addresses, clients must be grouped into levels of resolution rather than simply providing the ideal resolution to each client.

It is due to these restrictions a direct application of saliency mapping as discussed in Chapter 2 is unfavourable. Asides from the overheads of performance multiple pass image analysis on each frame, such an application would prove inefficient since only clients need be considered as salient objects (world geometry, for example, is replicated across clients within our model). This may be construed as leading to weaknesses in Beeharee's [AWH03] approach; when attempting to filter based on object properties the developer is faced with a difficult decision regarding implementation. Full-scene analysis on a per-pixel basis, of each rendered frame, is computationally highly expensive. This is compounded by the processor-intensive nature of many other aspects of virtual environments, such as physical modelling and the rendering process itself. An alternative approach which relies on object properties being inherently attached to objects (thus allowing an entity to be analysed purely by referring to its pre-defined properties) offers the potential to overcome such a problem, at the cost of accuracy. Such a method is adopted by Beeharee.

However, an important factor in current and future environment design is the notion of *composability*. Zyda [ZS00] defines composability as "...the ability to dynamically import models and behaviours, developed in one virtual environment, into another virtual environment." (p. 288). Composability is highly desirable since it circumvents the need to recreate content for every new environment, and allows a much greater degree of customisability and interaction. Requiring objects to possess pre-defined characteristics in order to perform interest management imposes restrictions on the composability of content. Conversely, restricting the interest management system to only use available properties constrains its effectiveness. Whilst certain characteristics of entities (primarily relative position within the environment) are ultimately required by all approaches, minimising the needed characteristics presents advantages in defining content and compatibility.

Given these drawbacks, consider the spotlight model described in Section 2.4.2. In its simplest form, a spotlight approach need only consider the positions of objects relative to the viewer, and hence is highly composable. As per the rendering approaches detailed in Section 3.2 the assumption can be made this spotlight exists near the centre of the visual display, and a precise focus and filtering thus performed accordingly. Since this would not require per-pixel analysis, or detailed analysis of object properties, it may be performed in a computationally inexpensive fashion. Moreover, since perceptual approaches such of that of Beeharee seek to define salience as a value independent of the hardware driving the interest management process (again, building upon the separation discussed in Section 3.3.1), the saliency results for a spotlight filter could potentially be summated with the results for a top-down and bottom-up analysis, providing further enhancement. In the next section, these notions are further explored towards the end of creating a model for spotlight-based interest management.

3.5 Summary

Whilst the volume of research on visual attention, together with the frequent ambiguity of results precludes any attempt to provide a 'definitive' model for visual attention-based interest management, a variety of potential enhancements to existing approaches may be suggested. By considering existing systems described in the previous section, alongside potential weaknesses in the current implementation of visual-attention based interest management as described by Beeharee, the previous section suggested several potential advantages of a spotlight-based approach. Such advantages are also emphasised in a 2005 paper [DW05b] (included in Appendix IV). In the next chapter, transcribing a spotlight model into an effective means for interest management is discussed, building upon the model described in Chapter 2. Advantages in both comparatively low processing overhead, and compatibility with the approach of Beeharee are outlined. The section concludes with a brief discussion of the context and scope of such an approach, and outlines the advantages of the model.

An additional task in forming a proposed interest management system is to identify the context and scope within which the proposal will operate. Most importantly, with reference to attempts to draw parallels to theories of perception based on real-world observation, a limit must be imposed on considering only virtual environments seeking to emulate the real world to some degree. Defining such a limit is partly an implementation issue, and thus described in more depth in Chapter 5. The remainder of the chapter discusses the potential benefits of a spotlight-based approach, and subsequently presents a novel method for interest management, which builds upon existing aura-based methods, providing perceptual refinement.

So far in this chapter, a review of the current state of the art regarding interest management has been presented, alongside existing applications of visual attention to virtual environments. In the previous section, the potential for a spotlight-based approach given existing weaknesses was suggested. Before discussing the potential for such a model further, it is useful to provide a summary of the previous chapters and re-emphasise the basis for such an approach.

As noted earlier, with respect to interest management, the current state of the art may be interpreted as a hybrid of grid and aura based approaches, with grid-based interest management operating as a large-scale, coarse resolution filter, whilst aura-based methods work within grid cells to further filter redundant data. Whilst current commercial environments typically filter purely based on proximity to the user, this simple approximation leads to easily demonstrable issues with overcrowding. Research has included refinements such as visibility culling, and Beeharee [BWH03] has adopted perceptual considerations in developing a bottom-up based means for interest management. Approaches that contradict, rather than complement this state of the art may expect considerable difficulty in presenting a universal approach effective enough to replace existing methods.

It is thus essential that a proposed approach consider its role within this current state of the art, with a view to complementing existing methods wherever possible. Perceptual refinement may be considered to operate on the aura-based level, since grid based filtering, such as that employed by NPSNET-V, is typically of a much coarser resolution and removes clients which are so distant from the user as not to be visible. Henceforth the method proposed in this thesis functions within the aura-based system, on the assumption that within an implementation context it would be employed within grid cells. For simplicity, it is possible to consider only a single grid cell, under the assumption the method could easily be duplicated within other cells.

The proposed spotlight-based approach to interest management builds upon a series of assumptions inherent to spotlight models of attention:

- Distance alone is not an accurate measure of relevance. The spotlight may be focussed on a distant object as well as nearby objects.

- That the zoom-lens notion, discussed in Chapter 2, implies the ability to focus on a small area with high resolution, or a large area with coarser resolution.

- That there is a central point of focus for the spotlight, around which exists a region in which change is perceived more readily.

It builds on these assumptions under the basis that such concepts may be modelled in a computationally inexpensive fashion. The assumption is also made that the implied depth of objects within a virtual environment, ultimately rendered to a two-dimensional display, may in fact be interpreted as a three-dimensional scene by a viewer. Whilst more sophisticated hardware may generate a true three-dimensional image, it is worthwhile exploring the potential within a typical two-dimensional display. Nagata [Nag91] suggests the capability to induce and enhance depth perception in two-dimensional images; and by assuming such depth perception it becomes possible to draw parallels between real-world studies of the spotlight of visual attention and the context of a user viewing a two-dimensional display. Such a relationship has also been demonstrated in other applications of perception to virtual environments, such as the aforementioned work of Funkhouser and Séquin [FS93].

One advantage when applying models of visual attention within interest management, as opposed to rendering, lies in the fact that the system need only consider entities, rather than the entire rendered scene. Whereas systems seeking to enhance rendering are forced to review the entire scene, a spotlight-based interest management approach may restrict itself to considering entities alone, since in the case attention is fixated on an area of terrain rather than an entity, the limitations of networking will not be perceived. A spotlight-based approach may offer particular advantages specifically where large, visible crowds of entities exist. This represents a common situation where existing methods break down as noted by Minson [Min05]. Approaches such as aura-based techniques, typically update more distant entities at lower resolution, and in the case of a viewer searching or interacting within a large crowd, offer inaccurate results, as evidenced by the breakdown described by Minson. By adopting a spotlight, it should become possible to more accurately identify scenarios wherein visual attention is deployed to a distant object or individual within the crowd, and significantly enhance perceived quality. This thesis now goes on in Chapter 4 to define a proposed technique that utilises the 'spotlight' approach, alongside a means of establishing focus from user input. This is subsequently combined with a means for including visual cues as a factor, to create a novel interest management paradigm for large-scale networked virtual environments.

4 A Spotlight-Based Interest Management Approach

In the previous chapter, the limitations of current virtual environments with respect to capacity and performance were related directly to the concept of interest management. When considering theories of visual attention alongside the current state of the art with respect to interest management, the potential for further consideration of attention, and its integration into the interest management process becomes apparent. This chapter seeks to define a proposed enhancement to interest management, which employs a spotlight model of attention alongside saliency mapping to create a means for interest management within large scale distributed virtual environments.

Having identified the potential for consideration of a spotlight model that builds upon the current state of the art, this chapter describes the creation of a theoretical model intended to provide interest management incorporating the spotlight concept. This consists of two major tasks. Firstly, it is necessary to create a means to generate saliency values for objects within the virtual world, and only those whose characteristics are affected by networked data (for example, the method should aim to consider user avatars, but not scenery which is replicated across all clients prior to run-time and is immutable). In many respects this process is identical to that used in perceptually oriented computer graphics to define saliency maps for images, although with the caveat that only certain objects within the world need by considered. Furthermore, rather than use a perceptual model identical to that deployed in graphics by researches such as Funkhouser [FS93] and Reddy [Red97], this chapter seeks to exploit the strengths of a spotlight-based approach in terms of computational simplicity and composability, and take advantage of the fact that only entities need be considered. The subsequent implementation, including practical concerns and psuedocode, forms the basis for Chapter 5.

4.1 Applying a Spotlight Model

The most widely available display technology creates a two-dimensional image on a flat surface by rendering a three-dimensional scene from a given viewpoint, rather than a true three-dimensional image. If such output methods are taken to be the focus of this study, a user's physical focus will be upon a single region of pixels, rather than a three dimensional object in space. This can be considered an advantage, though, provided the assumption is made that the object to which those pixels correspond is the focus of attention. Even so, this leaves us with the question of how to determine what a user is gazing upon with limited technology available. Approaching the problem firstly from a mechanical perspective, whereby the physical operations of the eye are examined to determine its focus, hardware capable of tracking eye movement based on corneal reflection has been available from as early as 1978 [CS78]. To date, though, such systems have little widespread use due to their limited application, and difficulty to use (some require physical contact with the eyeball, and hence the presence of an ophthalmologist, whilst accurate, non-invasive methods involve costly MRI or laser technology [KGH99]) Additionally, and more significantly, research within the multi-layered model of attention has shown that an object becomes the focus of attention *before* the saccade (eye movement) to look at it occurs [The93], and thus eye-tracking may not necessarily prove a panacea for the problem of determining the focus of a user's attention within a virtual environment. It has, however, been used successfully as a selection tool [Jac91]

To resolve this problem, it is possible to move away from the mechanical problem of determining the state of the eye, and look towards the concept of 'training' a user to *input* their attentional focus, as part of their interaction with the environment. Clearly, such input needs to be derived from normal user behaviour, rather than requiring the user to enter it explicitly, since the idea of a time consuming data entry for each object of focus is of little appeal. Hand [Han97] provides a good summary of interaction techniques within virtual environments, establishing that the majority of multi-user environments use keyboard and mouse interaction, and fall into a number of categories. Some, such as the Xerox 3D ROOMS system [MCR90], use mouse input to mimic head movement. Whilst in the

ROOMS system the mouse must be held over an icon and dragged to induce such movement, other systems, such as DIVE [CH93], have adopted and refined this approach to allow for a combination of keyboard input to control lateral movement (walking) whilst the mouse is used to look around. This successful combination has subsequently been utilised by an entire genre of games, and is particularly suited to military simulations where the client is required to both move and aim simultaneously.

Critically, this means a means for estimating user focus exists – if the 'head' is being moved by the mouse, then the user will tend to centre the focus of attention on the screen, provided they have sufficient experience with the system to move naturally. This simplification is made frequently when creating perceptually optimised rendering methods; the work of Reddy discussed in Section 3.2 makes such an assumption with successful results. This scenario applies particularly to the aforementioned applications of military simulations and games, wherein the target must be centred on the screen in order to hit. Regardless of application, it is plausible to assume that some approximation of attention may be derived from the orientation of the user's viewpoint. Thus the interaction device can be seen as approximating overt attention (as described in Section 2.3), whilst covert attention is modelled by analysing the objects surrounding the focal point in a spotlight-based fashion.

Change blindness can be construed as having an important implication; that the focus of a user (and hence interest manager) is likely to be upon a single object at any given time. However, this does not necessarily simplify the problem, since this attention may shift extremely rapidly, required a substantial network overhead in breaking and creating new connections to clients determined to be of interest. Thus practicality requires that the objects *likely* to be the foci of attention remain specified at high-resolution, whilst the user's actual attention rapidly moves between them. Furthermore, the necessity emerges to consider precisely how to extrapolate this target of attention given available user input devices.

4.2 Defining the Spotlight as a Field

The proposed approach uses a multi-layered field to describe the area around a user in terms of likely salience. This has several advantages over a more direct approach that seeks to define interest between two objects as a Boolean expression. Firstly, the approach is scalable, since the weighting of the perceptual field at a given point may be compared to a series of threshold values to determine interest group subscription. These thresholds may be modified on the fly to maintain a constant level of bandwidth consumption or match available bandwidth more closely. Secondly, it allows for additional weightings to be added easily, as a combined field.

To explain this description in more detail, let us first examine the simple case of a radial aura around a user, as shown in Figure 23. For the purposes of this example the client's avatar in virtual space is referred to as the *user* and the client being tested for relevance as the *entity*. Entities differ from the rest of the world geometry since their behaviour is governed by data received via the network. Many theories of visual attention recognise a tendency in humans to define a viewed scene as a collection of objects [Fel03], which lends itself well to the entity-based, hierarchic implementations of the majority of virtual environments. The known values in this case are the absolute position of the user relative to the universe centre, and the absolute position of the object (this position may be inaccurate depending on latency, although the approach remains valid).

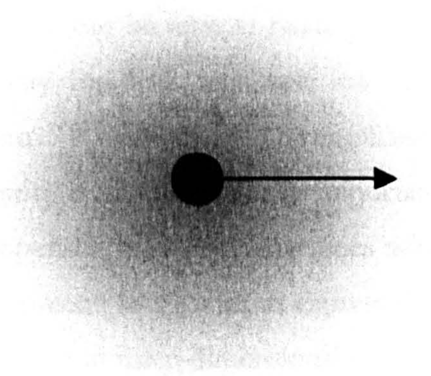


Figure 23: A radial aura.

In this case a simple description of perceptual relevance is assumed as decreasing with distance. The rate of decrease may be described as linear or exponential; since typically entities extremely close to the user have a very high level of relevance (primarily due to collision detection), an exponential approach appears the most valid. Exploring the effects of varying this decrease within the final system may prove interesting, and offer further potential for refinement.

The simplest case of exponential decrease, the inverse square law, is observed in nature for both gravitational and electrostatic forces. Adapting it to our perceptual field, as an approximation loosely based around a Weber-Fechnerian response (as described in Section 2.2.) to perceptual stimuli around the focus of the spotlight, it can be stated that our saliency, or *perceptual-relevance value* (P-value) is given by:

$$P = \frac{R_{obj}}{r^2}$$

Where R is a *relevance value* inherent to the object (this allows for further refinement as objects may be given implicit relevance values – to take an example of a military simulation, it may be preferable to give friendly troops lower relevance as the user is unlikely to be carefully aiming at them). These can be seen as relating back to the work of Beeharee, in that they provide the capacity for both high and low level object-based measures of perceptual relevance. The term "P-value" is used to emphasise the fact that the value is *not* a true probability of the entity being perceptually important; rather it is a *relative* value, which must be used in comparison to the other entities within the environment. Due to the nature of a virtual environment, it is useful to express this in terms of the positions of the two entities (user and object) in three-dimensional Cartesian coordinates, in a universe with centre at 0,0,0. Hence this becomes:

$$P = \frac{R_{obj}}{r^2} = \frac{R_{obj}}{\left(\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}\right)^2} = \frac{R_{obj}}{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Where x_1 is the x component of the vector describing the user's position in space, x_2 the x component of the entity position, and so on through y and z co-ordinates. Now consider taking into account not only the space around the user, but also the point around their visual focus. This is illustrated in Figure 24.

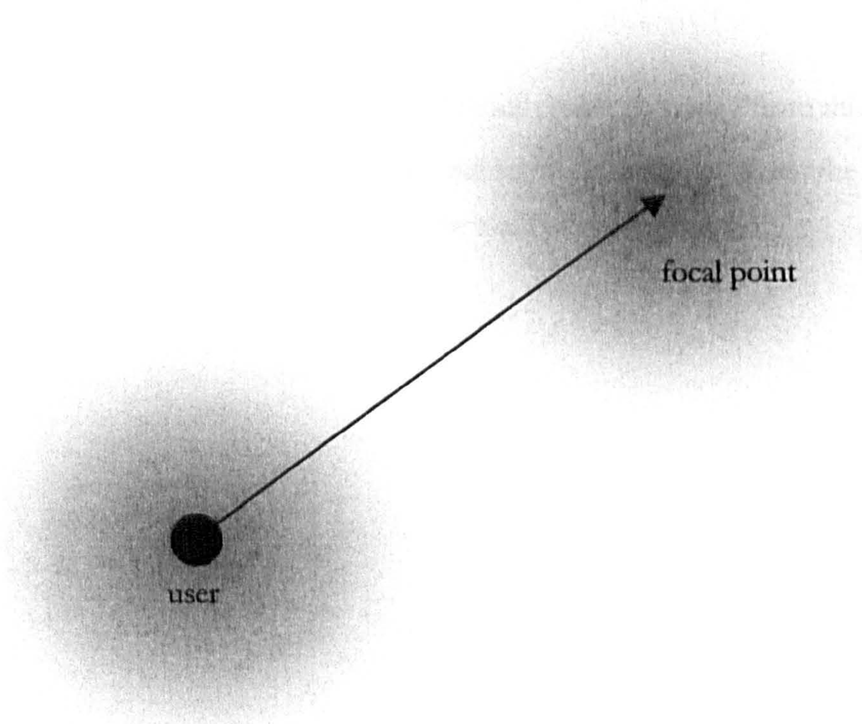


Figure 24: Including focus. Darker regions illustrate higher P-values.

In this case, another inverse square law is applied around the focal point, and the average the two fields taken. Since the environment is in 3 dimensions, the distance between user and focal point is again taken as a vector with x, y and z components. The opportunity is also taken to introduce a relevance weighting for the focal point, R_{focus}, which permits the importance of the focus compared to the space directly around the client to be scaled – e.g. a value of 0.5 would result in a heavy bias in importance to the space directly around the user, whilst a value of 2 would give the focus predominance.

Thus:

$$P = \frac{\left(\dfrac{R_{obj}}{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} + \dfrac{R_{focus} R_{obj}}{(x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2} \right)}{2}$$

$$\therefore P = \frac{2R_{obj}}{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} + \frac{2R_{focus} R_{obj}}{(x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2} \quad \textbf{(1)}$$

Where the centre of focus is at (x_3, y_3, z_3). The obvious challenge arising from this is the need to establish this point of focus based on information available from the user's interaction within the virtual environment. This is the topic of the next section.

4.3 Establishing Focus

In order to provide a rough approximation of a user's focus, an approach is adopted that utilises the user's input response to the scene, as discussed previously. By casting a pick segment along a normal from the centre of the visible display and into the scene, it can easily be determined which object corresponds to the pixel at the centre of the screen, but this method alone is insufficient, since frequently this point will not correspond to another user's entity but scene geometry. Hence a function is used to determine the object most *likely* to be the centre of focus based on their proximity to the user and distance from the pick segment. This is necessary to accommodate the fact that, for example, a very distant object within 10m of the segment is far more likely to be the focus than a very near object 5m away. The net result resembles a conical segment projected from the user – similar, but narrower, than their field of view.

Establishing the distance of an entity from the pick segment itself is a deceptively complex problem, given the need for computational efficiency. An illustration of the situation is shown in Figure 25.

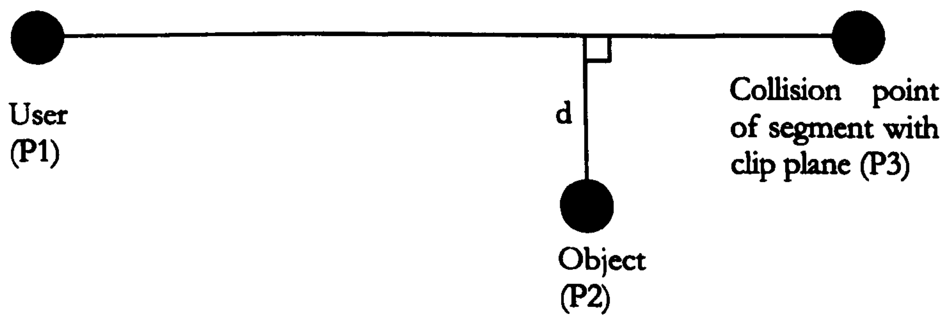


Figure 25: Distance of an Object from a Pick Segment

Taking again the point of the client as (x_1, y_1, z_1), position of the object as (x_2, y_2, z_2) and in this case the collision point with the universe bounds as (x_3, y_3, z_3), the vector between client and bounds can be expressed as:

$$\underline{V_{cb}} = \begin{bmatrix} x_1 + (x_3 - x_1)t \\ y_1 + (y_3 - y_1)t \\ z_1 + (z_3 - z_1)t \end{bmatrix}$$

Thus the distance between a point on the line at t and the object is given as:

$$d^2 = ((x_1 - x_2) + (x_3 - x_1)t)^2 + ((y_1 - y_2) + (y_3 - y_1)t)^2 + ((z_1 - z_2) + (z_3 - z_1)t)^2 \qquad \textbf{(2)}$$

Where:

$$t = -\frac{(P_1 - P_2) \bullet (P_3 - P_1)}{(P_3 - P_1)^2} \qquad \textbf{(3)}$$

The minimum distance may be found by differentiation; since at the minimum point:

$$\frac{d(d)}{dt} = 0$$

Differentiating the above and solving for t gives:

$$\begin{aligned} d^2 &= ((x_1 - x_2) + (x_3 - x_1)t)^2 + ((y_1 - y_2) + (y_3 - y_1)t)^2 + ((z_1 - z_2) + (z_3 - z_1)t)^2 \\ &+ 2t((x_3 - x_1)(x_1 - x_2) + (y_3 - y_1)(y_1 - y_2) + (z_3 - z_1)(z_1 - z_2)) \\ &+ t^2((x_3 - x_1)^2 + (y_3 - y_1)^2 + (z_3 - z_1)^2) \end{aligned}$$

(Which can be further simplified using cross/quad products; however since the equation will be transposed into an algorithm it remains simpler and more efficient to leave it in this inelegant state)

Combining this with the coefficients a and b, intended to allow weighting to control more precisely the relationship between proximity to user and distance from segment (such that in the condition a = b the aura around the user and aura around the focus are of equal intensity), gives us a final value for the probability of an object being the focus;

$$P_{focus} = a\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} + b\sqrt{\begin{aligned}&((x_1 - x_2) + (x_3 - x_1)t)^2 + ((y_1 - y_2) + (y_3 - y_1)t)^2 + ((z_1 - z_2) + (z_3 - z_1)t)^2 \\ &+ 2t((x_3 - x_1)(x_1 - x_2) + (y_3 - y_1)(y_1 - y_2) + (z_3 - z_1)(z_1 - z_2)) \\ &+ t^2((x_3 - x_1)^2 + (y_3 - y_1)^2 + (z_3 - z_1)^2)\end{aligned}} \quad \textbf{(4)}$$

Subsequent to the target object probabilities being established, the focal point is shifted to match the position of the object with the highest probability of being the focus. It is hoped that users will be able to adapt their behaviour to shift their viewpoint to the point of interest to mirror a natural, overt response to stimuli (thus allowing the system to continue to make approximations regarding the deployment of covert attention), although input devices often impose restrictions upon the speed and accuracy with which users can shift their view.

In summary, the following have been derived:

- An equation (4) stating the probability of an object being the focus of a user within a virtual environment, based on its proximity to a pick segment projected along the centre of the users field of view;
- An equation (1) resulting a perceptual relevance values for objects around a user, based on a combination of an aura around the user and an aura around their focus as determined by (4);
- A group of coefficients used to control the weightings of objects and properties within these equations. These fall into 2 groups; those attached to objects (R_{obj}), and those which are constant regardless of object (R_{focus}, a, b). The determination of these coefficients is discussed in the subsequent chapter.

The only other values within the equations are derived from the x, y and z values of points in space, which are obtained at run time. The equations are sufficiently succinct to run with a minimum of iteration and hence processing overhead. However, it will also be important to note they at most require execution once per update arrival via the network, which is substantially less frequent than frame rate. Coupled with dead reckoning, a 100ms update rate, which represents a typical minimum afforded by current Internet hardware, results in 10 updates per second, whilst interactive framerates are more in the region of 30 updates per second. In addition, the cost of forming new connections between clients may suggest an even slower update rate would offer improved performance with little decrease in quality. This is discussed in more depth in the next chapter, and explored through the experiments described in Chapter 6.

4.4 Summary and Illustrations of Fields

In this chapter, a spotlight model has been presented via two sets of field equations. The model presented in Section 4.3 attempts to describe the likelihood of an entity being the focus of a user's spotlight of attention by considering relative distance from the user, and relative distance from the centre of the screen. Figure 26 illustrates the resulting field in two-dimensions. In the illustration, the client is shown as a red dot, with their direction of view denoted by a red arrow. A simple filter eliminates the chance of clients outside the visual arc from being considered (otherwise the distance from line and distance from client equations presented in Section 4.3 would be capable of incorrectly considering clients behind the viewer). This could potentially be refined with additional visibility filters such as that of Hosseini [HPG02], but to provide a purer evaluation of the proposed method a simpler approach is used that eliminates entities outside the visual arc. The regions of highest probability are shaded white, fading to black for regions of zero probability. Of note is the broadening of the field near the user, as the radial distance from viewer component combines with the distance from the line. In effect, the focus is modelled as a three-dimensional cone, whose form is defined by the a and b coefficients as described previously. Determining the values of these coefficients is described in Section 5.6.1.

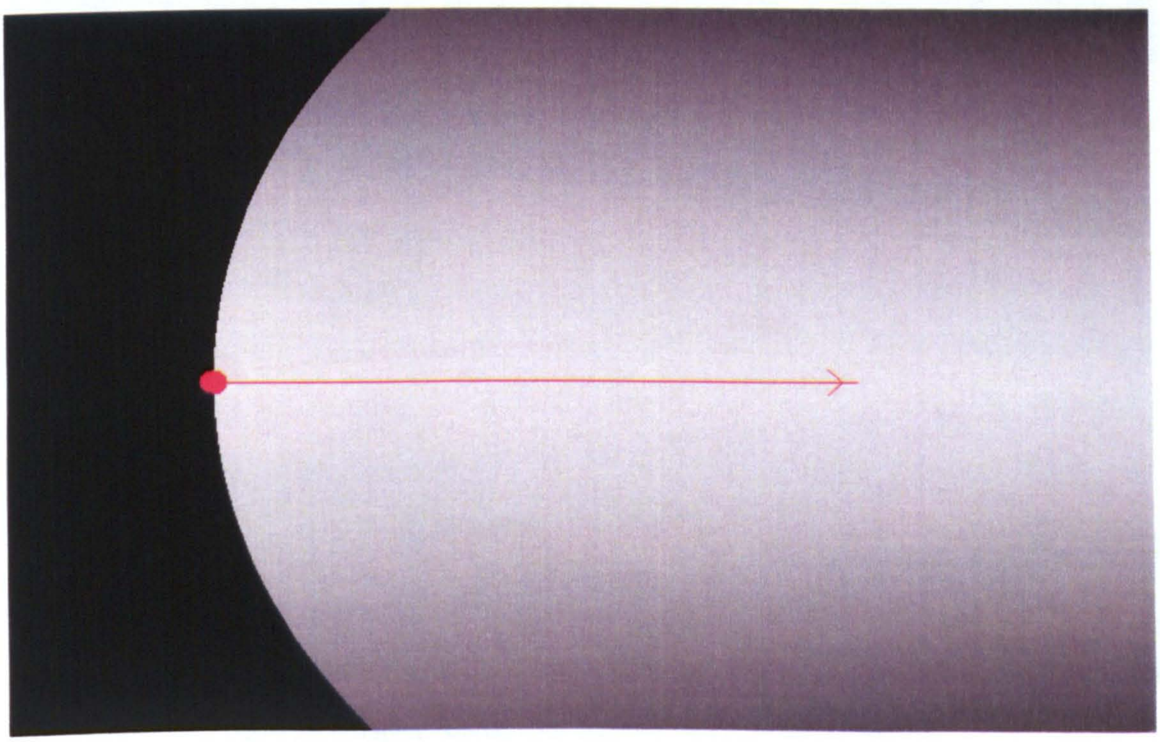


Figure 26: Focal field generated by the proposed technique.

With focus determined and placed upon a selected entity (with the highest probability of being focus), the second field is then established by combining radial auras around both client and focus. Figure 27 illustrates this field. The viewer is marked on the left, with an arrow indicating direction of focus to the selected client on the right. This is, in effect, our three-dimensional spotlight of attention. The zoom lens is accommodated in the fact more distant – and thus, when transcribed to the two-dimensional display, smaller – clients have the same rate of attenuation, which when mapped to the two dimensional display proves a more tightly focussed aura. Some weighting is still given to objects nearer the client in order to accommodate the possibility of them rapidly moving into view, or attempting to interact with the user. However, unlike existing approaches, the majority of resources are shifted around the focal target. The form of this field is controlled by the ratio of R_{focus} to R_{obj} – an increase in R_{focus} results in the focal target being given more emphasis, whilst conversely, increasing R_{obj} results in the area around the client generating higher saliency values. The values themselves may potentially be computed by analysing top-down and bottom-up characteristics of the focal selection; though this falls into the realm of existing works and is thus beyond the scope of this method. The illustration shown is for the case $R_{focus} = 5R_{obj}$.

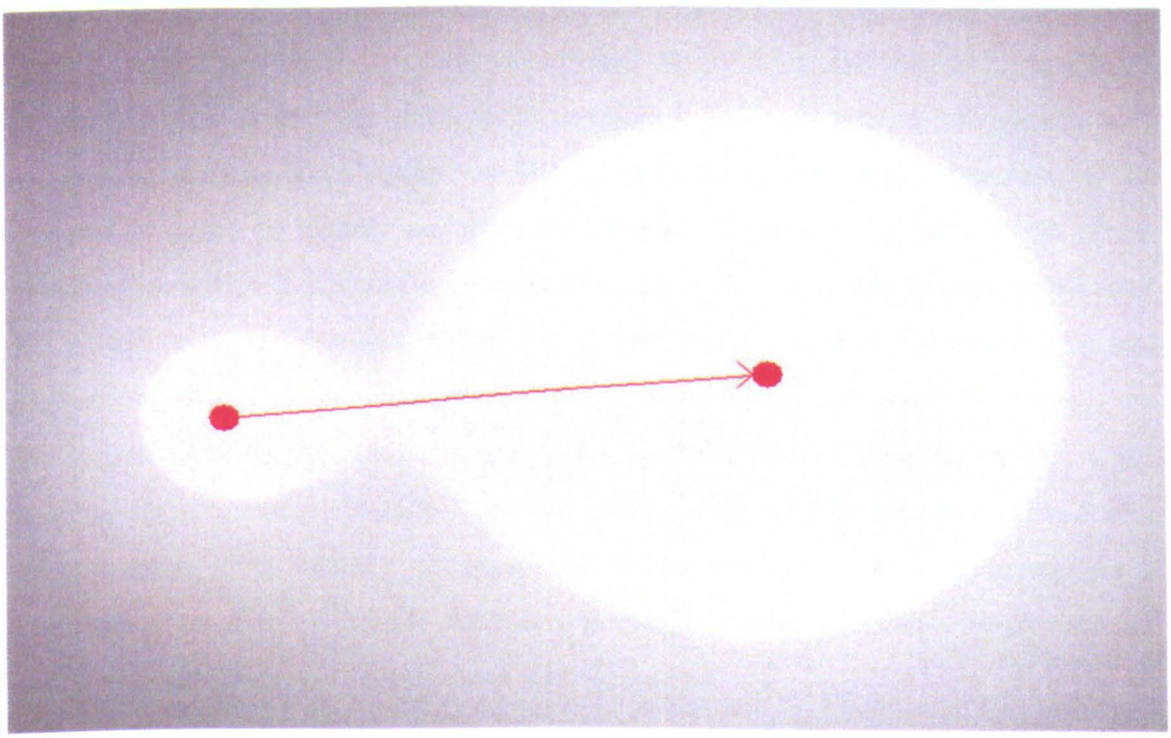


Figure 27: Perceptual (saliency) field generated by the proposed technique.

5 Implementation and Testing Methodology

This chapter details the implementation of the method proposed in the previous chapter. In the first section these challenges are outlined in more detail, leading towards the conclusion that the ideal methodological approach to evaluation involves the creation of an environment custom-built purely to perform experiments involving interest management and human perception. Such a system is then briefly outlined with specific regard to novel features and general design issues, and the resulting implementation illustrated. It is the intent throughout this section to avoid in-depth discussion of technical issues and instead focus heavily on the concepts behind the system. Separating the conceptual side of an IM technique from practical issues is not necessarily a simple task, and if done naively can result in questions regarding the impact practical concerns should have on design and raise doubts regarding the validity of experiments that choose to ignore them. Thus this section places particular emphasis on cases where practical concerns have directly impacted the effectiveness of the proposal described in Chapter 4, such as the limitations imposed by a multicast architecture. The chapter concludes by means of a series of illustrated examples providing evidence the approach functions as described in Chapter 4. This leads to the more rigorous, comparative evaluation of the technique, which forms the basis for Chapter 6.

5.1 Introduction

In order to provide an adequate analysis of the method discussed in the previous chapter, it becomes necessary to provide some platform upon which to directly compare interest management paradigms. One potential approach is to look at implementing the proposed method within an existing environment, and directly compare it with the current method. A platform such as DIVE [FS98], mentioned in Chapter 3, would be an ideal candidate in this case since it has been extended frequently to allow modification and metrics, a direct consequence of its research function. Such an approach offers the benefits of a rapid development timeframe, and the provision of a robust system for the many other necessities of a virtual environment, such as content development and user interaction.

Whilst adopting such a platform for implementation thus has advantages, there are also several significant drawbacks. Firstly, restrictions exist within the toolkit with regards to networking and modelling of visual attention. They may be overcome by rewriting sections of code, yet this raises questions regarding the impact on the existing interest management process, and the fairness of any subsequent comparisons. More significantly, however, the current interest management techniques applied using DIVE (or any other significant toolkit), are often proprietary and difficult to adapt. It becomes apparent that in any case, some effort must be taken to replicate the current state of the art in comparison to the proposal in order to identify any genuine successes or failures. In the context of the proposed method, it is possible to consider this state of the art in very specific terms; since the spotlight may be integrated with grid, aura, and even the multimodal model proposed by Beeharee [BWH03], the only aspect of this state of the art that needs emulating is the component incompatible with the proposed approach. This component is the model equating relevance with distance from the viewer alone, and hence the spherical aura approximation. Thus whilst a comprehensive implementation of existing approaches is unnecessary, creating an interest management system with the flexibility for this specific aspect to by modified and experimented with is essential.

It is the need to provide this direct and specific comparison that leads to the notion of creating a test environment purely designed to be capable of comparing interest management paradigms. Rather than comparing separate implementations, whose performance may be biased by unrelated factors such as efficiency in coding or rendering engine overheads, we may seek to extract the *concept* behind the current state of the art in its purest form, and then compare it directly to the proposed method. It becomes vital in this case to provide evidence that the proposal would be interchangeable with the current state of the art in existing implementations; otherwise the case may be argued that the proposal is impractical.

With particular regard to the perceptual weighting of the proposed technique, it proves inadequate to provide a simple comparison on bandwidth consumption between the proposed and current techniques. Whilst certainly, a user's inability to detect change alongside a reduction in bandwidth is a strong measure of success, and frequently used in experiments such as those by Beeharee, a more rigorous study would imply an attempt to provide a metric whereby the precise threshold for change detection occurs. Understanding such a threshold offers the potential to allow for a subsequent optimal configuration of any proposed technique. Additionally, any system intended for a heterogeneous, Internet-based scenario must be capable of dealing with fluctuation in available bandwidth. In this case simply standing methods alongside one another with identical bandwidths is insufficient.

Another, more practical, concern lies with the feasibility of evaluating methods intended to provide massive-scale interest management when practicality only affords a handful of test subjects and hardware. It immediately becomes apparent that in order to provide a thorough and realistic evaluation as possible, the functionality to add simulated users to the virtual world is a necessity. Whilst several existing environments including DIVE offer some means for simulating additional users, this simulation affords little or no consideration for the specific purpose of testing interest management. Thus simulated clients are manipulated locally, and removed from the network data stream where possible to provide better performance. Whilst this is a logical approach when attempting to create an efficient environment, it has consequences for the evaluation of interest management when simulated clients exist outside of the interest manager's control.

Creating a custom system from scratch affords the ability to tailor rendering performance to available hardware, and thus ensure framerate has a measurable impact on testing, an impact that may ideally be eliminated. It also allows for the environment content to be configured as providing as fair and thorough a test as possible – e.g. geometry may impact the effectiveness of an interest manager by obscuring certain clients and thus should remain constant between tests. Additional metrics may also be implemented transparently without fear of immeasurable overheads indirectly introduced by a specific toolkit. Note that, so far, the precise experiments used to judge the effectiveness of a given interest management paradigm remain undefined. In creating the test environment we focus purely on providing a platform within which such paradigms may easily be implemented. By standing methods alongside one another on a conceptual level (as opposed to comparing practical implementations), it becomes possible to evaluate them more directly. Thus the first step in the methodology used for evaluation is to create this test environment, then, once implemented, to consider the approach used to evaluate interest management approaches directly alongside one another.

5.2 Context and Scope

In order to create an interest management system centred upon visual attention, it is important to define the context and scope of the virtual environments to which the system would be applied. The term 'virtual environment' encompasses a massive range of systems with many diverse applications. In order to narrow the scope, and referring back to the examples of environments given earlier in this chapter, two primary assumptions are made about the nature of the environment:

1. The target environment is a *large-scale, heterogeneous, Internet-based* system. This assumption has a wide range of potential implications. Firstly with regards to scale; not only is it necessary to accommodate a large number of users (in this regard we assume 'large-scale' to imply users numbering in the thousands, since this is the current limitation of majority of Internet-based systems), but it also implies a certain scalability in providing continued acceptable levels of service as the number of users and their distribution within the environment fluctuates. Secondly, the concept of heterogeneity requires we consider clients with varying bandwidth and latency capacities within the environment. Whereas a local area network-based system may assume all users have similar hardware and network service, providing a heterogeneous system necessitates including the capacity to scale to provide the optimum quality of service to each user. Finally, the Internet-based criterion imposes a number of practical constraints upon the environment. It may be argued the term alone implies heterogeneity; and certainly the requirements of a heterogeneous environment are important in any Internet-based scenario. Moreover, it forces consideration of available bandwidth on a more quantifiable level, along with latency.

2. The target environment aims to provide a *realistic* virtual world. Referring back to Chapter 2, and the notions of believability and immersion in virtual environments, this assumption encourages the stipulation of further criteria which appear common throughout virtual environments which seek to mirror some aspect of the real-world:

   - Users are represented by a single 'avatar' – a graphical embodiment of their presence within the virtual world. This is typically a three-dimensional model of a human, although more primitive environments often use simpler avatars in order to improve performance and lessen development time, or to suit a specific application.

   - A physical model exists within the environment, which seeks to approximate that within the real world. This generally implies avatars and other objects in the world exist on a horizontal plane, and are affected by forces such as gravity.

   - Graphical content seeks to emulate the real world. Subsequently the environment can be assumed to be a region of terrain, or building, or other locale.

   - Network data pertains only to avatars, and their behaviour. Certain environments such as ActiveWorlds [AW06] seek to transfer information regarding world geometry and other aspects of the environment at run time. A deliberate restriction in scope is made to consider only avatars, primarily since this is the case in the majority of Internet-based environments where it is ideal to cache world geometry on each client and thus reserve all available bandwidth for avatar interaction.

To provide some grounds for comparison, environments exist wherein none of the above rules apply. These often have more specific applications; collaborative working, for example, where users simultaneously work on a prototype without the presence of avatars or indeed any view of other users bar their impact on the prototype. In such situations providing realistic physical and graphical constraints offers no benefit and thus is absent. Providing interest management for these systems represents a significantly different challenge and is thus deemed beyond the scope of this project.

These two assumptions do not necessarily preclude the application of the proposal to environments outside the scope, although imposing such limitations offers advantages in providing clearer grounds for evaluation, since we may consider the success of the method in a more specific application. Additionally, they allow for the underlying network architecture available for the method to be approximated more precisely, alongside giving a clear definition of the information available to the interest manager based upon the content of the environment.

5.3 Criteria for a Test Environment

Drawing upon the previous section, the conclusion is reached that an environment created solely for the purpose of comparing interest management approaches would offer the optimum capability for experiment and evaluation. Given this basis, the criteria with which such an environment needs to comply, in order to provide a fair and accurate platform for evaluation, warrant some discussion.

In the particular context of interest management and its evaluation, the facility must be provided to perform experiments based upon measures of networking demand and human perception. Due to the qualitative nature of perceptual research, the implication of the involvement of real-world users within the environment must be considered. Primarily this involves developing the facilities common to many environments - sufficient graphical and simulation functionality must be provided to allow users to distinguish their environment and other users. Referring back to the notions of immersion and believability discussed in Chapter 2, it is important to provide a simulation sophisticated enough to allow subjects to interact as would be expected in a typical large-scale virtual environment. This requirement can be further broken down into two main categories, as per the concept of immersion discussed by Kim et al.

- *Presentation* – The graphical content needs to be adequately detailed to offer users some degree of immersion within the environment. Critically, it must be recognisable as a facsimile of the real world.

- *Interaction* – A limited interaction model must allow for basic movement of the user avatar and viewpoint.

These aspects of the environment must be sufficiently developed in order to create the *immersion* defined by Kim et al., which represents a primary goal of the environments the proposed method seeks to serve. Failure to develop either aspect adequately may conflict with a perceptually oriented method as users fail to exhibit perceptual traits due to irreconcilable differences between the environment and real world.

A more practical limit is imposed by hardware; in particular it must be possible to ensure a framerate sufficiently high to avoid interference with perceptual experiments. The point at which this occurs is commonly termed the *"flicker fusion threshold"* [Pro13]. Typically, this is taken as 36 frames / second since higher values are generally indistinguishable by humans, although the precise value has remained a subject for debate since the advent of early motion pictures, and shown to be related to a large number of factors, such as alertness [Par82]. This relates back to the interest management implementation; the true minimum for the framerate to ensure a fair test must equal the minimum interval between network updates. Ideally this should also be above the flicker fusion threshold. Monitoring this leads to another essential criteria – the need to provide the capacity for clear and transparent performance metrics of all aspects of the system. Specifically concerning interest management; the ability to seamlessly change the interest manager on the fly offers the potential for change-detection type experiments. In general interest management must be implemented in a dynamic fashion to accommodate the necessity to evaluate various approaches within the environment. With the aims established, the next section details the development of such a system in more detail.

5.4 Developing a Test Environment

In the previous section, we defined key criteria as being presentation, interaction, the facility to provide metrics and a dynamic approach to interest management.

By carefully considering the location of the interest manager within the system and network infrastructure, it becomes possible to create a design whereby the concept behind the interest manager may be changed whilst keeping the networking layer isolated. Doing so requires a design emphasising the separation between networking and rendering; in effect compartmentalising the system such that components may be interchanged seamlessly. Existing virtual environments, such as those described in Chapter 3, may be seen to contain several common components as shown in Figure 28.

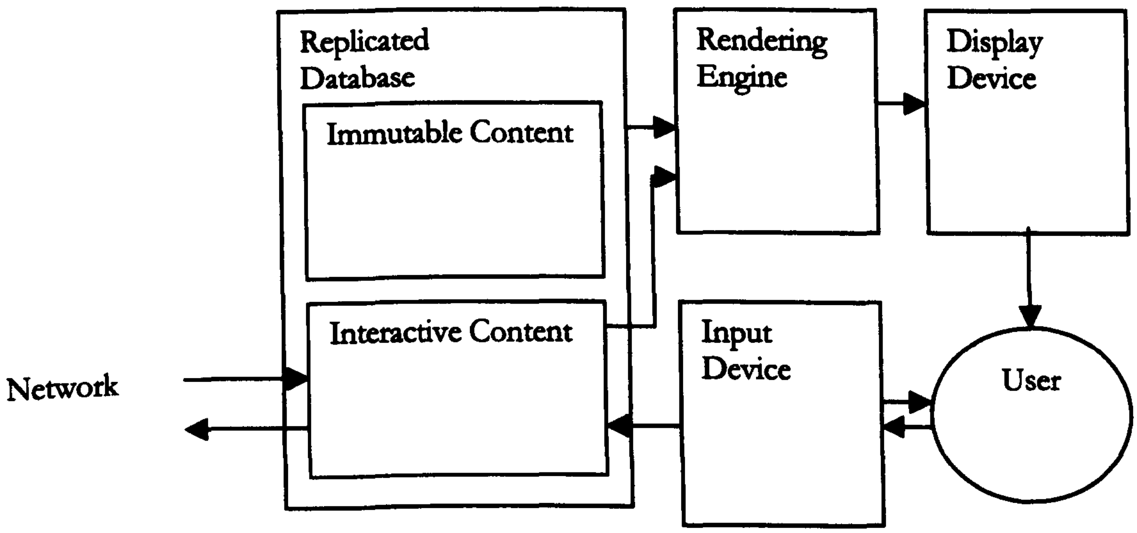


Figure 28: Model for a Generic Distributed Virtual Environment

Central to the environment is a store of data commonly referred to as a *replicated database*. Under ideal conditions, this should be simultaneously identical for each user within the VE. In reality, network latency and bandwidth restrictions result in portions of the database being inconsistent. A common distinction can also be made between *immutable* and *interactive* content. Immutable content remains the same for all clients throughout the life of the simulation. This typically represents the majority of the data in kilobyte terms; for example, the geometry and textures required to represent a vehicle or region of terrain. Whilst more sophisticated environments aim to integrate immutable and interactive content more closely (for example, users being able to drive the vehicle across terrain with subsequent tyre tracks left on the terrain and damage to the vehicle), many existing environments incorporate a large amount of immutable content. Interactive content, by comparison, consists of the aspects of the environment that may be modified by the user. Whilst the geometry and textures for the aforementioned vehicle may remain static, the location and state of the vehicle may not. For interactive content to appear consistent to all users, it must be distributed via the network. The most common example of interactive content is user avatar position, orientation, and state (although, once again, the geometry and texturing of these avatars are commonly immutable).

Another feature common to all virtual environments is a rendering system that serves to translate the raw information in the replicated database to a rendered frame. This is in turn displayed to a user who responds via an input device to modify interactive content within the replicated database and thus interact with the simulation. Such interaction typically consists of reorientation or movement of their associated avatar within the virtual world, alongside application-specific actions such as opening doors or communicating with other avatars. From this generic model, it becomes evident that the database aspects of the environment may be effectively separated from the rendering process. Interest management may thus be seen as lying between the replicated database and network; utilising information within the replicated database to best determine the most critical updates to propagate and request via the network.

Referring back to the environments briefly described in Chapter 3, a peer-to-peer design is advantageous in representing a progression towards larger scale, Internet-based environments. Whilst certain existing systems successfully adopt a client-server approach, these systems scale poorly in comparison to peer-to-peer approaches as illustrated by DIVE [FS98] (a successful peer-to-peer environment), as client-server approaches prove ultimately limited by server capacity and centralised network load. Whilst it remains possible to consider interest management on a conceptual level independently of underlying network architecture, in practice an implicit link exists between networking and interest management technologies - such a link is evidenced by the enhancements to interest management made possible via multicasting. In the case of this implementation, a logical decision is made to adopt a peer-to-peer architecture on the basis that such an approach will form the foundation for future large-scale Internet-based environments. The decentralised load is also beneficial when considering the hardware available for testing purposes.

5.5 Overview

Figure 29 on p.122 shows the framework and data flow for the proposed architecture. A replicated database approach is adopted, which emulates the approach common to existing environments as illustrated in the previous section. The system is created using a combination of Java and Java3D, with the latter API being used to provide a simple and efficient means to create a rendering engine for the simulation, allowing design and development emphasis to be given to other areas of the system. A replicated, immutable database provides a terrain mesh and other geometry used within the simulation. The local client data store shown in Figure 29 represents interactive components alone, in the case of this simple environment this being solely the position and orientation of the user and other clients within the terrain.

A discrete configuration utility used prior to run time affords some degree of dynamicism in supporting varying numbers of clients and interest management paradigms, plus affording some scalability within the rendering engine to enable it to run on a wider range of hardware (by providing both windowed and full-screen modes at various levels of resolution). Multithreading is used extensively at run-time, with a number of modules running independently:

- The Java3D rendering engine utilises a multithreaded approach within the API. This is isolated from the other aspects of the system, with updates on the locations of avatars and viewpoint orientation being read from the local world database and passed to the renderer each frame.

- Additional clients, used to simulate overcrowding, are designed to run in individual threads. These client simulation and emulation engines, discussed later in this chapter, are intended to run either alongside a real-world client, or independently on a stand-alone machine to distribute the simulation load.

- User avatar control interfaces directly with the local database to update the user's avatar position and orientation, with subsequent effects on rendered frames. Again, the goal in providing user interaction is to emulate the functionality common to existing VEs in order to facilitate as generic an evaluation as possible.

The outgoing data stream handles low-level networking tasks such as packet construction and scheduling, alongside multicast group assignments as specified by the interest manager. Similarly, the incoming data stream reads and processes packets as they arrive, performing subsequent updates to the replicated world database.

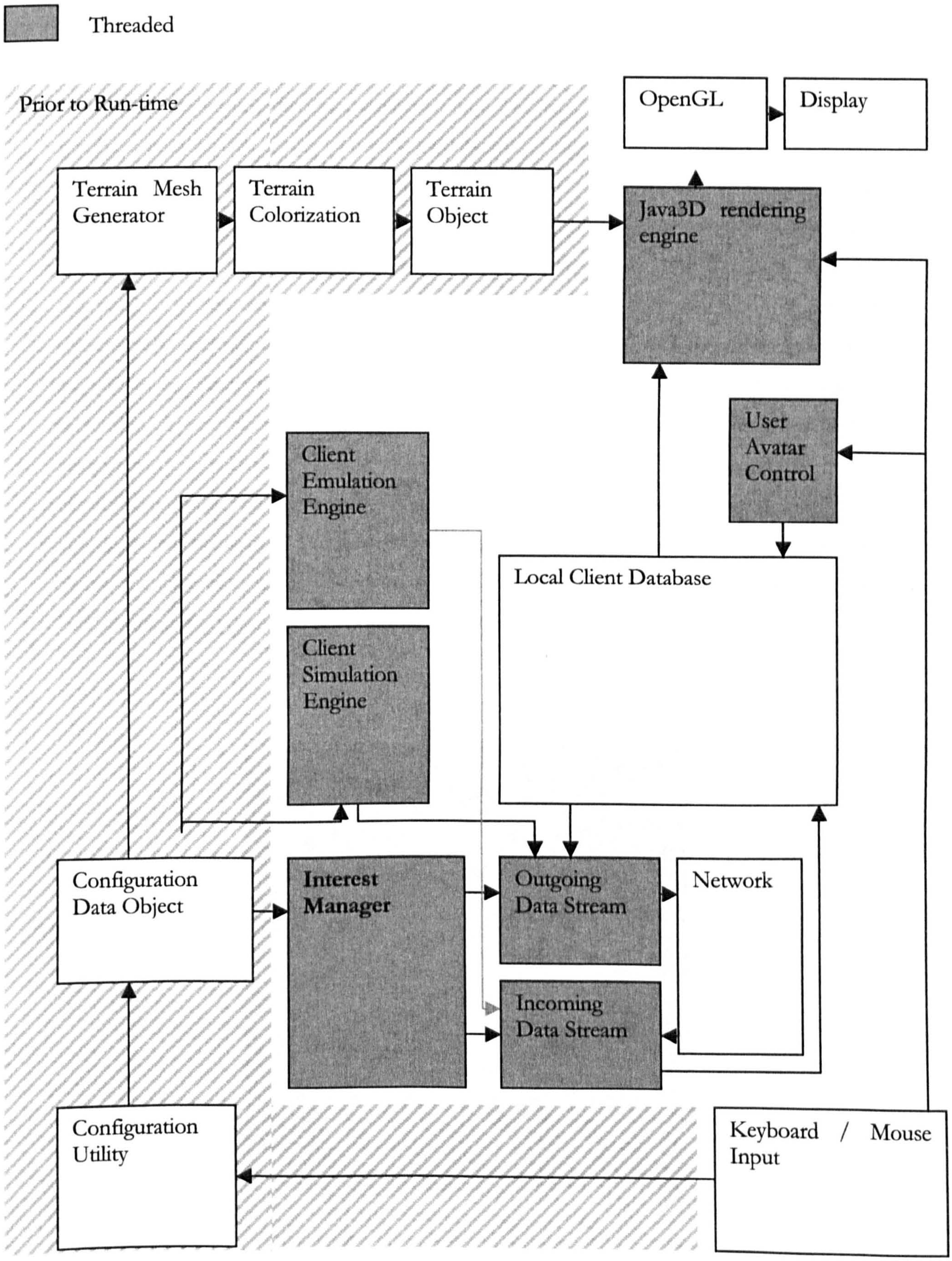


Figure 29: System Overview

5.5.1 Terrain Generation and Following

In meeting the criteria discussed earlier in this chapter, a key aspect of the system is the graphical content that serves to provide an immersive virtual environment. Central to this is the terrain generation system, which acts as a platform for the avatars within the environment. Believable terrain plays an important role in allowing users to make a clear recognition of the real-world parallel to the simulation, as identified by Kim et al.. Relating this back to the central goal of providing a platform to evaluate interest management, the terrain itself must be considered when performing experiments, since terrain occlusion and its effects on perception have a direct impact on the perceived effects of a filtering technique. A versatile terrain generation system is developed, which relies upon a source bitmap image to construct a height map. A sample image is shown in Figure 30. The image is processed with brightness values for each pixel (within an hue/saturation/brightness colour model) being used to construct a 2-dimensional array of height values. This method allows for different regions of terrain to quickly be created using a simple paint package, and interchanged within the environment to provide different conditions.

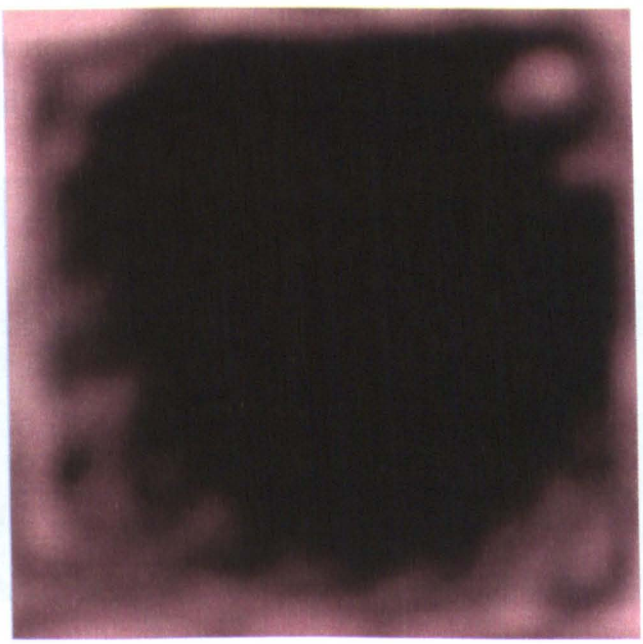


Figure 30: The bitmap image used to generate the terrain height map.

The region of terrain shown within Figure 30 is created such that a large, central open area is surrounded by raised hills. An important issue to consider is the effect of visibility on the effect of the interest management technique being evaluated. Since interest management methods already exist which incorporate visibility culling, such as the work of Hosseini [HPG02], discussed in Chapter 3, a deliberate attempt is made to avoid scenarios where visibility may affect the effectiveness of the method being evaluated. The proposal is intended to serve within situations where visibility culling alone is inadequate, and thus in order to provide adequate evaluation it is important to define the terrain such that this is indeed the case. Furthermore, visibility culling remains entirely compatible with the proposal and may be considered a component of the current state of the art, and thus by eliminating it from both interest management implementations some simplification can be made with regards to both implementation and evaluation.

On a final, more technical note, the height map produced from the image is processed into triangle strips, aligned to provide a seamless terrain mesh. This is a standard approach within terrain visualisation, and although the method lacks many refinements such as octree-subdivision to improve performance, the resultant mesh may be rendered at an acceptable framerate on the available hardware. The mesh is additionally vertex-shaded based on the height map to provide enhanced realism, and finally rendered as a solid region. The mesh and resulting visualisation are shown in Figure 31.

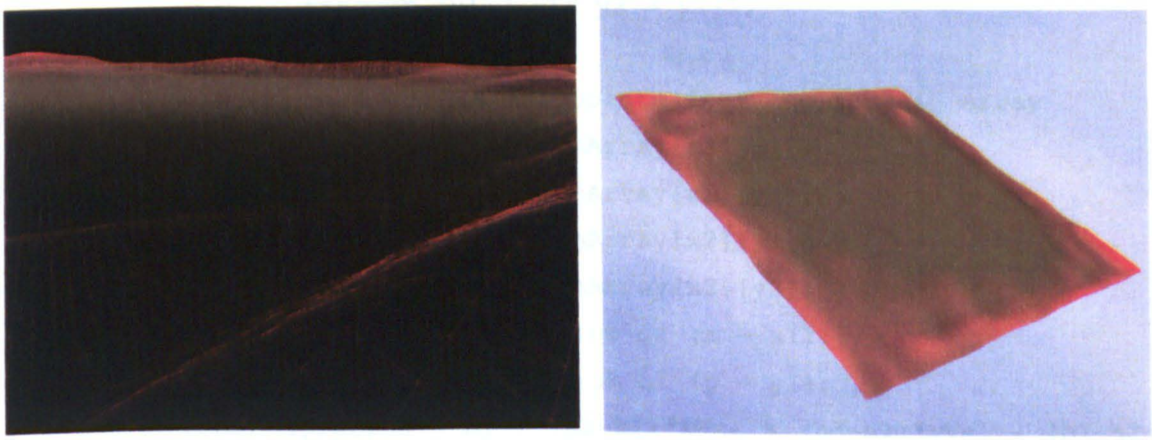


Figure 31: Rendering Terrain. Terrain mesh (left) and final rendered terrain area (right), using the source image shown in Figure 30.

The height map is again exploited in order to provide a terrain-following solution for clients within the environment. By retaining the height map in memory following terrain generation, the appropriate vertical co-ordinate position for any avatar within the environment may be simply extrapolated by considering their horizontal co-ordinates. The extrapolation method considers the proximity of a client to the four nearest points within the height map and produces a measure of the appropriate height given the slope of the mesh. It proves important to ensure this functionality is successfully implemented since unexpected client motion due to movement over rough terrain may easily be misinterpreted by an observer as a consequence of network latency and hence interest management. The method uses bilinear interpolation, as shown below (for notes on the psuedocode format used throughout this structure, refer to Appendix III):

```
method getValue(float x, float y) //gets the height value at lateral
co-ordinates x, y by interpolation from terrainArray[][]
        {
         int x1 = round to nearest integer(x);
         int y1 = round to nearest integer (y);
         int x2, y2;
            //by rounding to the nearest integer, check which point
            the given co-ordinates are nearest to
                 if (round to nearest integer (x1)!= round to nearest
                 integer (x1+0.5)) x2 = x1+1;
                 else x2 = x1-1;
                 if (round to nearest integer (y1)!= round to nearest
                 integer (y1+0.5)) y2 = y1+1;
                 else y2 = y1-1;
         //extract the 4 nearest points from the terrain array
         float h1 = (float)terrainArray[x1][y1];
         float h2 = (float)terrainArray[x1][y2];
         float h3 = (float)terrainArray[x2][y1];
         float h4 = (float)terrainArray[x2][y2];
         float xVal = absolute value of (x - x1);
         float yVal = absolute value of (y - y1);
         float solution = h1 + (h2-h1)*xVal + (h3-h1)*yVal + (h1-h2-
h3+h4) * (xVal*yVal); //bilinear interpolation
         return solution;
        }
```

When implemented, the method is capable of being executed once per frame with minimal impact on performance on the test system. Whilst potential issues with precision occur through the use of floating point values, the calculation is sufficiently accurate to provide no visible artefacts as the accuracy subtends the distance shown by a single pixel in normal use of the environment.

5.5.2 Client Simulation

Another prerequisite of the system is the ability to simulate large numbers of clients within the environment. Due to inherent complexities of human behaviour, weighed against the constraints of development time and available processing power, such simulation inevitably must prove a limited approximation of human avatars, rather than a realistic simulation. Figure 32 shows the integration of the client simulation approach within the architecture.

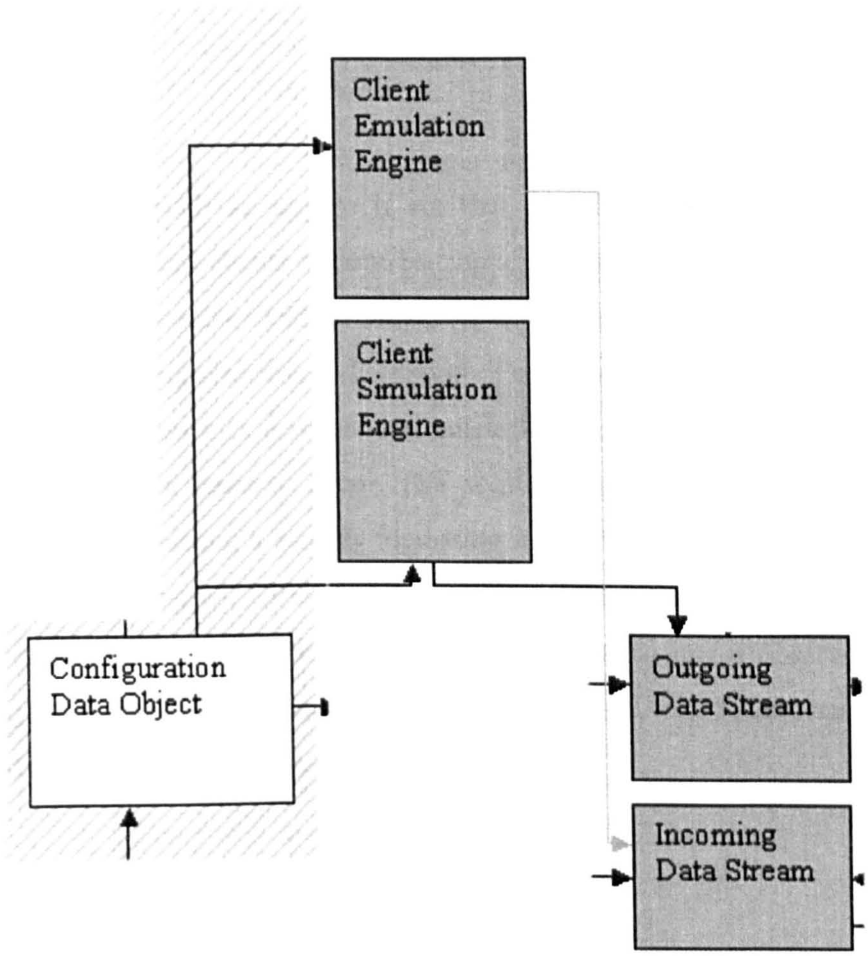


Figure 32: Client Simulation Engine and directly connected components

Two distinct types of client are created; *simulated* and *emulated*. Simulated clients are intended to mirror real-world users are closely as possible; existing as independent entities which multicast information on state to the network. Simulated clients are assigned names and network ID codes identically to real world users, yet avoid the rendering subsystem. They also incorporate interest management, applied from their viewpoint as it would be to a real world user. However, this is implemented in an upstream-only fashion, since running the interest management and associated database functionality at real-time for multiple simulated clients results in performance degradation to the extent it becomes impossible to run more than a handful of clients on a single machine. The need to provide simulation of a large number of clients, coupled with limitations in available hardware, result in this approach being adopted.

The consequences of such an approach are that global bandwidth measures become invalidated, since simulated clients are not placing a downstream demand on the network. Approximations may however still be made by considering trends in performance for individual real-world clients (for more discussion of these metrics see Chapter 7). With this compromise made, it becomes feasible to simulate 25 clients on a single machine before performance degradation starts to occur. The point at which such degradation occurred was determined by simulating a steadily increasing number of clients, as shown in Figure 33 overleaf. Whilst doing so, CPU load was measured using standard software supplied with the operating system. As would be expected, linear behaviour is observable, with 25 clients being a conservative estimate, to ensure no additional load from system processes during run-time resulting in overloading.

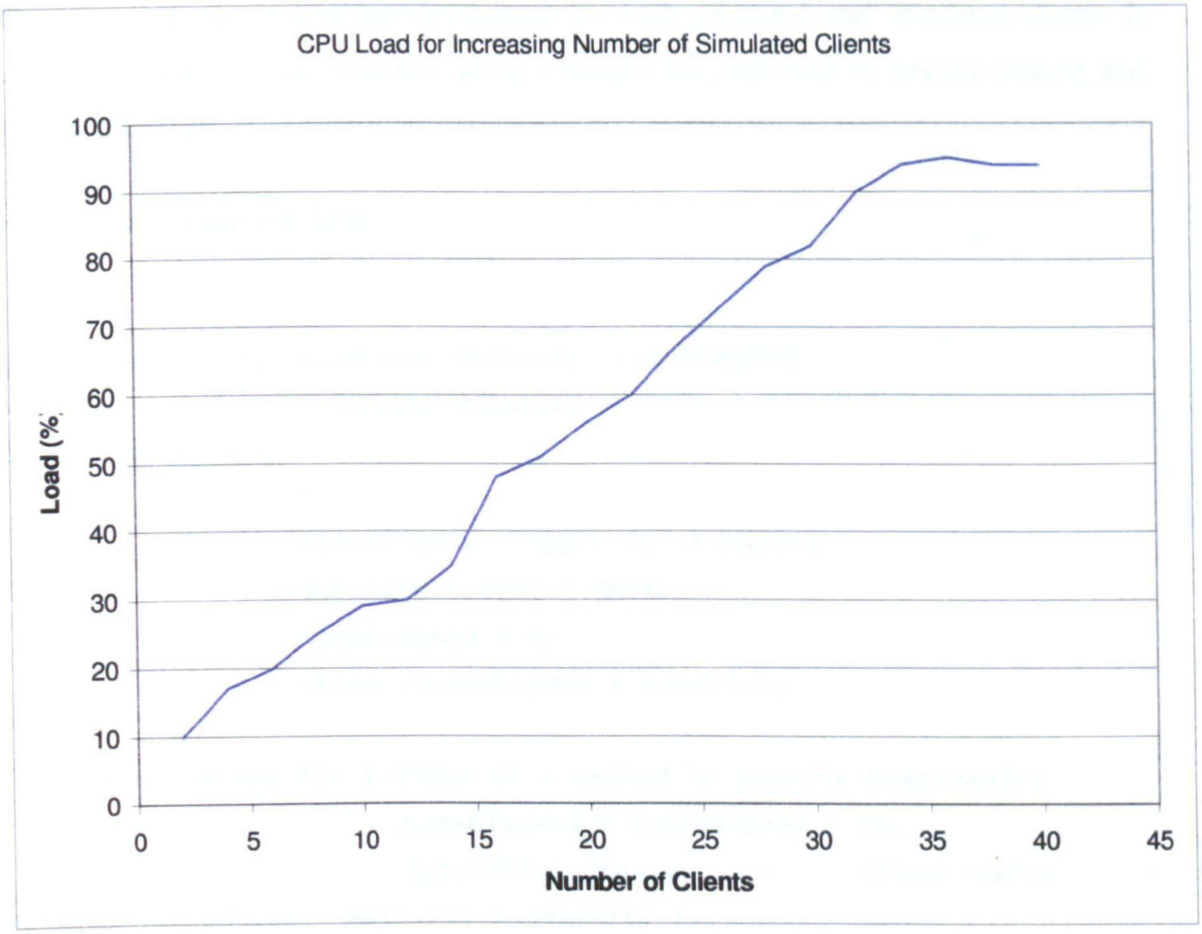


Figure 33: Effects of multiple simulated clients on CPU load of a standard PC used for the implementation, measured using Windows Task Manager.

The client *emulation* engine is implemented to enable more simple evaluation of the implementation on a single machine. Emulated clients run on a single host, and are looped-back through the incoming data stream, in effect bypassing the network yet appearing to the interest manager to be identical to sources of network information. This allows for an environment to be rapidly populated with a large number of locally simulated clients. This method may be used to produce clients that may be used to test the interest management process without the presence of a network.

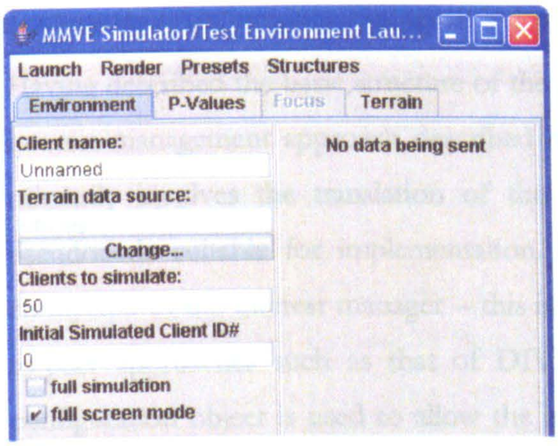
The client behaviour algorithm is identical for both emulated and simulated clients. It involves random periods of motion along a straight line, followed by random turning and stationary phases:

method moveSimulatedClient

```
            {
double direction, duration, velocity, timeElapsed
velocity = 0.8;  //constant velocity
loop indefinitely
                  {
                  direction = (2*Math.PI) * RNG();
                  duration = 8000 * RNG();
                  timeElapsed = 0;
                  while (timeElapsed < duration)
                  {
pause this thread for 1/100th of a second to prevent overloading
                        timeElapsed = timeElapsed + 10;
                        lateralCoordinate     =     (float)(xPos     +
Math.sin(direction + Math.PI) * velocity );
                        longitudinalCoordinate    =    (float)(zPos    +
Math.cos(direction + Math.PI) * velocity );
                        verticalCoordinate                           =
extrapolateFromTerrain(lateralCoordinate, longitudinalCoordinate)
                  }
                  }
            }
```

Where RNG() returns a pseudorandom number between 0 and 1.

This approach adopts the same terrain following method described in the previous section to extrapolate the vertical co-ordinate of the client from the progressively modified position. The method itself runs within a thread, which may be paused when the client remains stationary. Configuration of these clients is possible via the configuration utility shown in Figure 34.



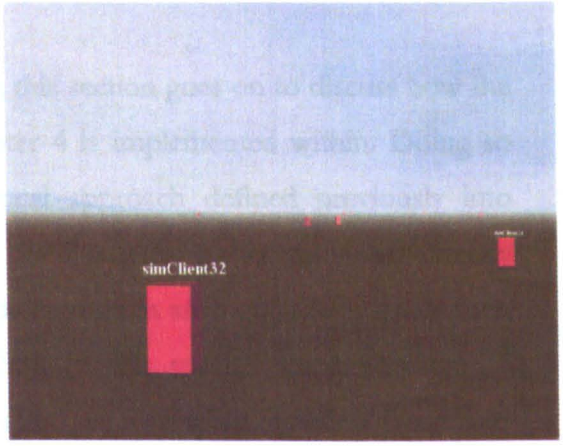


Figure 34: Simulating Clients. The client simulation configuration utility (left), and simulated clients within the virtual world (right)

The most crucial question regarding simulated clients is whether their behaviour is sophisticated and realistic enough to provide valid evaluation of a method intended for use within environments populated by thousands of real-world users. Referring back to the notions of top-down and bottom-up attention – from a bottom-up perspective they possess the same characteristics as a real-world client, and have the capability to generate attention cues through their random motion. However, they lack the interactivity to perform within tasks or activities, and hence are less satisfactory with respect to top-down behaviour. Subsequently, it is necessary to devise experiments that incorporate multiple real-world users to provide a more rigorous analysis – a single real-world user in an environment filled with such simulated clients would be unable to participate in any task-driven behaviour that involved interacting with other users.

Essentially, the simulated clients are a necessary approximation given the difficulties in obtaining hundreds of real-world subjects for experiment. They effectively simulate a large crowd, which may behave independently of the tasks of real-world users, and thus create an overcrowding scenario intended to stress the interest management technique.

5.6 Implementing the Proposed Method

Having described the basic structure of the system, this section goes on to discuss how the interest management approach described in Chapter 4 is implemented within. Doing so primarily involves the translation of the theoretical approach defined previously into pseudocode suitable for implementation. Figure 35 illustrates the components directly connected to the interest manager – this is an implementation analogous to existing peer-to-peer approaches such as that of DIVE [CH93]. In this specific implementation, a configuration object is used to allow the interest manager to receive dynamic parameters for configuration.

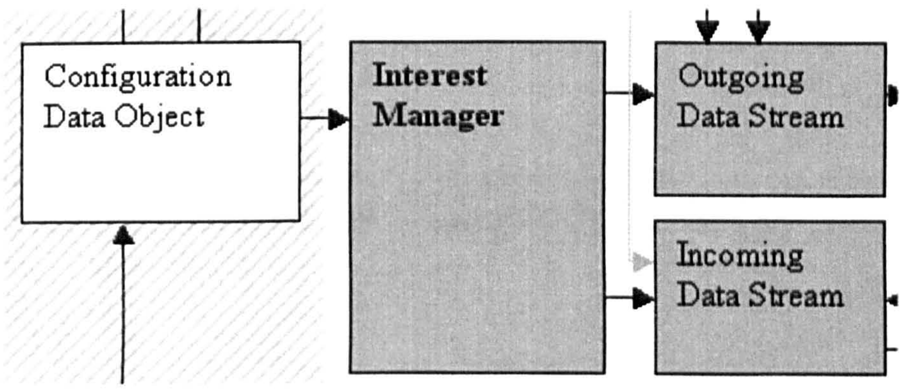


Figure 35: Interest Manager and Directly Connected Components of the System

Recalling the system architecture diagram in the previous section (Figure 29), note that the interest manager runs autonomously, in a thread separate from network and visualisation components. By acting upon the incoming and outgoing data streams at run-time, the multicast groups used for outgoing and incoming data may be changed dynamically. Thus the key challenges is implementing the proposed approach include describing the field-based approach in as computationally simple a means as possible; quantising the field results to define interest groups; and providing the logical multicast group assignments required to apply these groups to the network data streams. The first step in this sequence involves establishing a means to calculate the focal probability for an object relative to a user, and the subsequent perceptual relevance (p)-values for all objects in the world.

Thus firstly for focal probability a method is defined:

```
method calculateFocalProbability
{
define userPosition as vector
define userDirection as vector
define focusPosition as vector
define targetPosition as vector

focusPosition.y = userPosition.y; //simplify the method by working in
two-dimensions

focusPosition.x=Math.cos(2*Math.PI(double)clientOrientation.y        -
Math.PI/2));
focusPosition.z=(float)(Math.sin(2*Math.PI-(double)clientOrientation.y
- Math.PI/2));

focusPosition.x=userPosition.x + focusPosition.x;
focusPosition.z=focusPosition.z + userPosition.z;
...
```

Thus far the absolute position of the focal point of the user has been determined, based on the assumption their visual focus is at the centre of the screen described in Chapter 4. A simplification of the situation is also made in our simulation by only considering the x and z-axis, and not the vertical y-axis. This assumption remains adequate for any terrain-based situation where there is no probability of users existing with identical x and z-coordinates and differing y-coordinate values (i.e. a client stood directly above another within a building). A more rigorous, computationally-intensive solution would be relatively simple to implement for more complex environments by simply extrapolating y values based on symmetry to the x and z calculations. However, for the simulation used for evaluation, a simple flat area of terrain is used and hence the y-axis co-ordinate is irrelevant.

The next stage involves casting a line from the users position through the focal point, and thus calculating focal probabilities based on a point-line distance approach as described in Chapter 4. The point-line distance d may be expressed as:

$$d = \frac{|(\mathbf{x}_2 - \mathbf{x}_1) \times (\mathbf{x}_1 - \mathbf{x}_0)|}{|\mathbf{x}_2 - \mathbf{x}_1|}.$$

And hence this must be translated to psuedocode where x_2 is the client position, x_1 the target position and x_0 the focal point. It is also advantageous to include the a/b ratio described in Section 4.3 (as `focusAttenuationCoefficient`) to allow weighting of focal probability based on distance:

```
...

double focusAttenuationCoefficient

Vector X2 = clientPosition
Vector X1 = targetPosition
Vector X0 = focusPosition

Vector X2X1 = X2 - X1;
Vector X1X0 = X1 - X0;
Vector crossProduct = crossProduct(X2X1, X1X0);

DistanceFromLineComponent =
cross.length() / X2X1.length();
DistanceFromUserComponent =
focusAttenuationCoefficient*X2X1.length();

return (DistanceFromLineComponent + DistanceFromUserComponent)
} end method
```

Using this method a value for the focal probability of any client within the locale may be obtained. Before describing how this method is applied by the interest manager, it is also worthwhile to define a similar method used to calculate the p-value of any client within the environment based on the established focal point.

```
method calculatePValue
{

vector clientPosition
vector focusPosition
vector toCalcPosition

double x = clientPosition.x - toCalcPosition.x
double y = clientPosition.y - toCalcPosition.y
double z = clientPosition.z - toCalcPosition.z
double x1 = focusPosition.x - toCalcPosition.x
double y1 = focusPosition.y - toCalcPosition.y
double z1 = focusPosition.z - toCalcPosition.z

p          =          1/Math.sqrt((x*x)+(y*y)+(z*z))          +
clientTargetAuraRatio*(1/Math.sqrt((x1*x1)+(y1*y1)+(z1*z1)))

return p
}
```

In this method, distances x, y and z are defined as the distance between user (clientPosition) and a given point (toCalcPosition). The distances x1, y1 and z1 are included as distances between the client selected as the object of focus (focusPosition) and the given point. p may then be expressed as a combination of the length of the vectors described by (x,y,z) and (x1,y1,z1). By introducing the coefficient clientTargetAuraRatio it becomes possible to shift the p-values weighting to predominantly objects near the target in the case clientTargetAuraRatio>>1 and objects near the user in the case clientTargetAuraRatio<<1.

Between these two methods, we have a means to describe the fields proposed in Chapter 4 by calculating their intensities as return values at any given point. In the case of the method to establish focal probability, this intensity represents the relative probability of an object being the object of focus. In the case of the method to calculate p-value, the intensity represents the relative perceptual importance of the object. The next stage is to express how these methods may be applied iteratively to all clients in order to calculate the p-values for a given distribution of clients within a virtual environment. Once this is complete, the values may be assigned to multicast groups and network subscriptions performed.

In order to illustrate this iterative application of these methods, the proposal is now defined in psuedocode on the highest level, using the two methods previously discussed to calculate focal probability and subsequent p-value:

```
method applyInterestManagement

//select the client with the highest probability of focus

for all clients within locale
{
focalProbability = return of calculateFocalProbability
if focalProbability > currentMaximum
              {
currentClient = selectedClient
currentMaximum = focalProbability
}
}

//generate a perceptual field based around this client, calculating p-
values for each client in the environment

define pValues[] as an array

for all clients
{
pValue[]  =  return  of  calculatePValue(where  selectedClient  =
focusPosition)}
```

5.6.1 Field Form Coefficients

In developing the method, some value must be determined for the ratio between *a* and *b* coefficients (i.e. the value of the variable `focusAttenuationCoefficient`) as described in the Section 4.3. This ratio effectively controls the weighting with respect to focal probability, between proximity to the centre of the display, and proximity to the users viewing point. As defined in the previous sections, these coefficients observe behaviour such that when b >> a:

$$\frac{a}{b} \to 0$$

In this case, the focus is placed entirely on the object most central on the screen, regardless of its relative distance to the viewer. This is undesirable since it promotes rapid changes in focus as the user moves the viewpoint (as a consequence of fluctuating between distant objects), and also in that it is prone to identifying barely-visible distant objects whilst ignoring highly prominent nearby clients which fail to intersect with the centre of the screen. Conversely, in the case that a >> b:

$$\frac{a}{b} \to \infty$$

This situation implies the only governing factor is distance from the viewer, and thus the focal prediction is effectively based around the current state of the art, which assumes distance from user to be the predominant measure of relevance. Optimising the method implies striking a balance between a and b such to gain the advantages of both approaches.

Figure 36 illustrates the effect of varying the a/b ratio of the focal point. In the image on the left, a > b and hence the nearby client, offset from the centre of view takes prominence (note that further increasing a would result in the foremost client becoming the target of focus). In the right-hand image, a is reduced, resulting in the focal selection being the more distant, yet central client.

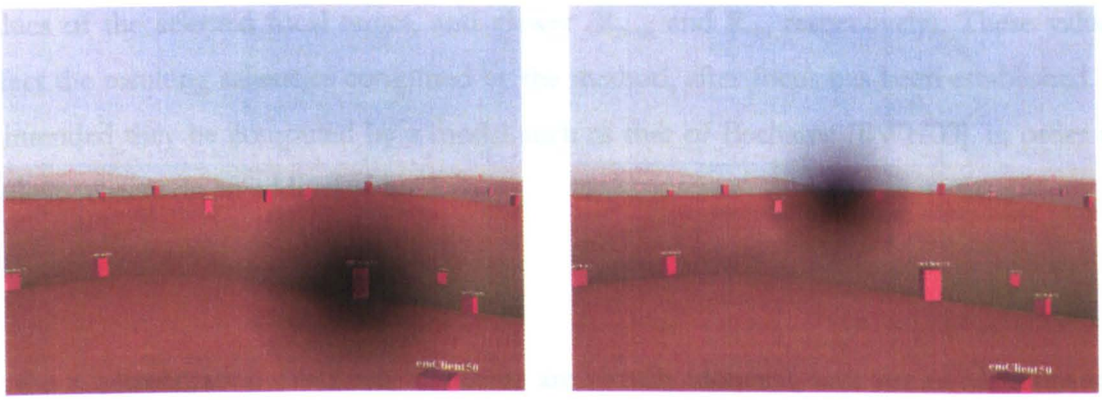


Figure 36: The effects of a/b ratio on focal point.

Determining an optimal value for the ratio, and perhaps even extending the value to be dynamic dependant on context, represents a substantial challenge, noted in Chapter 7 as future work. For the purposes of providing a functional implementation for which to perform an analysis, an approximation of the value is made based on a simple series of observations and practical experience with the method. To allow for experimentation, the capability to adjust the ratio between the two values by means of keystrokes is implemented – in effect holding one key increases the value, and the other decreases it. It then becomes possible to observe the effects whilst using the environment, and adjust the values in an attempt to achieve the best quality of simulation.

Experience within the environment with the values suggests optimum performance within a situation where b > a, and thus the majority of the emphasis is placed upon the distance from the centre of the screen. This is a favourable outcome with regards to the proposed method; the current state of the art works upon the assumption a >> b, since it filters on distance from the user alone. Whilst the sample of subjects used to evaluate this value remains too small to be conclusive, more rigorous evaluation of the proposed method described in Chapter 6 is hoped to provide further argument in favour of this approach. Drawing from these observations, a value is implemented of a/b = 0.03.

Again referring back to Section 4.3, the method is additionally weighted by the relevance values of the selected focal target, and viewer (R_{focus} and R_{obj} respectively). These values affect the resulting saliencies computed by the method, after focus has been established. It is intended they be computed by a model such as that of Beeharee [BWH03], in order to evaluate clients on an additional level.

In the implementation described, all clients are visually identical, save for an identification label above the avatar. Thus the assumption is made that R_{focus} remains constant for all clients. Whilst this may not be the case in a more sophisticated environment, it remains true for this simple implementation, since all avatars have identical characteristics. Again, the functionality is added to change the ratio dynamically, and the effects of changing weighting from client to target observed. Effectively, having a value too low for R_{obj} promotes the possibility of a very near, yet offset client being given low saliency. Additionally, clients behind the user that rapidly move into view are poorly supported since the interest manager fails to accommodate them until they enter very close proximity. Using too great a value results in an emulation of the traditional aura-based approach, wherein only nearby objects are considered and no weighting given to the focal selection. Thus once again a short experiment series is performed, with the value varied and best approximation selected. This again indicates results in favour of the proposed approach, with a value for R_{focus} being chosen as thirteen times greater than the value for R_{obj}. Again, it should be stressed this is an approximation enforced due to experimental constraints, although the general evaluation of the proposed method in Chapter 6 gives some insight into the effectiveness of this configuration.

5.6.2 Quantisation and Multicast Assignment

Ideally, the update rate of each individual client should be defined entirely as a function of salience. This would effectively allow for the most salient clients to be given a maximum rate of updates, with the frequency of updates decreasing in a continuous fashion to less-relevant clients. Unfortunately, the nature of networking implies some compromise must be made between defining update rates as a function of salience, whilst simultaneously attempting to obtain the benefits a multicast architecture allows. Defining values on a per client basis implies a new multicast group must be formed for each client-to-client communication; effectively reducing the network structure to unicast and negating the advantages of multicast. Conversely, placing all clients within a single multicast group eliminates the ability to utilise saliency, and is prone to the issues common to existing virtual environments, such as overcrowding. Thus although the proposed method generates saliency values that are continuous in nature, a multicast architecture implies that these values must be transcribed into a discrete number of multicast groups, the total of which seeks to strike a balance between perceptual gains and technical limitations. In effect, the most salient clients must be grouped into a high-resolution group, and additional tiers, containing lower saliency clients created. In practical terms, this means all clients must be set to broadcast at *n* levels of resolution, and the interest manager tasked with subscribing to the correct level of resolution for each client, given the calculated saliency.

To determine an optimal value for *n* accurately would entail a substantial series of experiments, wherein perceptual impact was measured for a wide variety of configurations. Due to constraints regarding the practicality of such a task within the scope of this thesis, an approximation is made. This is based upon research such as that by Sears and Pylyshyn [SP00], which suggests that, within a spotlight of attention, the visual system is typically only capable of tracking five simultaneous objects. To allow for some leniency, and in an attempt to avoid unintentionally inducing an attentional response by rapid changes in priority and thus update rate for "borderline" objects (i.e. objects whose saliency fluctuates rapidly), groups are configured into six layers of relevance; a first, global layer, which encapsulates all clients within the locale at a low level of resolution, and five increasingly higher tiers, with reducing numbers of clients. This configuration is summarised in Table 1.

Group Update Rate (ms)	Total Clients	Total Packets / Second
2400	190	79.17
800	2	2.50
500	2	4.00
100	2	20.00
80	2	25.00
50	1	20.00

Table 1: Multicast Configuration

With this configuration, the total packets / second for all groups is approximately 150. This corresponds to the downstream rate for a single client within the virtual environment. Given a reasonable packet size (for a practical virtual environment) of 512 bytes, this represents a level of bandwidth consumption comparable to existing Internet-based virtual environments.

With the approach detailed in concept, the issue is worthy of some practical discussion. Given the array of p-values representing the perceptual value of each client relative to the user (as determined in the previous section), multicast groups must now be assigned based on these values and the assumptions stated above. In order to perform this assignment, we need to first declare a protocol by which IP addresses will be assigned to clients. A simple approach is adopted wherein the address space 224.1.xxx.xxx is used, with the first 0-256 space used to define the client number and the second the level of resolution, which represents the update rate for the group. As an example, 224.1.4.0 would correspond to client ID#4 at the highest level of relevance, whilst 223.1.1.5 would represent client ID#1 at the 5th level of resolution.

At this point, a number of technical issues present themselves that whilst conceptually irrelevant to the proposed method place restrictions on the implementation. The reader will no doubt have noticed that by using IP address space for each client, a limit is placed on the total clients in the environment by the IP protocol itself. However, whilst this simple implementation restricts itself to 256 clients, fuller use of the available addresses by adopting the range 22x.xxx.xxx.xxx would allow for far more (655,350) clients without interfering with the majority of common network and Internet services. Whilst this is certainly adequate for any environment in the near future, the additional benefits of the likely adoption of IPv6 [LJS96] would easily allow for sufficient addresses to allow an environment with more clients than the current global population of around 7 billion. Thus it is reasonable to eliminate address space as a major cause for concern.

Additional issues primarily involve the cost of multicast (un)subscription, particularly in the situation where an attempt is made to subscribe to a group where a subscription already exists. Since the act of subscribing itself contains network overheads, any situations where multicast assignments may change extremely rapidly has the potential to induce overheads which negate the benefits of the proposed method. To prevent this, a number of mechanisms are implemented which serve to place restrictions on the rate at which attempts are made to change subscriptions.

The first such mechanism simply involves adjusting the rate at which the applyInterestManagement method is applied to the match the fastest network update rate present. In this situation whilst there is no degradation in the quality of the interest management (since the state of the client database remains constant in the interval between updates), the computational load is decreased by running the method less frequently and the network overheads associated with performing multicast assignments are reduced as the frequency of interest management applications decreases. Secondly, by directly keeping a local store of current group assignments, attempts to make multiple subscriptions to the same group can be prevented without resorting to polling the network API or operating system. This is performed by keeping a two-dimensional array of Boolean values that represent the current subscriptions in terms of client number (the first dimension) and resolution (the second). Each time the interest manager iterates, the array of required subscriptions is first cleared:

```
for(each client)
{
for(each level of resolution)
              {
toSubscribe[this client][this level of resolution] = false
              }
}
```

Then subsequently, the information from the pValue[] array is utilised to set the values in the toSubscribe[] array to true where appropriate:

```
integer totalToAssign = total number of clients

for(each client)
              {
integer currentIndex = NthHighestIndex in pValues[] array where n =
current value of totalToAssign

for (each level of resolution)
                   {
if (n<maximum permitted clients per group at this resolution) and
(pValues[this client] != -1)
                        {
toSubscribe[currentIndex][current level of resolution] = true
                        decrement totalToAssign by 1
                        break from this loop
                        }
                   }
}
```

In this code, we utilise a generic method, NthHighestIndex, which searches an array for the n^{th} highest value (where n is an integer), and returns an integer corresponding to the index in the array of this value. This is a relatively common programming task, although the solution adopted is included below for completeness:

```
method NthHighestIndex(integer n, array toSearch[])
```

```
{
array sortedArray[]
for(each array element)
                {
                sortedArray[f] = array[f]
                }

sort sortedArray by element values

                unsortedArray[] = toSearch[]

float val = sortedArray[length of array-n]
                for(m = 0 to m = length of sorted array)
                     {
                     if(unsortedArray[m]==val) return m;
                     }
                return -1 //if no match
                }
             }
```

This method in turn utilises a sort method assumed to be capable of sorting an array into an ordered series of values. The majority of programming languages (including the Java language used for implementation) provide a generic method for this task and thus we omit the code used for this task on the assumption the API provides an adequate and well-tested function. It is worth noting that in the situation of multiple identical values, the method returns the lowest possible value; e.g. a call NthHighestIndex(4, {2,4,4,4,4,5,6}) will return 2 and make no attempt to notify the calling method there are multiple occurrences of the same value. Fortunately in the context of this interest management technique the situation where two objects have equal values is near impossible, since it would require two clients occupy the same position (or positions with a line of symmetry defined by the vector between user and focal point) to a degree of accuracy such that floating point co-ordinate values in all three dimensions were equal. Even in this situation, the interest management would remain effective since the error would only occur for a single update provided one client was moving; and if all three clients were static the position updates would become irrelevant.

Returning to the discussion of the previous section of code employing the `NthHighestIndex` method, the approach can be seen to be iterating through two nested loops, in order to analyse each client at each level of resolution. Whilst the method iterates through the array of p-values in the order they appear in the array, before checking through the levels of resolution a call is made to `NthHighestIndex` to establish the current index to assign. Thus assignments are made in order of priority, with the most significant client being assigned first. For each level of resolution a reference is made to the maximum permitted clients in this level and if this is this is less than the current number of clients left to assign an allocation made. The end result is a Boolean array `toSubscribe[]` stating the recommended multicast assignments. Having built the two Boolean arrays, multicast assignments may now be simply performed by comparison, and subscription / unsubscription calls made where there are discrepancies between the two arrays. The state of the `alreadySubscribed[]` array is preserved between iterations of the interest manager, preventing multiple subscription calls being made where unnecessary:

```
for(each client)
{
for (each level of resolution)
                {
if (alreadySubscribed[this client][this level of resolution] is false)
and toSubscribe[this client][this level of resolution] is true)
                    {
subscribe to 224.1.[this client].[this level of resolution]
alreadySubscribed[this client][this level of resolution] = true
                    }
if (alreadySubscribed[this client][this level of resolution] is true)
and toSubscribe[this client][this level of resolution] is false)
                    {
unsubscribe from 224.1.[this client].[this level of resolution]
alreadySubscribed[this client][this level of resolution] = false;
                    }
                }
}
```

With the interest management infrastructure in place, the information sent via the network is worthy of some consideration. The packet structure (Figure 37) remains constant regardless of interest management approach, representing the data needed to update the position of an avatar in the virtual world:

TCP/IP Header	Avatar ID	Avatar Name	Position	Orientation	Flags
~60bytes	32 bits (integer)	~15 bytes (byte/char)	96 bits (3 floats)	96 bits (3 floats	16 bits (short int)

Figure 37: Packet Structure.

The bulk of the packet consists of the IP header, which contains information required to correctly transmit the information across the network (including source and destination information, and sequencing data). This is typically in the region of 60 bytes. A unique ID number used to identify each avatar is included, alongside a string representing the avatar's displayed name within the environment. Two three-component vectors representing position and orientation form the vital information used for positional updates. A final field is used for flags that may be manipulated to mark clients or induce global events such as a change of interest management technique.

In the context of commercial distributed virtual environments, optimising packet size is crucial in providing as efficient a system as possible. Furthermore, sophisticated environments typically have a greater packet size, as avatars and their interactions require additional information to be transferred. However, since the proposed method works independently of packet size, emphasis is given on providing a simple structure that transfers all information relevant to the simulation. Consequently, the header information required by the IP protocol is significantly larger than the data within the packet. Larger packet sizes would have a direct impact on the bandwidth required, with the update interval between packets increased to compensate. In this respect the proposed method may be considered to accommodate varying packet sizes since update rates of multicast groups as defined in the previous chapter may be changed to compensate. Additionally, UDP offers clear advantages over TCP/IP in decreasing packet size, since additional IP header information such as sequencing is irrelevant in this context. TCP/IP has advantages in simpler implementation, however, and since performance within the system is measured in relative (comparison between techniques) rather than absolute terms, performance optimisation is unnecessary.

5.6.3 Hysteresis

An additional phenomenon presents itself when implementing a spotlight-based approach, which mirrors an issue discovered by Funkhouser and Séquin [FS93] when attempting to apply visual attention to the rendering process. The issue centres upon objects with near-identical saliency values, arising due to extremely close proximity, but may also be induced by rapid re-orientations of the field of view. Effectively, objects rapidly oscillate in priority within the system. In the context of the work by Funkhouser and Séquin, this is seen to result in rapid changes in level of detail for threshold objects, which evoke an unintended response from the viewer. In the case of this implementation, however, the change is much harder to perceive since the level of detail remains constant, with only the frequency of updates regarding position changing. Such rapid shifts in priority are, however, extremely expensive with regards to multicast subscriptions.

In order to accommodate this issue, a hysteresis-based approach similar to that used by Funkhouser and Séquin is adopted. Rather than simply limit the multicast subscriptions permitted per second arbitrarily, an approach is taken which compares the relative saliencies of objects and only induces a shift between multicast groups if this saliency exceeds a certain threshold. The value of this threshold is minimised in an attempt to ascertain a balance between an optimal implementation of the spotlight concept, and the overheads implied by the multicast architecture. Again the capacity to modify the value at run-time is implemented; with experience within the environment weighted against multicast subscription rate (the method used to measure subscription rate and results are later described in Section 6.2.2). In the absence of the facilities to perform more rigorous evaluation regarding the optimum value, an approximation is again made which is subsumed into the evaluation process described in Chapter 6 when comparing the configured proposal against an approximation of the current state of the art. Since the oscillation only becomes a significant issue regarding performance when the multicast subscription calls increase dramatically to several hundred per second (in the rare case that two clients coexist in the same virtual space), a low threshold value of 0.5% is applied, with no apparent impact on quality.

Having described the practical implementation of the proposed technique, it becomes possible to consider configuring the method within the environment. The next sections detail the approach used to allow configuration of interest management both before and during run-time, and subsequent analysis of the method, in order to confirm it to be working as intended before evaluation.

5.6.4 Interest Management Configuration

A central goal of the system is providing a dynamic, configurable approach to interest management. In addition to the proposed method outlined previously, the facility to configure other interest management approach for evaluation purposes is provided, leading to the experiments described in the next chapter. It is also necessary to provide the facility to fine-tune the proposed approach by adjusting the variables defined in the previous chapter. In order to provide these facilities, a configuration utility is developed and integrated as shown in Figure 38, to allow interest management presets to be created that may then be switched between during run-time. Run-time switching presents an issue with the peer-to-peer nature of the system - consistency must exist in the interest management approach across all clients, or attempts to address non-existent multicast groups result. Thus run-time switching of techniques must result in the propagation of a request via the network to perform a simultaneous change between methods. This is implemented by adding the facility for a client to change between presets by setting a network flag.

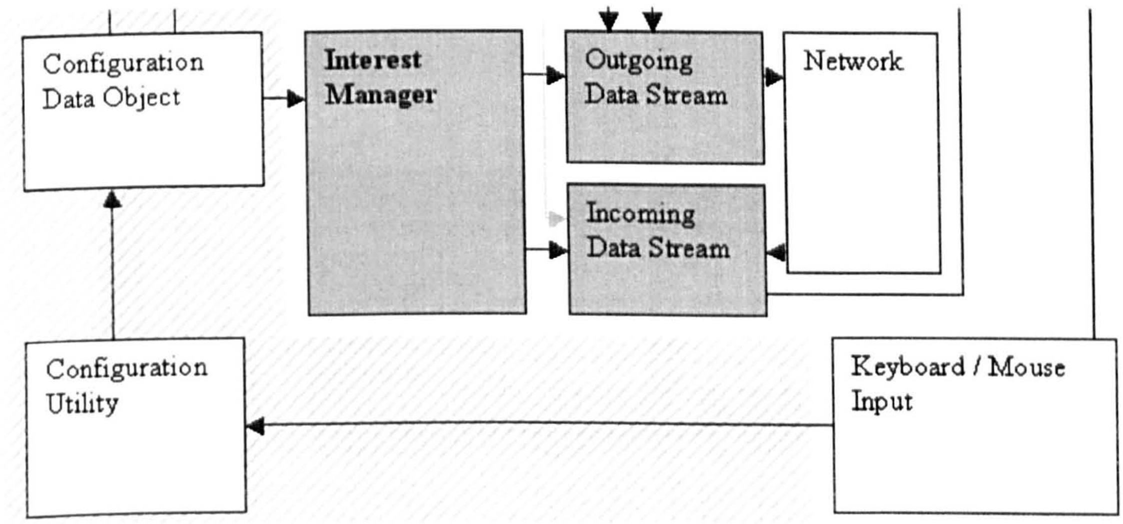


Figure 38: Configuration Utility and directly connected components

Presets are specified via a configuration utility, shown in Figure 39. The mode pane (top left) allows the generic type of interest management to be selected. Global update rate (top right) allows the specification is milliseconds of the delay within which all clients are mutually updated. This is typically in the region of several seconds, and allows aura-based interest management to be performed. Aura configuration (bottom left) allows for different spotlight attenuation models to be examined. Groups per client (bottom right) enables the creation of higher update-frequency multicast groups which are individually configurable (inset).

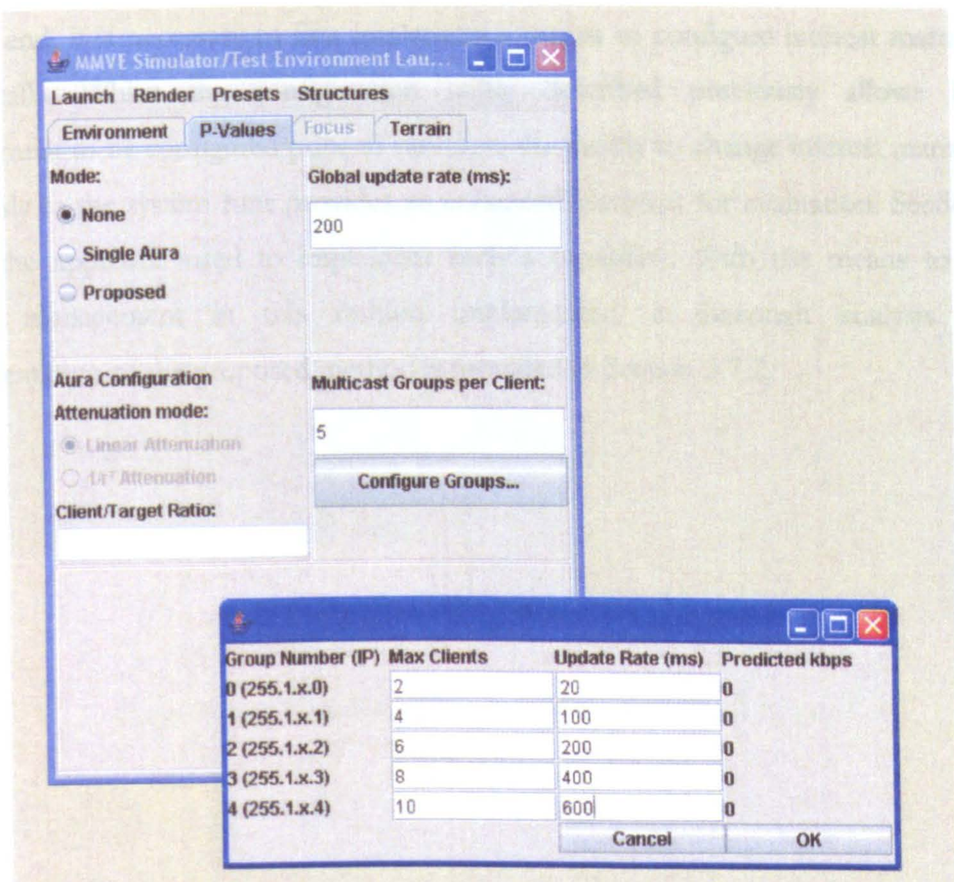


Figure 39: Screenshot of the IM Configuration Tool.

5.7 Testing the Implementation

With the model proposed in Chapter 4 implemented, the final issue before attempting to evaluate its success is to verify the method works as described conceptually in Chapter 4. Such verification has already been done partly within the previous section (the demonstrated impact of adjusting a and b values, for example, gives some evidence the focal approach is correctly implemented). However, it is important to provide a more rigorous assessment before moving on to describe perceptual experiments and perform measures of bandwidth consumption in Chapter 6.

To this end, it is necessary to first implement a means to configure interest management dynamically. Whilst the configuration utility described previously allows for the environment to be configured prior to run-time, the facility to change interest management seamlessly as the system runs provides an enhanced platform for evaluation. Section 5.7.1 details the approach used to implement such a capability. With the means to modify interest management in this fashion implemented, a thorough analysis of the implementation of the proposed method is provided in Section 5.7.2.

5.7.1 Dynamic Interest Management

Figure 40 illustrates the user interface used to configure interest management. Run on a single client, this allows the interest management system used by all clients to be changed dynamically. This is achieved by using the offline configuration utility described in Section 5.6.4 to create a number of stored presets. Each preset contains information on the multicast group arrangement, global update rate, interest management approach used, and parameters used to configure the interest management (such as the a and b coefficients). Presets are stored locally on each client within the environment.

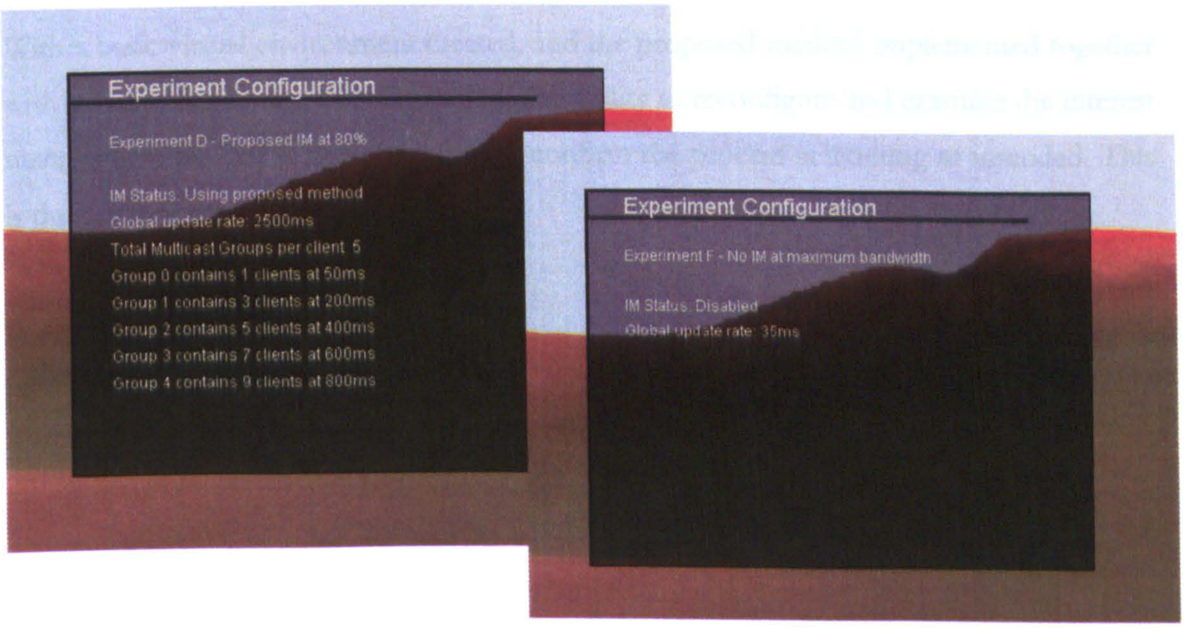


Figure 40: Two illustrations of the interest management configuration panel.

When the user changes the interest management paradigm by use of the interface shown in Figure 40, a flag is set in outgoing packets from that peer to change method, and propagates to all clients via the network. Subsequently, multicast groups are reformed and the environment re-synchronised. The entire process takes in the region of several hundred milliseconds, and appears seamless to users. Only one client need be used to trigger the change; acting as a control client for experiments.

Allowing such a capability is advantageous when performing perceptual experiments, as outlined in Chapter 6. Integrating the dynamic capability within the interest management process itself, so the interest management approach could adapt in response to network load or other factors would appear an interesting avenue for further research (noted in Chapter 7), although much of the success of the dynamic implementation lies in the fact the large-scale environment is merely simulated on a high-performance local area network. Attempting to reconfigure large numbers of multicast groups in such a fashion in a truly distributed, Internet-based scenario would result in a much larger delay due to hardware constraints. Hence the dynamic capability is used purely as a tool for evaluation, and each defined interest management approach is held static during evaluation.

With a basic virtual environment created, and the proposed method implemented together with a means for dynamic configuration, this ability to reconfigure and examine the interest management process is now exploited to confirm the process is working as intended. This is the focus of the next section.

5.7.2 Visualisation of Results

Inspecting the approach presents something of a paradox; the proposal intends to exploit the limitations of human perception and utilise change blindness to filter less relevant clients, although in order to visually confirm the system is functioning the changes must be in some way perceived. Using the dynamic method described previously allows for some capacity to change systems rapidly in order to note any changes, although this still makes confirmation of a correct implementation difficult. To solve this problem, the capability is added to shade entities based on their saliency (p-values) determined by the proposed approach. This has the effect of illustrating how relevance is being assigned. Figure 41 shows a screenshot of the environment with shading in effect. Clients are shaded on a scale (show in the top left of Figure 41) with red representing zero relevance and blue maximum, shades in between representing interim values.

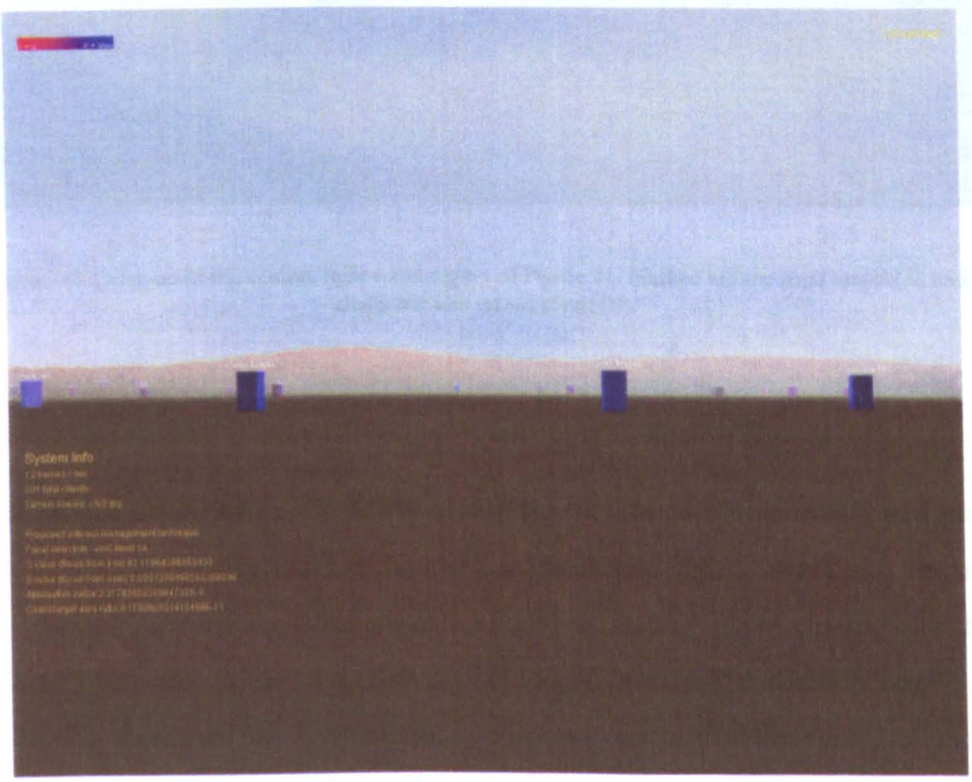


Figure 41: Relevance Shading

In Figure 41, it can be seen that the most prominent clients are shaded dark blue, enforcing the notion that the visible clients are correctly calculated as the most salient. A closer inspection of the central right hand region of the image in Figure 41, shown in Figure 42, demonstrates the background clients towards the side of the image to be given lower saliencies, since they are away from the focus. However, high priority is given to the focal target, marked in Figure 42 as client A, visible on the far left (and central to the overall display). Nearer, yet offset objects on the right-hand side of the image such as client B can be viewed to have lower relevance.

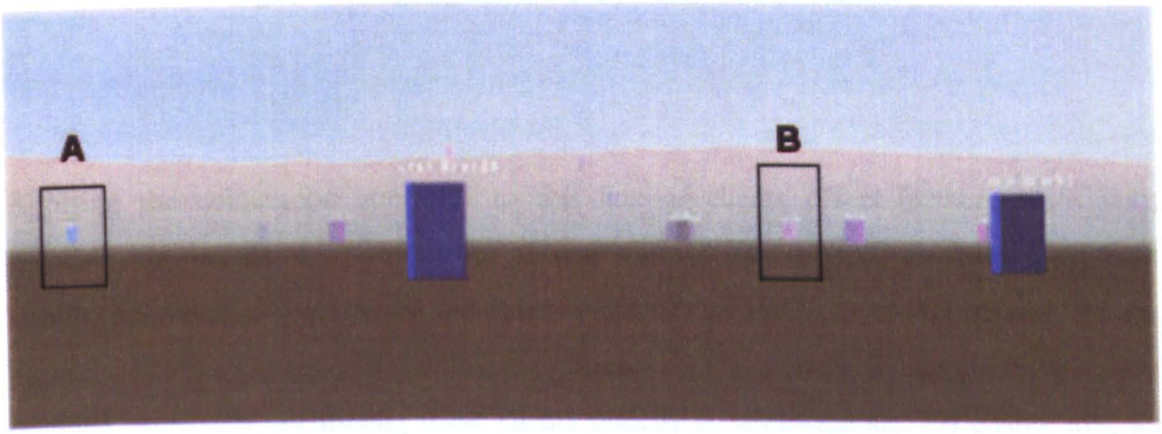


Figure 42: Close-up of the central, right-hand region of Figure 41. Marked are the focal target (A) and a closer but less salient client (B).

To offer further clarification, the ability to use the simulated clients mentioned previously to stress the interest management technique may be considered. A worst-case scenario (and the best example) for the focal technique is a long, oscillating line of clients, with the client stood roughly parallel to the line. This is due to the fact such a scenario rapidly induces change in the object of focus, as clients oscillate in and out of the centre of the screen. Figure 43 overleaf illustrates this configuration via a group of 100 simulated clients. By observing the behaviour of the proposed method in this situation it is easier to identify the object of focus.

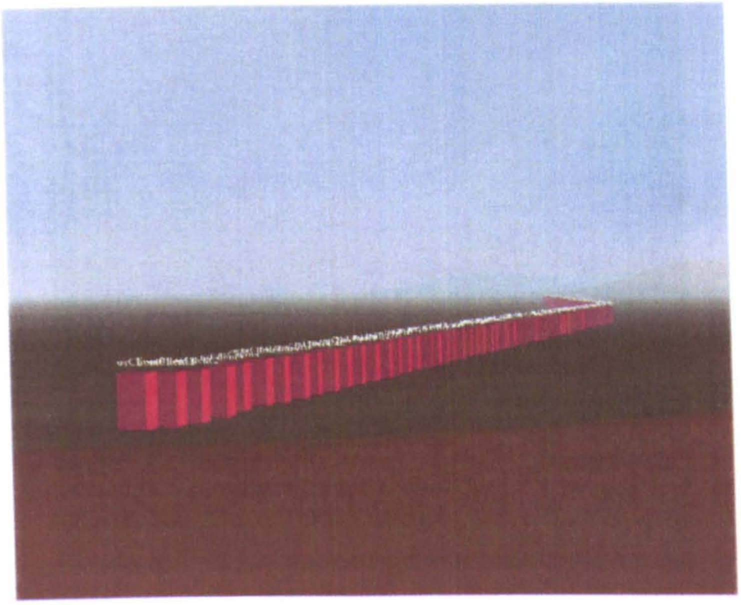


Figure 43: Configuring simulated clients to stress the implementation.

Applying the colouration approach to this line of clients offers further clarification; however, due to the large volume of clients, some difficulty exists in using a gradient shading approach to demonstrate the order of saliency values. Since the clients with the ten highest saliencies are devoted the majority of resources, as per the multicast configuration outlined previously, identifying these clients directly is advantageous. To enable this, an adapted colouration technique may be utilised which colours objects more distinctly. In Figure 44, such an approach is shown for a client viewing the line of simulated clients. The five most salient clients are coloured red, and subsequently lower saliencies coloured orange, yellow, green and blue in that order.

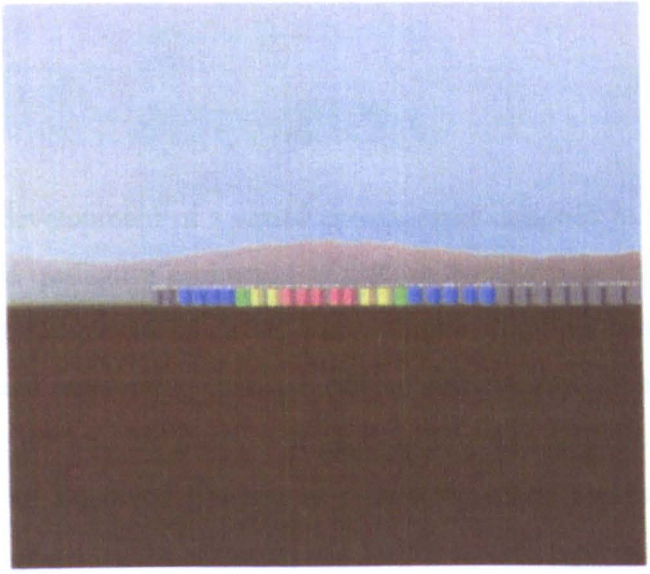


Figure 44: Multicast Shading and Point of Focus

With this method, it becomes much more readily apparent that the focus is centred on the point at which the line intersects with the centre of the field of view. By comparison, the traditional aura-based approach would tend to favour the clients on the extreme right of the image, since they are closer to the viewer. An interesting test of the proposed method is to reallocate focus to the end of the line. In this situation, saliencies should be assigned in decreasing value along the line. Figure 45 shows that this is indeed the case (note that this image is cropped to illustrate the saliency assignment; the centre of the screen is, in fact, on the left of the image).

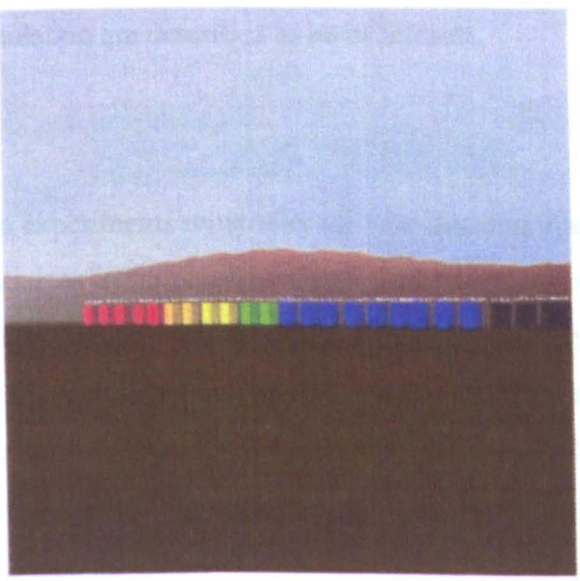


Figure 45: Confirming Multicast Assignment

5.8 Summary

In this chapter, the development of a virtual environment designed to provide a means for both quantitative and qualitative evaluation of interest management has been documented. Due to the inherent complexity of virtual environments, drawing upon diverse computer graphics, database, and networking concepts, several assumptions are necessary, and where possible this chapter has justified them to as great an extent as possible. Fundamentally, the environment provides a general (though unoptimised) approximation of many existing environments, providing a region of terrain and simple models representing human avatars.

Having created such a foundation, the method proposed in Chapter 4 has been implemented and configured. Whilst many aspects of environment design and content may have an effect on the experience perceived by a user, by implementing an approximation of the current state of the art alongside the proposed method, comparative evaluation may be performed within the system, with the only variable being the interest management paradigm used. Such an approach is necessary since the perceptual nature of the system means analysis of bandwidth characteristics alone is insufficient. Additionally, a goal of the proposed method is to allow lossy reduction in quality of simulation in situations where overcrowding occurs, and thus experiments in reducing bandwidth and noting effect on perceived quality of simulation are described to be of interest.

With emphasis given to experiments most relevant to evaluating interest management, and considering limited resources in terms of hardware and tester availability, experiments are now defined which aim to provide qualitative insight into the success of various techniques. The next chapter also presents the results of these experiments, alongside measures of bandwidth consumption for various methods, to provide a rigorous analysis of the proposed method.

6 Experimental Design and Results

This chapter details the development of metrics intended to evaluate the success of the proposed method, via the implementation discussed in the previous chapter. In order to do so, it is necessary to consider the success of the method in obscuring the effects of limited bandwidth to users, in comparison to the current state of the art. In the particular context of the method proposed by this thesis, an additional consideration is imposed by the perceptual nature of the technique – evaluation must include some consideration of the human end-user and their response to the method. Doing so involves the creation of experiments that bring together a significant number of real-world users collaborating within a virtual environment. To this end, a series of perceptual experiments are devised, which involve a group of users interacting within a sequence of tasks. By illustrating to users differing update rates for clients within a virtual world, and thus providing them with a reference, they are asked to estimate the quality of simulation by means of approximating the delay between updates. Section 6.1 explains the rationale behind these experiments and their implementation, with particular regard to the approach used to emulate and compare the current state of the art to the proposed approach.

It is also necessary to provide an analysis of bandwidth consumption and multicast subscription rates, in order to fully appreciate the network load within the various experiments. In Section 6.2 the nature of these measures is explained and their results for each experimental configuration presented. Two instances of the experiment session are performed, in part to ascertain the impact learning and experience has on the effectiveness of the system; the first instance is with a small sample of users experienced with the system, and the second with a larger, random sample of subjects. Results for these two sessions of perceptually-oriented experiments with 5 and 10 subjects are shown in this chapter, and analysed in an attempt to ascertain a direct comparison between the bandwidth consumed by the system and resulting quality of simulation. These results, presented in Section 6.3, suggest the ability to reduce the measured bandwidth substantially using the proposed method, with little impact on the perceived simulation.

6.1 Experiment Design

Whereas the previous chapter concerned itself with the implementation of the proposed method, and centred around experiments intended to illustrate the proposal performing as expected under a set of controlled variables, this section moves on to answer the crucial question of whether the method provides tangible benefits over existing interest management techniques. In order to achieve this, real-world users are introduced to the environment alongside simulated clients, and used to evaluate three simulations: the first with no interest management, the second an approximation of the state-of-the art, and finally the proposed method. This introduces several complex questions, most notably how to provide an implementation of the current state-of-the art when interest management is typically application-integrated with few, if any, standards existing. This section seeks to address these issues, and also present the rationale behind the experiments detailed in this chapter.

6.1.1 Evaluating Perceived Quality

Chapter 3 provides evidence that the majority of interest management approaches are evaluated with respect to their reduction in bandwidth consumed. Hence the simplest measure of success for an interest management technique is to prove a reduction in bandwidth whilst quality of simulation remains unchanged. However, the proposed technique has at its core the notions of scalability that are often described as key to internet-based environments. Subsequently, we aim to both provide this simple measure of success as a means of comparison to existing techniques, but also to look more closely at the nature of the relationship to quality of simulation and how minor reductions in this respect may lead to worthwhile decreases in bandwidth consumption, or its reallocation to actually enhance the quality of simulation over locale or aura-based methods (by providing greater update rates on the most significant clients).

Thus having loosely established the framework within which we are evaluating interest management, we may now go on to discuss how existing interest management approaches have tended to be evaluated with respect to their impact on bandwidth, with regards to developing a method for evaluation for the proposed system

As noted in Chapter 3, quality of simulation itself is a much harder variable to measure due to both its relationship to human perception and a relatively smaller body of existing work. We effectively need to answer two questions when comparing interest management techniques in terms of their impact on quality of simulation; firstly is the difference noticeable to users, and secondly, if so, by what margin does one technique provide better quality of simulation over another? Furthermore, we must also consider the behaviour of a single technique as available bandwidth is varied and clients move throughout the environment gathering and dispersing, effectively asking these same two questions.

The first question, of simply noticing a difference between two approaches, is by far the easiest to answer. Drawing from basic experimental psychology, and relevant experiments into effects such as change blindness [RRC97], it immediately becomes apparent that utilising as large a group of testers as possible is desirable. Furthermore, from other experiments into the nature of top-down attention [CEY04], it becomes apparent that a user's task within an environment is liable to affect their visual attention greatly. Since creating virtual environments with practical application implies some type of inherent task-based nature, to simply ask users to randomly move throughout an environment and identify discrepancies alone would appear insufficient. A factor also noted frequently by researchers performing any type of test within a virtual environment is the close relationship between familiarity with the interface (both in terms of input and output), and the effectiveness with which users both perform tasks and detect changes within the environment – this may be related back to the notions of immersion and presence defined in Chapter 3 [KDE00].

Whilst theoretical predictions and practical measures of bandwidth consumption can offer some means for evaluation, the perceptual nature of the proposed technique warrants a qualitative aspect to evaluation involving real-world users. To devise experiments aimed at such evaluation, advantage is taken of the ability to dynamically change interest management at run-time, as discussed previously. The need to provide a rigorous and thorough evaluation of the proposal in a wide variety of conditions is balanced by the constraints inherent to gathering a large group of testers for a prolonged series of experiments. Additionally, since the evaluation must be comparative (taking the proposed technique alongside other interest management approaches), experiments must be repeated multiple times. This presents a challenge in creating sessions that provide meaningful evaluation within a limited timeframe.

Th ough assigning users interactive tasks within an environment is expected to affect their perception and hence subsequent response in evaluating interest management, the actual effectiveness with which they perform the task is often used as a metric for the success of a perceptual technique. This has been used as a measure for the effectiveness of rendering-based approaches described in Chapter 3 such as that of Sundstedt et al. [SDL05]. It has drawbacks, however, in the context of interest management since interaction and devising meaningful tasks proves complex, particularly when additional variables arising from the unpredictable nature of human interaction within the environment are taken into account (when first presented with the environment described in Chapter 5, many user's focus automatically on finding one another and interacting, regardless of task). It is thus advantageous to define additional means of measuring the effectiveness of a given approach rather than task completion alone.

Drawing from the evaluation methods used for existing techniques ([BWH03], [MLS05], [GB95], [OG02]), we can determine some general themes of test:

- Change detection between techniques. Essentially, can a change be noticed between interest management technique A and interest management technique B within the same environment?

- Change detection between bandwidth. Given the same interest management technique, but different configurations to consume less bandwidth, is a change noticeable?

- Quality. In the event change can be detected, which method 'feels' more natural to the user and provides the better-looking simulation?

Within these three themes, we can also define two types of experiment which apply to both:

- Non-objective based experiments. This is simply allowing a user to roam within a virtual environment, under different interest management approaches and bandwidth conditions.

- Objective based experiments. This involves giving a user a task to do, to alter their perception and hence observe what effect this has on their response to changes in interest management technique and bandwidth. These objectives may be further broken down into two categories:

    - Environment-centric. An example would include finding a static geometry object within the scene. Essentially, the interest management approach has no effect on this test except for providing a background.

    - Entity-centric. This would typically entail interaction between two clients within the environment. Hence interest management plays a role in this interaction and subsequently should have a far greater effect on the test than in the environment-based case.

Given limited resources with which to test, it is logical to place an emphasis on objective based, entity-centric experiments, since they provide the most direct evaluation of the success or failure of interest management, particularly in a visual-attention based context. Two such entity-centric experiments are devised. These centre on giving users clear, objective-based tasks that involve interaction with other users within the environment. They are chosen particularly to stress and evaluate different aspects of the interest management process, and comprise:

- A *"grouping activity"*. The most common cause of breakdown in interest management techniques is the gathering of a large number of users in a small region, and the consequent increase in load. In this task, users are randomly distributed throughout the environment and instructed to find other real-world users through a crowd of simulated clients. The users identify one another via labels attached to the top of their simple avatar, as shown in Figure 46. These identifiers become legible at approximately the same distance facial or other recognition would be possible with more sophisticated avatars. This has the goal of providing a transition from low to high client density within a small area, and subsequently stressing the interest management.

Figure 46: A single client and identifier

- A *"tag game"*. This mirrors the simple real-world game of tag, wherein one user is randomly selected as being tagged, and must search out and tag other users, again amongst a crowd of non-participating simulated clients used to induce overcrowding. The game is implemented in a 'last man standing' fashion wherein tagged players remain tagged until only a single player remains, at which point the game ends. To enable tagged users to catch others within a reasonable timeframe, it is necessary to increase their movement speed slightly (a value of 20% is used). In order to allow users to identify, and thus try to avoid, tagged users, the avatars of tagged users are shaded blue. This promotes a range of behaviours in participants, including evasion, searching, and distribution throughout the environment as users seek to evade or tag other users.

These two activities result in a variety of phenomena intended to stress test interest management. In the case of the grouping activity, the environment starts with evenly distributed users, and then gradually they form into a single location, creating the overcrowding problem noted in Chapter 3. Users are also constantly performing a visual search task in order to find other users, with consequences for their deployment of visual attention. The tag game again seeks to explore different distributions of users, as opposed to the grouping activity users are initially inclined to scatter, and gradually redistribute throughout the environment. In avoiding tagged users they are also typically attempting to identify and elude more distant clients in a way similar to online games and military simulations such as NPSNET. Conversely, tagged users are attempting to avoid the crowd of simulated clients in order to track down other participants.

Within these two activities, it is necessary to provide a comparison between the proposed method, and current state of the art. One of the most difficult questions posed is how best to represent the current state of the art, in order to allow such comparisons. Rather than attempt to emulate a comprehensive approach such as those detailed in Chapter 3, it is instead preferable to focus on the key characteristics that differ from the proposed method, to allow a more fundamental comparison. Returning to Chapter 4, the method is described as a field defined by radial auras around both client and focus. Recalling equation 1 from Chapter 4:

$$P = \frac{2R_{obj}}{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} + \frac{2R_{focus}R_{obj}}{(x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2} \quad \textbf{(1)}$$

Consider the case of the proposed method configured such that $R_{focus} = 0$. In this case the right-hand term is eliminated and thus:

$$P = \frac{2R_{obj}}{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad \textbf{(5)}$$

This represents a single, radial aura centred upon the viewer. Whilst existing methods apply additional functionality, such as a high-level grid partitioning, visibility culling, or consideration of objects on a top-down/bottom-up level, these methods all imply either this behaviour, or a more simplistic equal-relevance measure ignoring distance from user (i.e. P = 1). No method applies an additional aura around the target of focus, which is the reserve of the proposed approach. Thus, it becomes possible to consider the state of the art as this special case of the proposed method, and provide direct comparisons. The P = 1 case may be eliminated from consideration since distance based filtering has been repeatedly shown to offer improved performance in both interest management and other scenarios [HPG02], [FS93] and hence the current state of the art is approximated using equation 5.

Whilst research such as that of Beeharee also serves to provide attention-based enhancement to interest management, and may in this respect be viewed as the state of the art regarding attention-based interest management, it centres on top-down and bottom-up models that adjust the relevance of objects - R_{obj} in the case of this model. By keeping this value constant within the experiments performed on the method, it is possible to focus only on the aspect of this current state of the art incompatible with the proposed technique – i.e. the relationship between distance from viewer, focal point, and resulting salience. Thus the experiments are intended to offer a direct comparison in these respects alone, with a view towards creating a method compatible and complimentary to the current state of the art to as great an extent as possible. It is unnecessary to provide evaluation of those aspects that the proposed approach does not seek to modify (and attempting to do so would impose other variables which may compound the results), and hence a complete implementation of all aspects of the current state of the art is deemed unnecessary.

With two experiment activities defined and a basis for comparison established, the next section discusses the precise configurations of interest management utilised throughout the tag and grouping activities, in order to compare differing approaches and configurations. Arising from this is a set of twelve experiments (the two activities described through six interest management configurations), forming the basis for evaluation.

6.1.2 Experiment Configuration

Arising from the discussion in the previous section, evaluation requires that entity-centric tests be devised, seeking to test and quantify the differences between approaches. Using the dynamic interest management capabilities of the system described in Chapter 5, it is possible to define interest management configurations prior to an experiment session, and change them dynamically. This includes versions of the proposed method, and a current-state of the art approximation created via the configuration described in the previous section. In order to validate the experiment it is also necessary to prove users capability to estimate update rates to within such degree of accuracy that obtained results may be justified following analysis of experimental error. To do this, additional iterations of each experiment are included using no interest management, with both high update rates (within 30ms) and low (using the same bandwidth available to the proposed technique but without applied interest management). Hence for purposes of experimentation, six different configurations of the interest management process are considered. Throughout the remainder of this section, these headings are used to refer to the individual configurations:

- A *"maximum-performance"* scenario, which uses all available bandwidth. Since the environment developed in Chapter 5 is deployed over a local area network for testing purposes, this results in an extremely high update rate on all clients and subsequently fluid motion for all entities within the environment.

- A *"no interest management"* configuration that throttles the available bandwidth to a more realistic value for an Internet-based, large-scale virtual environment. The subsequent result is less fluid motion for all entities as the delay between packets increases. The delay between updates is set at 1330ms, resulting in a total of ~150 packets / second. The following configurations for the proposed method and current state of the art approximation seek to use this same packet rate, but allocate resources more efficiently to provide a higher quality of simulation.

- A *"proposed at 100%"* configuration. In this case, the bandwidth between nodes remains the same as the no interest management case. However, the interest management approach is deployed to more efficiently allocate this available bandwidth. The coefficients used to configure the method are used as determined in Section 5.6.1 and seek to provide the best performance possible. Multicast groups are configured as per Chapter 5, restated here in Table 2 for comparative purposes with Tables 3 and 4 (total for this configuration ~150 packets/sec):

Group Update Rate (ms)	Total Clients	Total Packets / Second
2400	190	79.17
800	2	2.50
500	2	4.00
100	2	20.00
80	2	25.00
50	1	20.00

Table 2: Update Rates used for Proposed Method at 100%, and Current State-of-the-Art Approximation.

- A *"current state of the art approximation"*. The proposed technique is modified as described in the previous section to eliminate the spotlight-based aspect, and hence represent the interest management process with the spotlight enhancement removed. This uses the same multicast arrangement as the proposed method at 100%.

- A *"proposed at 80%"* configuration. This reduces the available bandwidth by roughly 20% from the no interest management, proposed at 100%, and current state of the art approximations. The exact percentage is measured by metrics discussed in the next section. This enables some consideration of how well the system performs as network performance descreases. In order to enable this reduction in bandwidth, the multicast group update rates are decreased, to match the configuration shown below (total ~117 packets/sec):

Group Update Rate (ms)	Total Clients	Total Packets / Second
3000	190	63.3
1500	2	1.3
800	2	2.5
200	2	10
100	2	20
50	1	20

Table 3: Update Rates used for Proposed Method at 80%

- A *"proposed at 40%"* configuration, which reduces the available bandwidth still further, to 40% of the original amount, as shown below (total ~70 packets/sec):

Group Update Rate (ms)	Total Clients	Total Packets / Second
5000	190	38
2000	2	1
1500	2	1.3
700	2	2.9
300	2	6.7
50	1	20

Table 4: Update Rates used for Proposed Method at 40%

A total of 200 clients (the subjects involved in the experiment, and 190+ simulated clients) exist in the virtual world, which mirrors a single grid-cell as described in Section 3.3.2.1. This value is chosen since it is beyond the simultaneous, per-frame update capacity of existing environments (and will thus stress interest management), whilst remaining within the processing limitations imposed by available hardware. Deploying each of these six configurations across both grouping and tag activities results in a sequence of twelve experiments, which form the basis for the results presented in Section 6.3. To provide an evaluation of interest management during these tasks, several methods of evaluation are considered:

- Asking users to state a simple preference for the technique used in a certain experiment (which has drawbacks due to the sequence of the experiments impacting perceived results).

- Asking users to rate each experiment on an arbitrary scale (again, with drawbacks due to experimental sequence, and more critically being too vague an assessment method).

- Asking users to estimate the performance of the interest management more directly by estimating the delay between client updates within each experiment. Ideally in the case of the proposed method this should result in lower estimations of delay for the proposed method versus current state of the art approximations given equivalent bandwidth availability.

The third approach is adopted since it provides the most meaningful data for analysis, but has several drawbacks. Firstly, for inexperienced users, explaining the effects and meaning of various update rates presents a minor obstacle. Thus in order to obtain an approximation of users preference and ability to detect change between techniques, participants are shown an ongoing illustration of clients at a steadily oscillating update rate, varying between 30 and 2000ms over a period of 30 seconds. The current update rate is stated on the display. This is provided in response to preliminary testing wherein users expressed difficulties translating their view of the environment to an estimation of the delay between client updates. Figure 47 shows users participating in an experiment. The large screen shows an ongoing illustration of the effects of various update rates on client behaviour, and is used as a reference by participants for estimating the delay during various experiments. A handout and survey, given in Appendices I and II respectively, details the task and is used to collate their responses.

Figure 47: Users participating in an experiment within the shared environment.

6.2 Performance Metrics

Having established in the previous section the multicast configurations to be used for various experiments, this section now considers the methods used to analyse these configurations to determine the true network load induced. Whilst an approximation is possible by simply considering the total packets / second implied by the configurations shown in Tables 2, 3 and 4, it is important to evaluate them more rigorously to ensure no unforeseen factors have an impact on the effectiveness of the method. One such factor is the cost in subscribing and unsubscribing to multicast groups, detailed later in this section.

The most generic bandwidth consumption tests involve measures of packets sent between nodes / second and bytes sent between nodes / second, with the majority of experiments favouring a measure of packets due to the application independence of such an approach. Consequently, measures of both these factors are worthwhile experiments. Whether these factors may be determined mathematically depends on the nature of both simulation and interest management approach; the key factor being whether the movement of clients (essentially the only pseudo-random factor) affects the number of packets sent. As an example, one interest manager may update all clients within a given radius of one another, and will hence consume more bandwidth as clients pack into a small virtual space. By comparison, another might define sets of clients and various resolutions and place them into groups depending on proximity, while keeping bandwidth use constant, such as the proposed technique. Since we intend to evaluate basic aura-based approaches that fall into the first category, alongside the proposed method, it becomes necessary to implement bandwidth measurement tools within the system alongside mathematical predictions.

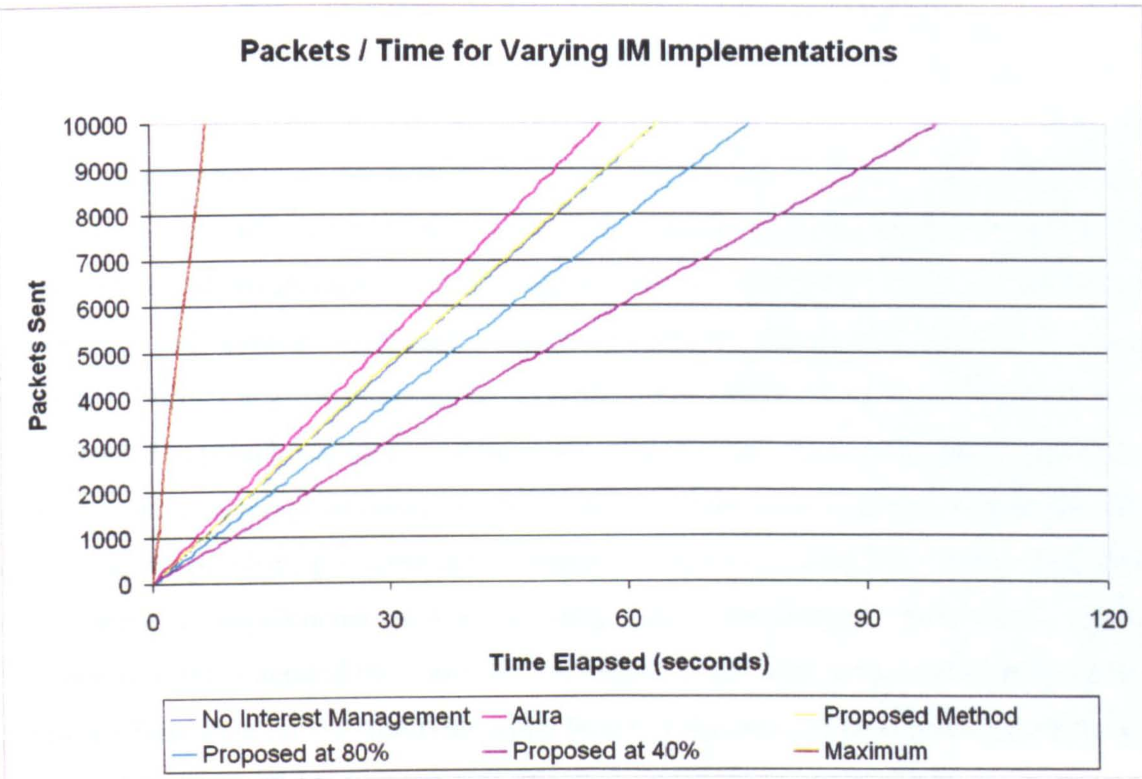
Performance metrics may be split into measures of distinct variables at run-time. Of these, the foremost is bandwidth consumption, since the experiments defined previously seek to perform an analysis of the proposed method against the current state of the art approximation, given the same bandwidth overheads. However, a generic measure of bandwidth consumption itself fails to provide sufficient information; since the underlying network infrastructure is also modified between techniques, with the proposed method likely to induce more multicast (un)subscriptions, which have an inherent cost. Thus two measures are presented with respect to network load: The bandwidth consumed, and multicast subscriptions / time.

- *Bandwidth consumed* is measured in terms of packets / second, rather than kilobytes / second. Since our simulation uses a packet size typically smaller than that of more immersive environments (see Figure 37 for an illustration of packet content), it proves misleading to supply results that could be erroneously compared to kb/second output from other systems.

- *Multicast subscriptions / time* is a measure of the additional network load imposed by rapid changes in the object(s) of interest to the client. Typically this value may be expected to increase in crowded situations or as a consequence of the user moving or reorienting their view quickly. Such behaviour is easily identified as a spike in the number of subscriptions – a critical comparison in evaluation is how well the proposed method handles such behaviour and how well it compares to the current state of the art approximation.

Additionally, measured performance of the simulation in terms of frames / second when using more rigorous interest management approaches allows consideration of the performance impact of the proposed method. Whilst this remains an issue partially in the realms of implementation and hardware (the multithreaded nature of the architecture would suggest improved performance on dual-core systems), it is important to demonstrate a consistent framerate above the threshold by which it may interfere with the perceived quality of the simulation. Failure to provide such a framerate may lead to results for experiments in the subsequent section being influenced. In order to ensure this is the case, all clients are modified such that the framerate must remain a stable 36 frames / second, and if this ceases to be the case, a warning is generated. To ensure a constant quality of simulation to all clients, identical hardware is used for all machines, and the maximum framerate capped at 36 frames / second.

6.2.1 Bandwidth Consumption



Results shown are for an environment with 100 simulated clients, plus one live client used to record the data. The live client follows a set path through the environment for each time, and the downstream data (packets arriving to the clients port) are recorded and time stamped. The resulting log file is parsed into the above graph. The six experimental configurations used in later experiments are thus measured in terms of the network data created, in order to compare them accurately in terms of bandwidth consumption. Whilst the bandwidth is intended to be set by direct configuration within the system, hardware and networking restrictions may place additional limitations on packet rates and thus a comparison using measured network data is worthwhile. Evidence for this can be first seen in the *maximum* configuration, which represents a simulation wherein the packet rate is limited only by available hardware. A distinct reduction in packet rate occurs after approximately four seconds, and may be accounted for by a bottleneck forming within the network. Analysis of this method over a longer time period suggests the rate stays constant beyond this initial reduction. Asides from this exception, all approaches exhibit highly linear behaviour with average correlation coefficients $1 \pm 1\text{x}10^{-5}$. However, the *aura* configuration, representing the current state of the art, suggests higher bandwidth consumption than the two *proposed* and *no interest management* settings. This is despite all three methods being intentionally configured to use a similar amount of bandwidth.

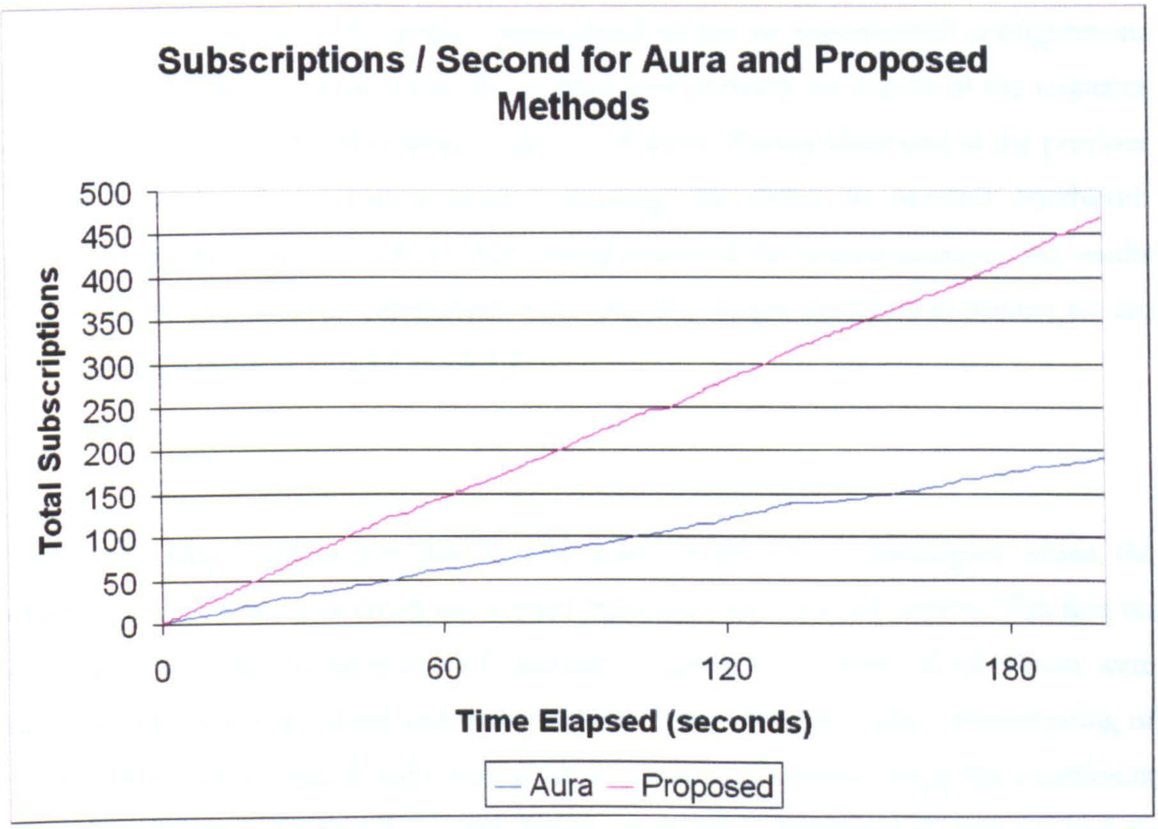
Such a discrepancy may be accounted for by several factors. Firstly the difficulty in configuring three differing approaches to consume identical amounts of bandwidth must be considered; although values used in configuration (see Appendix) are correct additional processing and network overheads, which are initially transparent, may have some influence. Further analysis indicates the increased consumption may be a consequence of the threaded approach resulting in update rates slightly larger than those directly specified, since execution time and shortage of CPU resources may have an effect. Regardless, this represents no problem, provided the increased bandwidth is considered when evaluating the perceptual experiments performed using this configuration. Additionally, slight variance from the intended 80% and 40% of original bandwidth examined when using the proposed technique can be observed, again as a consequence of underlying hardware and software limitations. Table 5 summarises the deviances from intended value.

Interest Management Method	Packets/ Second	% Bandwidth
Maximum Update Rate	1420.1	909.7%
No Interest Management	156.1	100.0%
Aura / Current State of the Art Approximation	175.9	112.7%
Proposed Method	155.9	99.8%
Proposed Method at 80%	133.2	85.3%
Proposed Method at 40%	101.1	57.5%

Table 5: Summary of total packets sent over time and resulting percentage bandwidth, when compared to the no interest management case.

To obtain measures of packets per second used in Table 5, a linear regression is performed (linear behaviour is expected, assuming the quality of service provided by the network remains constant) on each data set and the gradient of the resulting trend line observed. In all cases a low correlation coefficient indicates accurate approximations, to within 0.4 packets / second.

6.2.2 Multicast Subscription Rate



To measure multicast subscriptions per second, a test client again follows a preset route within an environment containing 100 simulated clients. Multicast subscription calls invoked by the interest manager are recorded and time stamped. Both approaches show a linear increase in subscription calls with time. Slight variance over short time periods may be accounted for by situations where clients are static or persist in the same region of interest, leading to a lower subscription call rate. In the case of the proposed method particular spikes can be accounted for by the viewpoint reorienting rapidly (and thus many new focal selections taking place). Due to the fact the proposed method promotes a more distributed aura, focussed upon a potentially distant target of focal selection, and may assign new groups based on a reorientation of the viewpoint, a higher subscription rate is induced. Sampling over a 3 minute time period suggests a rate of 2.3 subscriptions / second for the proposed method, and 1.5 subscriptions / second for the aura approximation – a notable 53% increase for the proposed method.

6.3 Experimental Results

With an understanding of the precise implications of the six experimental configurations described in Section 6.1 established, this section now presents the results of the sequence of twelve experiments for two separate groups of users. Having illustrated in the previous section a series of quantitative results, showing the ability to monitor bandwidth consumption and adjust it by appropriate configuration of the interest manager, the results for two series of perceptual experiments following the designs described in Section 6.1 are now presented in Sections 6.3.1 and 6.3.2.

Two series were used in part due to availability of subjects to participate within the experiment, but also to ascertain the impact experience had with the system. The first set of results, presented in Section 6.3.1, represent a group of 5 users, all of whom were familiar with the environment and user interface. Hence they had some understanding of the nature of the system, though were given no more information during the experiment than the subjects in Section 6.3.2. This larger group of 10 simultaneous users all had no prior experience of the environment before the experiment, and were given no information before the experiments (beyond that shown in Appendix I).

6.3.1 First Experiment Series

The first group of 5 subjects was comprised of subjects familiar with the research and relatively experienced in the user interface of the virtual environment. All 5 purported to have participated within virtual environments before either as a result of research or gaming / collaboration in the survey provided (included in Appendix). The results are shown in Figure 48.

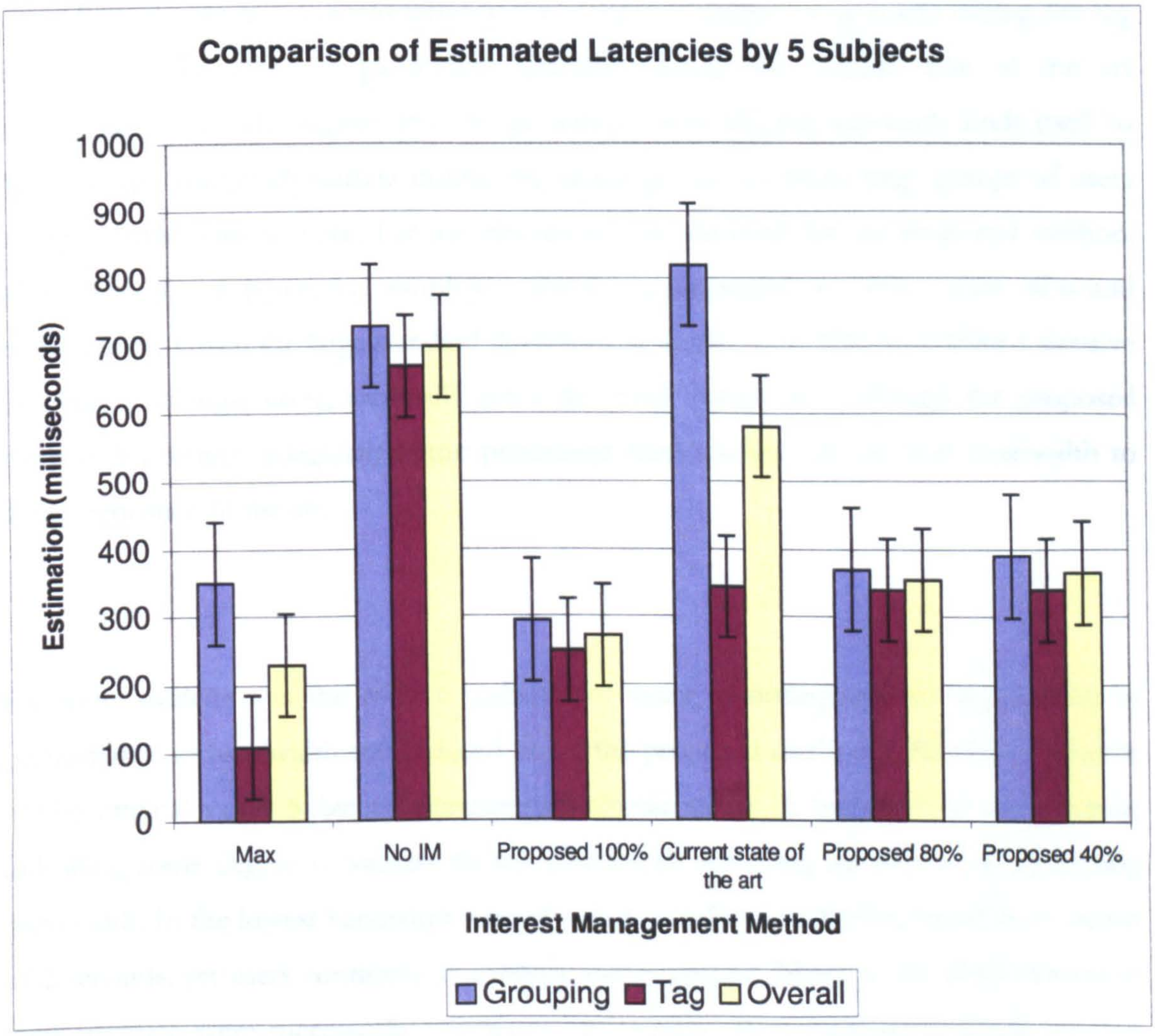


Figure 48: Results of the Interactive Experiment for 5 Subjects.

In this case, all subjects were familiar to a limited degree with the environment and nature of the research. However, no information was given regarding which interest management paradigm was used in which experiment. All subjects were able to distinguish between the "No IM" configuration (set at a global update rate of 1000ms) and the "maximum" configuration, at a 30ms update rate. However, whilst all were able to correctly identify the higher update rate, the accuracy with which they were able to do so varied substantially between subjects - the standard deviation of results for the grouping experiments was 320ms. Interestingly, this was reduced to 250ms for the second series of tag experiments, suggesting the ability of users to accurately measure the update interval was increasing with experience within the environment.

Of additional note is a uniform trend in the perceived latency being lower during the tag experiment. The effect is particularly dramatic during the current state of the art approximation, which suggests that the proximity-based filtering approach lends itself to the task, and conversely suffers during the grouping activity when large groups of users occupy a small virtual space. Far less discrepancy is observed for the proposed method, which due to its perceptual emphasis, should be expected to offer better all-round performance. Given the large standard deviations in results, it is difficult to infer a decisive preference amongst users, especially given the small sample set, although the proposed method did receive marginally better preference from users at an identical bandwidth to the current state of the art.

Far more evident was the relative inability for users to distinguish any degradation in performance as bandwidth was reduced under the proposed method to 80 and 40 percent of the original value. Subjects approximated similar values in both sets of experiments, indicating some degree of success by the method in obscuring the effects of decreasing bandwidth. In the lowest bandwidth case, the global update rate had increased to in excess of 2 seconds, yet users continued to estimate values around 300ms as the likely objects of their attention were successfully identified. The comparatively high prediction in the "no IM" case suggests that users were capable of identifying higher update rates where evident, lending credence to this conclusion.

6.3.2 Second Experiment Series

In order to obtain a larger sample, the experiment was repeated with a further 10 subjects, none of whom had any familiarity with the research or virtual environment used (although all but one subject expressed some degree of familiarity with virtual environments in an entertainment context). In this case, it was necessary to present the subjects with a brief presentation and allow them five minutes to familiarise themselves with the interface used to navigate the virtual world. Before beginning the session, all users were asked to verbally confirm they understood the task they needed to accomplish, but no information was given regarding the nature of the interest management within the system. Again, two series of six experiments were performed (grouping with six interest management approaches, and tag with six interest management approaches). The results are shown in the next section.

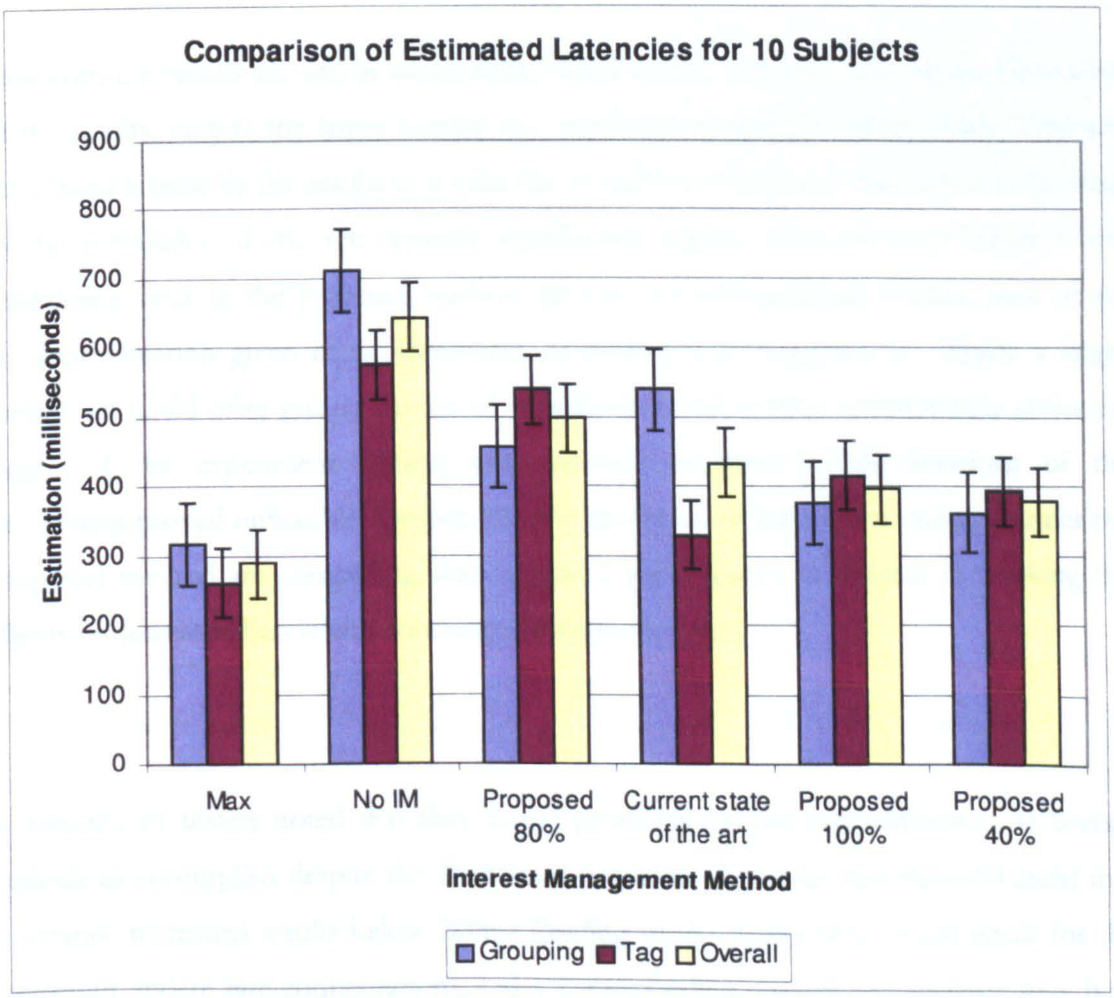


Figure 49: Results of the Interactive Experiment for 10 Subjects

All subjects were unfamiliar with the environment prior to the experiment series, relying on the illustration of the effects of latency described in the previous chapter as a basis for estimation. Again, the current state of the art approximation demonstrated a more significant variance between the two experiments, performing well in the tag game, but less effectively during the grouping activity. Subjects again demonstrated a clear ability to distinguish between the two extreme cases; although the trend of lower latency predictions this time shifted to favour the grouping activity in most cases. Once again a marked reduction in bandwidth using the proposed method to 40 percent of the original value went relatively undetected, with users in fact reporting *lower* latencies than for the 100 percent bandwidth scenario. Furthermore, the results for the proposed method at 80 percent of the original bandwidth suggested a higher latency than the proposed method at 40 percent – a scenario that should be impossible given ideal conditions for the experiment.

However, the results are within experimental error of one another. The standard deviation of the results, despite the larger sample size, remained around 270 milliseconds. This was influenced heavily by the results of a minority of subjects reporting values non-conforming to the remainder of the set; typically significantly higher. Such an error infers a low confidence value in the proposed method directly outperforming the current state of the art approximation given equal bandwidth, in both sets of experiments. Ideally a larger sample set could offer greater degree of confidence in the results, unfortunately given the length of the experiment session and practical concerns further iterations of the experiment proved unfeasible. Despite this, the results when latency was reduced under the proposed method are compelling, and suggest a good degree of success in masking the effects of decreased bandwidth consumption from the user.

A majority of testers noted that they found providing an accurate estimation of latency difficult to accomplish despite the illustration provided. Typically, this was evidenced in a reluctance to record results below 200ms (leading to the higher than actual result for the maximum update rate configuration), and a corresponding tendency to estimate very high latencies as being lower than the actual value. Thus it is important to stress the use of these values as an indication of preference, rather than an absolute measure of latency. Subjects uniformly reported that their estimations were based on relative comparison between iterations of the experiment, and thus a lower estimation, which implies 'smoother' motion experienced, may be taken as inferring a preference towards a given configuration. To further elaborate this concept, the results for both experiment series are combined, and user preference determined by comparing estimations for both the proposed method and current state of the art, taking the lower value to imply a preference. This is shown in Figure 50. A consistent trend in favour of the proposed method can be observed.

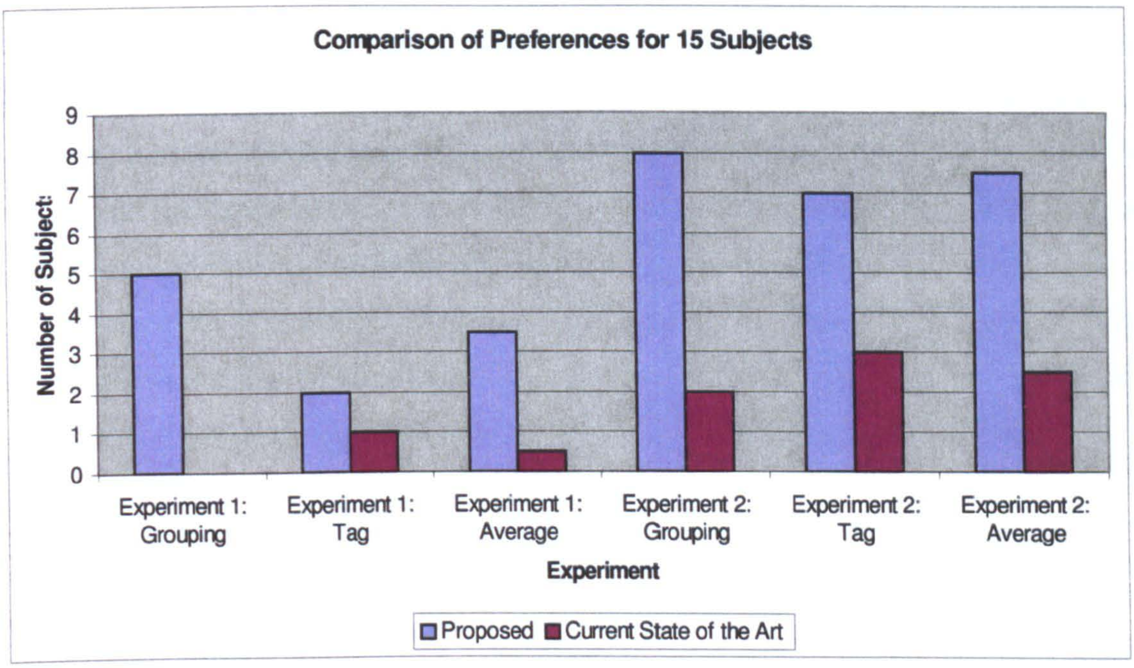


Figure 50: Preference expressed by subjects (based on approximation of latency).

Finally, it is worth noting that, at equal bandwidth, the users more experienced with the system appeared to rate the proposed method more highly than the inexperienced users. It was suggested that more familiarity with the user interface encouraged users to reorient the viewpoint more frequently and instinctively. Subsequently, this results in increased performance for the proposed approach since the user integrates their visual search task into the system, rather than moving their focus of attention around the monitor screen whilst keeping their viewpoint static. This may indicate potential further gains for the proposed approach in more sophisticated environments, wherein the user interface allows for more natural interaction. In general, the relationship between user interface and visual attention is worthy of considerable further research.

6.4 Summary

In this chapter, the implementation described in Chapter 5 has been analysed through a series of perceptually oriented experiments, reinforced by measures of bandwidth consumption. The goal has chiefly been to establish whether users were capable of detecting change within the implementation, and thus study whether change detection occurred between the proposed method and current state of the art, and in situations where available bandwidth decreases. The results have shown that users are capable of identifying a significant global change in the update rate, and are particularly able to note negative effects where no interest management is applied and bandwidth decreases. It has also been shown that users continue to rate the proposed method similarly as bandwidth is decreased substantially, due to the focus of resources upon the most salient objects. Whilst such results are inconclusive, due to factors including the subjective nature of human perception, and limited availability of participants, the majority of testers exhibited a preference for the proposed method over the current state of the art approximation. The final chapter now goes on to consider the implications of these results, and suggest means by which the proposed approach may be improved, alongside consideration of potential limitations.

7 Conclusions and Future Work

Having provided in the previous chapter documentation of a series of experiments intended to evaluate the spotlight-based approach to interest management, this thesis now concludes by considering the successes and failures of the proposed method in light of both this experimental evidence and with respect to the field in general. In Section 7.1, the contributions of this thesis to both interest management, and more generally the application of visual attention within virtual environments, are discussed. This includes consideration of the integration of the proposed method with existing techniques, and the benefits it provides. Section 7.2 considers the limitations of the method, in order to define the context and scope within which the proposed method may be expected to offer improved performance over existing techniques. Following on from this discussion, Section 7.3 considers potential avenues for future work, both with regard to the configuration of the spotlight approach and more generally with respect to interest management. Finally, Section 7.4 presents a final summary and the conclusions of the research.

7.1 Specific Contributions

This thesis has, throughout, emphasised the link between visual attention and interest management. Following a comparison of models of visual attention to existing interest management approaches, an initial hypothesis suggesting that a spotlight-based refinement of aura-based interest management would offer improvements over the current state of the art has been reinforced by experimental evidence. In particular, the potential gains of a spotlight approach in terms of compatibility, composability and processor efficiency have been discussed in Sections 3.4 and 3.5. The evidence obtained by experiment (as presented in the previous chapter) strongly suggests that users perceive an improved update rate within a system that allocates resources based on a spotlight model, compared to the current state of the art. This conforms to the logical argument presented in Chapter 3, which suggests that perceptual relevance is poorly approximated by current methods. All testers exhibited the ability to identify a difference between uniform update rates of 30ms and 1200ms. This suggests a common ability to estimate and detect latency. The accuracy of estimations varied substantially between participants; but lends credence to their ability to distinguish performance between different implementations. In light of this capability, failure to note a significant difference as bandwidth was reduced under the proposed method to 40% of the original value shows a promising potential to eliminate less salient clients from the network resource overheads, and support a substantially lower (and thus less expensive) global update rate.

The spotlight approach only requires information regarding the position of clients within the locale, which is advantageous when compared to the existing visual attention based approach of Beeharee described in Section 3.3.3. Zyda [ZS00] stresses the need for objects within virtual environments to be described in a compatible and ubiquitous form ('composability', as described in Section 3.4.2). In placing requirements on object definition to perform interest management, limitations are placed on the composability of objects within the environment, and the ability to import new content. The spotlight-based approach considers only relative position to the client as an extrinsic factor, and thus integrates well with the notion of composability. In disregarding many bottom-up characteristics whilst still offering positive results, the spotlight approach proves capable of providing effective filtering in the absence of such information. Furthermore, the approach described in Chapters 4 and 5 illustrates how a spotlight-based filter may be applied with a minimum of iteration and scene analysis, and thus infers little impact on overall system performance.

When considering integration with existing methods, the composability and minimal overheads associated with the method present further advantages. The method may be seen as operating under a locale or grid-based paradigm similar to that defined by Watsen and Zyda [WZ98]. In this case a replication of the process would be used within each locale – the experiments described in Section 6.1 can be seen to be an example of the method operating within a single locale. As such, it may be seen as a refinement of the aura-based technique that seeks to specifically affect the relationship between distance and salience implied by all aura-based approaches. Whereas existing implementations of aura-based interest management as those described in Section 3.3.2.2 fail to directly acknowledge the relationship to visual attention, this thesis has demonstrated a spotlight model of attention to offer benefits over a linear relationship between distance and salience. It also complements the work of Beeharee, who seeks to analyse characteristics other than relative position, and may be integrated alongside a multimodal modelling approach.

In implementing the proposed method, this thesis has demonstrated the successful creation of an architecture wherein saliency values may be explicitly mapped to levels of resolution within the network via multicasting. The architecture is only limited in capacity by available multicast addresses; comparing favourably to the client-server architecture common to existing commercial environments as described by Bauer et al. [BRS02], which has numerous demonstrable limitations, as discussed in Chapter 3. Notably, though, it achieves such capacity by the use of peer-to-peer design, which suffers limited uptake by commercial environments due to the decrease of security in moving from a client-server based approach, as noted by Hu and Liao [HL04]. Ultimately, however, the benefits of peer-to-peer approaches in scalability, coupled with advances in peer-to-peer security and trust, would suggest a long-term shift towards such architectures.

Hence, in summary, this thesis has contributed compelling evidence that a spotlight model of visual attention provides substantial scope for refining interest management. The advantages of a spotlight model have been exploited to describe a composable, computationally inexpensive refinement to existing radial aura-based approaches. By means of a successful implementation, the ability to correlate multicast groups to saliency values has also been demonstrated, alongside for the potential for integration into other perceptual techniques.

7.2 Limitations of the Spotlight Approach

Due to the relationship between the approach and visual attention, some limitations are implied with regards to application. Section 2.4.2 presents an example of a limitation of the spotlight model in the case of a user instructed to focus deliberately on a single area within a target. In this case, overriding top-down behaviour is shown to control the form and deployment of the spotlight. This is an issue common to all approaches that seek to exploit perception, including the rendering approaches described in Section 3.2. Whilst the methods are effective in the case of a user passively observing, or performing real-world tasks within the environment (such as the grouping and tag examples in Chapter 6, where the method is shown to be effective), supplying them with a task deliberately intended to focus their attention on entities deemed less relevant by the interest manager may cause the approximation of visual attention to break down. Thus, with regards to application, the approach may be seen as limited to virtual environments that represent the real world, wherein users are expected to exhibit real-world behaviour.

The system also places constraints on user interface devices. Having been developed for keyboard and mouse interaction, with the mouse used to realign the direction of view, a relationship exists between the ability of the interest management to detect covert attention shifts, whilst overt shifts are carried out by the user, who reorients their viewpoint via the mouse. A spotlight model would be expected to perform less successfully in situations where view shifts are impossible (e.g. observers watching a demonstration), since in the absence of the ability to deploy overt attention through interacting with the environment, observers would be expected to shift their focus between entities rather than reorient the viewpoint. Conversely, however, more sophisticated interaction tools such as head-mounted displays may offer improved performance, and represent an avenue for future work.

One demonstrated weakness regarding the performance of the spotlight method is an increased demand on the network placed by more frequent multicast subscription calls (see Section 6.2.2). Such subscription calls increase with the number of clients within the locale, since the target of interest is likely to change more frequently. Furthermore, when compared to the standard aura, the proposed method forces more rapid shifts in selected objects of interest, as a change in view between a distant and near client can invoke a substantial redeployment of resources. This is a problem common to other systems aiming to enhance interest management – though future improvements in multicasting, particularly with respect to group management via approaches such as that of Auerbach et al. [AGK03], should result in increased support for methods seeking to exploit large numbers of multicast groups. This issue is particularly significant when comparing the methods to those used by early systems such as MASSIVE, which were restricted by hardware limitations on multicasting that are no longer a widespread issue. Hence this drawback whilst noteworthy is justifiable in the context of advancing network technology. A more ideal solution, however, may seek to modify the total of available groups based on capacity – with regards to the proposed method this may be done by dynamic modification of the total levels of resolution provided by individual clients. Throughout the experiments this level was fixed at a total of five to provide a balance between performance and providing a realistic network load, resulting in a total of 5 multicast groups per client. Reducing this value with have a subsequent effect on the total subscriptions / minute generated by the simulation, at the cost of decreasing the available levels of resolution.

7.3 Future Work

This research has focussed purely on the visual aspects of perception, however, in a virtual environment, aural cues also play a significant role in drawing a users attention. Whilst the proposed method may serve adequately in the case that all avatars constantly emit equivalent sounds (e.g. footsteps), more detailed analysis of the relationship between hearing, vision, and attention in the context of perceptual optimisation should prove worthwhile. Some allowance for the need of the interest manager to register non-visible but audible clients is allowed for under the proposed method by the additional, radial aura centred around the client, is additional to the perceptual 'focus'. However, this is a simple approximation, effectively identical to that used in other aura-based approaches. Further refining the form of this aura, and additional consideration in its relationship to aural sources may provide an avenue for accommodating sound more rigorously. The need for such accommodation particularly applies to more sophisticated environments, where sounds are pivotal in providing immersion and interactivity.

The proposed method makes three main assumptions, which whilst backed by limited experimental evidence, should prove the three most important avenues for further enhancement.

- In specific regard to the effectiveness of the focal selection process, it would stand to reason more detailed analysis of the impact of the a and b coefficients which drive the process (see Section 4.3) may yield improved results. It is likely top-down aspects of task would affect the tendency to focus on more distant objects, and thus integrating models of attention into dynamic values for these coefficients would appear an interesting avenue for further work.

- The relative base relevance values of objects within the virtual world (R_{focus} and R_{obj_i}) are assumed constant for all clients regardless of context. Whilst this is valid for the experiments in the previous section, since all entities are identical, it raises questions regarding the suitability of the method to more complex environments, where entities may possess widely varying characteristics, ranging from the appearance of an avatar to their behaviour and context. In particular, the integration of the approach proposed within this thesis, alongside other proven perceptual techniques such as that of Beeharee [BWH03], may offer a solution for calculating R_{focus} and R_{obj} that builds upon the strengths of the approach within this thesis in dealing with large crowds, to add more bottom-up measures of relevance.

- Finally, the multicast arrangement described in Section 5.6.2 is arbitrary, and whilst based on perceptual research such as that of Sears and Pylyshyn [SP00], would require a significant amount of further work in analysing the perceived effects of different configurations on users to better optimise the proposed method. It may prove that a dynamic approach would offer improved results, although questions exist over the impact on performance of re-configuring multicast groups at run-time.

The spotlight model of attention applied by the proposed approach is a simplification of the complex process of visual attention. It may thus be argued that models that encompass feature-integration theory provide a more comprehensive and precise model for visual saliency. In particular, they encompass phenomena such as increased saliency for objects near the periphery of the field of view, with certain properties (as an example, moving objects in the periphery of the field of view are shown to have high saliency; whereas sensitivity to colour is reduced, as mentioned in Chapter 2). Such models, whilst offering the potential for more precise saliency value calculation, imply additional overheads in processing to accommodate such characteristics. Furthermore, they require tighter integration to a specific virtual environment, since avatar characteristics must be analysed to determine saliency. By comparison, the proposed approach filters based on relative position alone, and thus removes itself from application-related complications. Despite these drawbacks of more sophisticated models, the application of such approaches can be expected to offer improved bandwidth reduction, and is thus an obvious avenue for further research.

One additional issue may present itself when faced with environment involving more complex geometry. Whilst within the series of experiments described in Chapter 6, the objects exist on the same vertical plane, the addition of structures and other scenery would allow for a greater vertical offset. In this case, phenomena as described in Section 2.2 wherein users tend to focus on their current plane and give less regard to objects above and below may present an issue. The implementation described in Chapter 4 is based around a radial concept, where objects are given equal bias on both vertical and horizontal axis. Reducing the vertical component (in effect, 'flattening' the aura), may lead to enhanced performance in these situations by better accommodating this phenomenon.

An ultimate goal for future work would be the integration of the state of the art in visual-attention research, into a flexible, high-performance peer-to-peer framework applicable to any large-scale VE. The rapid evolution of this state of the art implies a degree of separation between attentional model and network framework, which this thesis has demonstrated as viable within the implementation discussed in Chapter 5. Effectively, any visual-attention model may produce saliency values, which may then be transcribed to multicast groups, or, more generally, network resource allocation.

7.4 Summary and Conclusions

This thesis has presented a novel enhancement to interest management, which utilises a spotlight model of visual attention. In Chapters 2 and 3, a background was provided which outlined the limited use of theories of visual attention by existing interest management paradigms, despite a close conceptual link. Subsequently, a spotlight-based metaphor was identified in Section 3.5 as possessing several potential benefits, whilst being largely compatible with existing approaches. The spotlight model described in Chapter 4 considers the interaction between user and environment to reorient their viewpoint to be an approximation of overt attention, and thus seeks to emulate covert attention by means of a spotlight, focussed on a single most likely target, whilst generating saliency values for remaining objects based on their relative position to both user and covert object of focus.

In Chapter 5, the implementation of the method was described via the creation of a distributed virtual environment intended to fully accommodate the facility to implement and evaluate interest management approaches. By populating this environment with a group of simulated clients, it was shown possible to demonstrate the effects of overcrowding without the need to accommodate large numbers of real-world users. Subsequently, the spotlight approach was evaluated against the existing, linear relationship between distance and saliency implied by aura-based approaches, and shown to be preferred by 80% of users attempting to perform tasks in an overcrowding scenario. Further experiments examined the ability to decrease bandwidth without perceived effect, with users continuing to report performance equivalent to the existing aura-based approach with bandwidth reduced by up to 60%.

The limitations of this study reflect the implicit difficulties of any quantitative research within the field of human attention. Certainly the wealth of research presented in Chapter 2 offers no definitive approach to accurately measuring attention and relevance, and thus the experimental results given in Chapter 7 are presented purely as evidence towards the effectiveness of the proposed technique, rather than as conclusive proof. Certainly the high percentage of users expressing blind preference for the proposed approach is compelling;

as is the notable difficulty they had perceiving change as bandwidth was reduced substantially. However, as is frequently the case regarding research into visual attention, a larger sample of users, though impractical, may still offer further insight. In particular, whilst users were assessed with regards to their vision and basic experience within virtual environments, a larger sample may offer the ability to explore additional characteristics that may impact the effectiveness of the method, such as gender and age (both of which have been shown to have some impact on attention [LLC03]).

Returning to the applications described in Chapter 1, perceptual refinements to the interest management process such as those presented in this thesis offer the ability to deploy available bandwidth with improved efficiency. Subsequently, this allows for additional bandwidth to be utilised in order to provide more depth of interaction, or support for greater numbers of users. In the case of the user socialising within an environment such as Second Life [Lin06], this promises the capability to view vast crowds of other users and participate in large online gatherings. For the user within the military simulation, it promises fluid motion and realistic behaviour of both allies of enemies in large-scale encounters. For the online gamer, it promises greater interactivity and games that support thousands of users without restriction. Throughout this thesis, parallels between human attention and interest management have been drawn; ultimately, both serve the same purpose – to eliminate perceptually irrelevant data. Hence, the convergence of theories regarding visual attention with interest management promises to offer a comprehensive solution to the problem of providing users of distributed virtual environments with truly immersive experiences. Such interest management may be seen as an essential component within future distributed virtual environment technology. This technology, over upcoming decades, promises worlds with imperceptible boundaries and limitless applications.

REFERENCES

[AGK03] AUERBACH, J., GOPAL, M., KAPLAN, M., & KUTTEN, S. 2003. Multicast group membership management. *IEEE/ACM Transactions on Networking*. Vol. 11, No. 1

[AS01] E. ARROYO & T. SELKER. 2001. Interruptions: Which are the Less Disruptive, *MIT WhitePaper.*

[AWZ98] ABRAMS, H., WATSEN, K., & ZYDA, M. 1998 Three-tiered interest management for large-scale virtual environments. *Proceedings of 1998 ACM Symposium on Virtual Reality Software and Technology (VRST'98)*

[BBN95] BLACKMORE, S.J., BRELSTAFF, G.,NELSON, K., & TROSCIANKO, T. 1995. Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception,* Vol. 24, pp. 1075–1081

[BCP03] BROWN, R., COOPER, L., & PHAM, B. 2003. Visual attention-based polygon level of detail management. *Proceedings of the 1st international Conference on Computer Graphics and interactive Techniques in Australasia and South East Asia.* GRAPHITE '03. ACM Press, New York

[Bea04] ARISTOTLE. On Sense and the Sensible. circa 350BC. (English translation by J. I. Beare, University of Adelaide, 2004)

[Bro59] BROADBENT, D. 1958. Perception and Communication. *New York: Pergamon*

[Bro82] BROADBENT, D. 1982. Task combination and selective intake of information. *Acta Psychologica,* Vol. 50, pp. 253–290.

[BRS02] BAUER, D., ROONEY, S., SCOTTON, P. 2002. Network infrastructure for massively distributed games, Proceedings of the 1st workshop on Network and system support for games, Bruanschweig, Germany. p.36-43

[BWA96] BARRUS, J.W., WATERS, R.C., & ANDERSON, D.B. 1996, Locales: Supporting Large Multiuser Virtual Environments, *IEEE Computer Graphics and Applications*, Vol. 16, No. 6, pp. 50-57

[BWH03] BEEHAREE A. K., WEST A. J., HUBBOLD. 2003. R.:Visual attention based information culling for distributed virtual environments, *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 213 - 222.

[Can86] CANNY, J. 1986. A Computational Approach To Edge Detection, *IEEE Pattern Analysis and Machine Intelligence*, Vol. 8 pp. 679-714

[CEY04] CONNOR, C., EGETH, H., & YANTIS, S. 2004. Visual Attention: Bottom-Up Versus Top-Down, *Current Biology*, Vol. 14, pp. 850–852

[CH04] CHUI, Y. & HENG, P. 2004. Attitude dead reckoning in a collaborative virtual environment using cumulative polynomial extrapolation of quaternions: *Research Articles. Concurr. Comput. : Pract. Exper.* Vol. 16, No. 15 pp. 1575-1599.

[CH93] CARLSSON, C., & HAFSAND, O. 1993. DIVE: A Multi-User Virtual Reality System. *Proceedings of the IEEE Virtual Reality Annual International Symposium*, pp. 394-401.

[Che53] CHERRY, E.C. 1953. Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America.* Vol. 25 No. 5, pp. 975-979.

[Che96] CHESHIRE, S. 1996. It's the latency, stupid. Online at: http://rescomp.stanford.edu/~cheshire/rants/Latency.html (Accessed 12/07/06)

[CK66] CAMPBELL, F.W. & KULIKOWSKI, J.J. 1966. Orientation selectivity of the human visual system. *Journal of Physiology*, Vol. 187, pp. 437-445

[CK81] CSIKSZENTMIHALYI, M. & KUBEY, R. 1981. Television and the rest of life: A systematic comparison of subjective experience. *Public Opinion Quarterly*, Vol. 45, No. 3 pp. 17-328

[Cla76] CLARK, J. 1976. Hierarchical Geometric Models for Visible Surface Algorithms, Communications of the ACM, 19, 10, 547-554.

[Cor75] CORNSWEET, T. N. & CRANE, H. D. 1973. Accurate two-dimensional eye tracker using first and fourth Purkinje images, *Journal of the Optical Society of America*, Vol. 63, No. 8, pp.921-928

[CS02] CORBETTA, M. & SHULMAN, G. L. 2003. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews,* Vol. 3, pp. 201–215

[CS78] CRANE, H. & STEELE, C. 1978. Accurate three dimensional eye tracker. Applied Optics, Vol. 17, pp. 691-705

[DB89] DRIVER, J., & BAYLIS, G. C. 1989. Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance,* Vol. 15, pp. 448–456.

[Deb98] DEBEVEC, P. 1998. Rendering Synthetic Objects Into Real Scenes: Bridging Traditional and Image-Based Graphics With Global Illumination and High Dynamic Range Photograhy, *Proceedings of SIGGRAPH 98, Computer Graphics Proceedings, Annual Conference Series, Addison-Wesley* pp. 189-198

[DSL06] DSL Reports. September 2006. Online at: http://www.dslreports.com/archive (Accessed 08/09/06)

[Dun84] DUNCAN, J. 1984. Selective attention and the organization of visual information. *Journal of Experimental Psychology: General,* Vol. 113, pp. 501–517.

[DW05] BRETT-KING, D. & WERTHEIMER, M. 2005. Max Wertheimer & Gestalt Theory. *New Brunswick (USA) and London (U.K.), Transaction Publishers*

[DW05b] DUNWELL, I. & WHELAN, J. 2005. Perceptually-Oriented Interest Management in Large-Scale Networked Virtual Environments. *In proceedings of EGUK Theory and Practice of Computer Graphics '05*, pp. 83-89

[DW97] DENSLEY, D. & WILLIS, P. 1997. Emotional Posturing: A Method Towards Achieving Emotional Figure Animation, *Computer Animation '97, Geneva, Conference Proceedings (IEEE June 1997)* pp. 8-14

[EH72] ERIKSEN, C. & HOFFMAN, J. 1972. Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics* Vol. 12 (1), pp. 201–204.

[EH73] ERIKSEN, C. & HOFFMAN, J. 1973. The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics* Vol. 14 (1), pp. 155–160.

[EK90] EYSENCK, M. W. & KEANE, M. T. 1990. *Cognitive psychology: a student's handbook, Lawrence Erlbaum Associates Ltd., Hove, UK*

[ES86] ERIKSEN, C. & ST. JAMES, J. D. 1986. Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, Vol. 40, pp. 225–240.

[Fel03] FELDMAN, J. 2003. What is a visual object? *Trends in Cognitive Sciences* Vol. 7, No. 6, pp. 252-256

[Fre53] FRENCH, R.S. 1953. The discrimination of dot patterns as a function of number and average separation of dots. *Journal of Experimental Psychology*, Vol. 46, pp. 1–9.

[Fry06] FRYE, C. 2006. Envisioning Cyberspace (A Review), Online at http://www.techsoc.com/envisioning.htm (Accessed 12/04/06)

[FS93] FUNKHOUSER, T. A. & SÉQUIN, C. H. 1993. Adaptive Display Algorithm for Interactive Frame Rates During Visualization of Complex Virtual Environments, *Computer Graphics (SIGGRAPH '93 Proceedings)*, Vol. 27, pp. 247–254.

[FS98] FRECON, E., & STENIUS, M. 1998. DIVE: a scalable network architecture for networked virtual environments, *Distributed Systems Engineering*. Vol. 5 pp. 91-100

[FS98b] Quote taken from FRECON, E., & STENIUS, M. 1998. DIVE: a scalable network architecture for networked virtual environments, *Distributed Systems Engineering*. Vol. 5 p. 93

[GB89] GILLIE, T. & BROADBENT, D. 1989. What makes interruptions disruptive? A study of length, similarity and complexity *Psychological Research*, Vol. 50, pp. 243-250

[GB95] GREENHALGH, C., & BENFORD, S. 1995. MASSIVE: a Distributed Virtual Reality System Incorporating Spatial Trading. *Proceedings of the IEEE 15th International Conference on Distributed Computing Systems*, Vancouver, Canada

[GB99] GREENHALGH, C. M. AND BENFORD, S. D., 1999 Supporting Rich And Dynamic Communication In Large Scale Collaborative Virtual Environments, Presence: Teleoperators and Virtual Environments, MIT Press. Vol. 8, No. 1, pp. 14-35

[GHF99] GILCHRIST, I. D., HEYWOOD, C. A. & FINDLAY, J. M. 1999. Saccade selection in visual search: evidence for spatial frequency specific between-item interactions. *Vision Research.* Vol. 39, pp. 1373–1383

[Gri96] GRIMES, J. On the failure to detect changes in scenes across saccades. 1996. *K. Akins (Ed.), Vancouver studies in cognitive science: Vol. 2: Perception, New York: Oxford University Press.* pp. 89–110.

[Han97] C. HAND. 1997. A Survey of 3D Interaction Techniques, *Computer Graphics Forum* Vol. 16, No. 5, p. 269

[Hav06] HAVOK Engine Official Website. Online at: http://www.havok.com/ (Accessed 11/09/06)

[HCP03] HO, C., CHENG, W., PAN, T., & WU, J. 2003. A user-attention based focus detection framework and its application. *Proceedings of the IEEE ICICS-PCM 2003*, Vol. 3, pp. 1315-1319.

[Hel10] HELMHOLTZ, H. 1910. Treatise on Physiological Optics. *Hamburg and Leipzig* (English translation by J. Southall. 1925). Electronic Edition 2001, *University of Pennsylvania,* Online at http//psych.upenn.edu/backuslab/Helmholtz. Accessed 04/06/06.

[Hel10b] HELMHOLTZ, H. 1910. Treatise on Physiological Optics. *Hamburg and Leipzig* (English translation by J. Southall. 1925) Electronic Edition 2001, *University of Pennsylvania,* Online at http//psych.upenn.edu/backuslab/Helmholtz. Accessed 04/06/06. p. 205

[HHF05] HAASCH, A., HOFEMANN, N., FRITSCH, J., & SAGERER, G. 2005. A multi-modal object attention system for a mobile robot. *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp.1499-1504

[HL04] HU, S., & LIAO, G. 2004. Scalable Peer-to-Peer Networked Virtual Environment *Proceedings of ACM SIGCOMM 2004 workshops on NetGames '04*, pp. 129-133.

[HLS97] HAGSAND, O., LEA, R., & STENIUS, M. 1997. Using spatial techniques to decrease message passing in a distributed VE *Proceedings of VRML'97 (Monterey, CA)*

[HNL06] HENNESSEY, C., NOUREDDIN, B., & LAWRENCE, P. 2006. A single camera eye-gaze tracking system with free head motion. *Proceedings of the 2006 Symposium on Eye Tracking Research &Amp; Applications* (San Diego, California, March 27 - 29, 2006). ETRA '06. ACM Press, New York, pp. 87-94.

[Hop96] HOPPE, H. 1996. Progressive Meshes, Proceedings of SIGGRAPH'96, 30, 99-108.

[HPG02] HOSSEINI, M., PETTIFER, S., & GEORGANAS, N. D. 2002. Visibility-based interest management in collaborative virtual environments. *Proceedings of the 4th international Conference on Collaborative Virtual Environments, Bonn, Germany.* pp 143-144.

[HR87] HUMPHREYS G.W. & RIDDOCH M.J. 1987. To see but not to see. A case study of visual agnosia. *Lawrence Erlbaum Associates, London*

[HY94] Hillstrom, A. & Yantis, S. 1994. Visual motion and attentional capture. *Perception and Psychophysics*, Vol. 55 pp. 399-411

[IEEE95] INSTITUTE FOR ELECTRICAL AND ELECTRONICS ENGINEERS. 1995. IEEE Standard for Distributed Interactive Simulation – Application protocols. *IEEE Standard 1278.1-1995. Piscataway, NJL IEEE Standards Press*

[IK99] ITTI, L. & KOCH, C. 1999. Target detection using saliency based attention," *Proceedings of RTO/SCI-12 Workshop on Search and Target Acquisition (NATO Unclassified)*, Utrecht, The Netherlands, pp. 3.1-3.10.

[Itt05] ITTI, L. 2005. Models of Bottom-Up Attention and Saliency, *Neurobiology of Attention, (L. Itti, G. Rees, J. Tsotsos, Eds.)*, pp. 576-582

[Jac91] JACOB, R. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What you Look at is What you Get. *ACM Transactions on Information Systems*, Vol. 9, No. 3, pp. 152-169.

[Jam90] JAMES, W. 1890. The Principles of Psychology (Volume 1). *Dover Publications*

[JWB01] JOHANSSON, R., WESTLING, G., BACKSTROM, A., & FLANAGAN, J. 2001. Eye-hand coordination in object manipulation. *The Journal of Neuroscience* Vol. 21, No. 17, pp. 6917 – 6932

[KDE00] KIM, H., DI GIACOMO, T., EGGES, A., LYARD, E., GARCHERY, S., & MAGNENAT-THALMANN, N. 2000. Believable Virtual Environment: Sensory and Perceptual Believability, *Believability in Virtual Environments, December Issue. Topic: Personality and Emotion Simulation*

[KDE91] KWAK, H. W., DAGENBACH, D. & EGETH, H. E. 1991. Further evidence for a time-independent shift of the focus of attention, *Perception and Psychophysics* Vol. 49, pp. 473-480

[Kel04] KELLY, J. N. 2004. Play Time: MMORPG Play as an Economic Activity. Online at: http://www.anthemion.org/Words/PlayTime0/playtime0.html (Accessed 20/09/06)

[KGH99] KIMMIG, H., GREENLEE, M., HUETHE, F., & MERGNER, T. 1999. *MR-Eyetracker:* A new method for eye movement recording in functional magnetic resonance imaging (fMRI). *Experimental Brain Research*, Vol. 126, pp. 443-449

[Kin60] KINGSLAKE, R. 1960. The development of the zoom lens. *Journal of the SMPTE* Vol. 69, p. 534

[Kle01] KLEINE-HORST, L. 2001. The bipolarity of figure quality. *Empiristic theory of visual gestalt perception. Hierarchy and interactions of visual functions (ETVG)*, Part 2, I.

[Leh03a] Figure taken from LEHAR, S. 2003. The World In Your Head. *Lawrence Erlbaum, Mahwah, New Jersey*. p. 52, Fig. 3.4. Reproduced with permission.

[Leh03b] Figure taken from LEHAR S. 2003. The World In Your Head.. *Lawrence Erlbaum, Mahwah, New Jersey*. p. 52, Fig. 3.3. Reproduced with permission.

[Lin06] LINDEN RESEARCH, INC. 2006. What is Second Life? Online at: http://secondlife.com/whatis/ (Accessed 12/09/06)

[LJS96] LI, Q., JINMEI, T., & SHIMA, K. 2006. IPv6 Core Protocols Implementation. Elsevier/Morgan Kaufmann Pulbishers.

[LL04] LUDWIG, I. & LACHNIT, H. 2004. Effects of practice and transfer in the detection of embedded figures, *Psychological Research,* Vol. 68, No. 4, pp. 277 – 288

[LLB06] LE MEUR, O., LE CALLET, P., BARBA, D., & THOREAU, D. 2006. A Coherent Computational Approach to Model Bottom-Up Visual Attention, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 28, No. 5, pp. 802-817

[LLC03] LEE, H., LEE, A., & CAMERON, D. 2003. Validation of a driving simulator by measuring the visual attention skill of older adult drivers, *The American journal of occupational therapy,* Vol. 57, No. 3, pp. 324-328

[LMD00] LEVIN, D.T., MOMEN, N., DRIVDAHL, S.B., & SIMONS, D.J. 2000. Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition* , Vol. 7 , pp. 397-412

[LMF89] LANIER, J., MINSKY, M., FISHER, S. & DRUIN, A. 1989. Virtual Environments And Interactivity: Windows To The Future. *1989 ACM Siggraph Panel Proceedings, ACM Press, New York.*

[MBD96] MORSE, K.L., BIC, L., & DILLENCOURT, M. 1996. Interest Management in Large-Scale Distributed Simulations. *In Preparation.*

[MCR90] MACKINLAY, J., CARD, S., & ROBERTSON, G. 1990. A semantic analysis of the design space of input devices. *Human Computer Interaction,* Vol. 5, pp. 145-190

[Min05] MINSON, R. 2005. Adaptive Interest Management for Distributed Virtual Environments. *Thesis Proposal, University of Birmingham*

[MLS05] MORGAN, G., LU, F. AND STOREY, K. 2005. Interest Management Middleware for Networked Games. *Proceedings of the I3D 2005 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, Washington, DC* pp. 57-63

[MMoG06] An Analysis of MMOG Subscription Growth. Online at http://www.mmogchart.com/. (Accessed 07/08/2006)

[MPZ94] MACEDONIA, M., PRATT, D., & ZYDA, M. 1994. A Network Architecture for Large Scale Virtual Environments, *Proceedings of the 19th Army Science Conference, Orlando, Florida*

[MRT00] MYSZKOWSKI, K., ROKITA, P., & TAWARA, T. 2000. Perception-Based Fast Rendering and Antialiasing of Walkthrough Sequences, *IEEE Transactions on Visualization and Computer Graphics*, Vol.6 No.4, pp.360-379

[MSS00] MOST, S., SIMONS, D., SCHOLL, B. & CHABRIS, C. 2000. Sustained Inattentional Blindness: The Role of Location in the Detection of Unexpected Dynamic Events, *PSYCHE*, Vol. 6(14)

[MT95] MILLER, D. & THORPE, J. 1995. SIMNET – The advent of simulator networking, *Proceedings of the IEEE* Vol. 83 No. 8 pp. 1114-1123

[MZ04] MASA, M. & ZARA, J. 2004. Interest Management for Collaborative Environments Through Dividing Their Shared State Cooperative Design, Visualization, and Engineering, Springer Berlin / Heidelberg, Vol. 3190/2004

[MZ79] MCCONKIE, G.W., & ZOLA, D. 1979. Is visual information integrated across successive fixations in reading? *Perception and Psychophysics*, Vol. 25(3), pp. 221–224.

[MZP94] MACEDONIA, M., ZYDA, M., PRATT, D., BARHAM, P., & ZESWITZ, S. 1994. NPSNET: A Network Software Architecture for Large Scale Virtual Environments, Presence, Vol. 3, No. 4.

[Nag91] NAGATA, S. 1991. How to reinforce perception of depth in single two-dimensional pictures. In *Pictorial Communication in Virtual and Real Environments*, S. R. Ellis, Ed. Taylor & Francis, Bristol, PA, pp. 527-545.

[Nav06] NAVE, R. The Human Eye. *Online at* http://hyperphysics.phy-astr.gsu.edu/hbase/vision/eye.html (Accessed 10/07/06)

[NHS95] NAKAYAMA, K., HE, Z., & SHIMOJO, S. 1995. Visual surface representation: a critical link between lower-level and higher level vision. *Invitation to Cognitive Science. M.I.T. Press*, p. 1-70

[OG02] OLIVEIRA, J., & GEORGANAS, N. 2002. VELVET: An Adaptive Hybrid Architecture for Very Large Virtual Environments, *IEEE International Conference of Communications, No. 1*, pp. 2491-2495

[Oli01] DEOLIVIERA, J. C. 2001. Issues in Large Scale Collaborative Virtual Environments, PhD Thesis, Ottawa-Carleton Institute of Electrical and Computer Engineering School of Information Technology and Engineering, University of Ottawa

[OPI05] MA, C. OSHERENKO, A. PRENDINGER, H. ISHIZUKA, M. 2005. A chat system based on emotion estimation from text and embodied conversational messengers Proceedings of the 2005 International Conference on Active Media Technology (AMT 2005). pp 546-548

[Pas98] PASHLER, H. E. 1998. The psychology of attention. *Cambridge, MA: MIT Press.*

[Par82] PARROTT AC. 1982. Critical flicker fusion thresholds and their relationship to other measures of alertness. Pharmacopsychiatria Vol. 15 No. 1 pp. 39–43.

[PB98] PAPPU, R & BEARDSLEY, P.A. 1998. A Qualitative Approach to Classifying Gaze Direction. *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan*

[Pet99] PETTIFER, S. 1999. An Operating Environment for Large Scale Virtual Reality *PhD Thesis, Dept. of Computer Science, University of Manchester*

[PG00] PURBRICK, J. & GREENHALGH, C. 2000. Extending Locales: Awareness Management in MASSIVE-3. *Proceedings of the IEEE Virtual Reality 2000 Conference* VR. IEEE Computer Society, Washington, DC, p. 287.

[PKK00] PARK, C., KO, H., & KIM, T. 2000. Multi-resolution spatial model for large-scale virtual environment. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (VRST '00). ACM Press, New York, NY, pp. 92-96.

[PPW97] PAUSCH, R., PROFFITT, D., & WILLIAMS, G. 1997. Quantifying immersion in virtual reality. *Proceedings of the 24th annual conference on Computer graphics and interactive techniques, ACM Press* pp. 13–18.

[Pro13] PROSZYNSKI, K. 1913. Lecture to the Royal Photographic Society – The Persistence of Vision. Cinema Notes, Amateur Photographer, 3rd March (Volume unknown)

[PS03] PETERS, C., AND SULLIVAN, C. O. 2003. Bottom-up visual attention for virtual human animation. In Computer Animation for Social Agents.

[PSB90] PERROTT, D. R., SABERI, K., BROWN, K., & STRYBEL, T. Z. 1990. Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics*, Vol. 48, pp. 214-226.

[PSD80] POSNER, M. I, SNYDER, C. R. R. & DAVIDSON, B. J. 1980. Attention and the detection of signals. *Journal of Experimental Psychology* Vol. 109 pp. 160-174.

[PSM05] POGGEL, D. A., STRASBURGER, H., & MACKEBEN, M.. 2005. Relative motion in the periphery of the visual field is a powerful cue for visuo-spatial attention. *Journal of Vision,* Vol. 5 No. 8, p. 162a

[RB83] REGAN, D. & BEVERLEY, K. I.. 1983. Visual Field Described by Contrast Sensitivity, by Acuity and by Relative Sensitivity to Different Orientations, *Investigative Ophthalmology and Visual Science,* Vol. 24, pp. 754–759

[RCD97] ROBERTSON, G., CZERWINSKI, M., & VAN DANTZICH, M. 1997. Immersion in desktop virtual reality. *Proceedings of the 10th annual ACM symposium on User interface software and technology, ACM Press.* pp. 11–19

[Red97] REDDY, M. 1997. Perceptually Modulated Level of Detail for Virtual Environments, *PhD thesis, Department of Computer Science, University of Edinburgh, UK*

[Red97b] REDDY, M. 1997. The Effects of Low Frame Rate on a Measure for User Performance in Virtual Environments. *Technical Report ECS-CSG-36-97, Department of Computer Science, University of Edinburgh*

[Ren00] RENSINK, R.A. 2000 The dynamic representation of scenes. Visual Cognition Vol. 7, pp. 17-42.

[Ren02] RENSINK, R.A. 2002. Internal vs. external information in visual perception. Proceedings of the 2nd ACM International Symposium on Smart Graphics pp. 63-70.

[RP03] REITSMA, P. S. AND POLLARD, N. S. 2003. Perceptual metrics for character animation: sensitivity to errors in ballistic motion. *ACM Trans. Graph.* Vol. 22, No. 3, pp. 537-542.

[RRC97] RENSINK, R.A., O'REGAN, J.K., & CLARK, J.J. 1997. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science,* Vol. 8, pp. 368–373

[SC94] SINGHAL, S., & CHERITON, D. 1994. Using a Position History-Based Protocol for Distributed Object Visualization, *Stanford University Computer Science Technical Report STAN-CS-TR-94-1505*

[SDL05] SUNDSTEDT, V., DEBATTISTA, K., LONGHURST, P., CHALMERS, A., & TROSCIANKO, T. 2005. Visual attention for efficient high-fidelity graphics, *Proceedings of the 21st spring conference on Computer graphics, Budmerice, Slovakia*

[SGB95] SNOWDON, D., GREENHALGH, C., & BENFORD, S. 1995. What you see is not what I see: subjectivity in virtual environments *Proceedings of FIVE'95: Framework for Immersive Virtual Environments (London)*

[Sim96] SIMONS, D.J. 1996. In sight, out of mind: When object representations fail. *Psychological Science,* Vol. 7(5), pp. 301–305.

[SLG06] SINGH, K. LOGAN, N., & GILMARTIN, B. 2006. Three-Dimensional Modeling of the Human Eye Based on Magnetic Resonance Imaging. *Investigative Ophthalmology and Visual Science.* Vol. 47, pp. 2272-2279.

[SKH02] SMED, J., KAUKORANTA, T., & HAKONEN, H. 2002. A Review on Networking and Multiplayer Computer Games, *Technical Report, Turku Centre for Computer Science*

[SLW05] SILVANTO, J., LAVIE N. & WALSH, V. 2005. Double Dissociation of V1 and V5/MT activity in Visual Awareness *Cerebral Cortex,* Vol. 15, pp. 1736-1741

[SP00] SEARS, C. & PYLYSHYN, Z. 2000. Multiple Object Tracking and Attentional Processing. *Canadian Journal of Experimental Psychology*, Vol. 54, No. 1, pp. 1 – 14

[SR05] SIMONS, D & RENSINK, R. 2005. Change blindness: past, present, and future. *Trends in Cognitive Sciences* Vol. 9(1) pp. 16-20

[SRS95] SMITH, J., RUSSO, K. & SCHUETTE, L. 1995. Prototype multicast IP implementation in ModSAF, *Proceedings of the Twelfth Workshop on Standards for the Interoperability of Distributed Simulation* pp. 175--178.

[ST91] SUTTER, E. E. & TRAN, D. 1991. The Field Topography of ERG Components in Man—I. The Photopic Luminance Response, *Vision Research*, Vol. 32 No. 3: pp. 433–446

[Ste55] STEVENS, S.S. 1955. The measurement of loudness. *Journal of the Acoustical Society of America*, Vol. 27 pp. 815-829

[Str35] STROOP, J. R. Studies of interference in serial verbal reactions. 1935. *Journal of Experimental Psychology*, Vol. 12, pp. 643-662

[Str35b] Illustration of the Stroop Effect. Online at:
http://www.apa.org/science/stroop.html (Accessed 04/09/06)

[SY02] SHOMSTEIN, S., & YANTIS, S. 2002. Object-based attention: Sensory modulation or priority setting? *Perception & Psychophysics*, Vol. 64, pp. 41-51.

[TG80] TREISMAN, A.. & GELADE, G. 1980. A feature-integration theory of attention, Cognitive Psychology, Vol. 12, No. 1, pp. 97-136

[The93] THEEUWES, J. 1993. Visual selective attention: A theoretical analysis, *Acta Psychologica*, Vol. 83, pp. 93-154

[Tre88] TREISMAN, A.. 1988. Features and objects: the fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, 40A, pp. 201-236

[TSL03] TSE, P. U., SHEINBERG, D. L., & LOGOTHETIS, N. K. 2003. Attentional enhancement opposite a peripheral flash revealed by change blindness. *Psychological Science*, Vol. 14(2), pp. 1–8.

[VBR98] VALDES-SOSA, M., BOBES, M., RODRIGUEZ, V., & PINILA, T. 1998. Switching Attention without Shifting the Spotlight: Object-Based Attentional Modulation of Brain Potentials. *Journal of Cognitive Neuroscience.* Vol. 10, No. 1, pp. 137 - 151

[VRL82] VIRSU, V., ROVAMO, J., LAURINEN, P. AND NÄSÄNEN, R. 1982. Temporal Contrast Sensitivity and Cortical Magnification, *Vision Research*, Vol. 33: pp. 1211–1217

[WCC06] WOOD, S, COX, R & CHENG, P. 2006. Attention Design: Eight issues to consider. *Computers in Human Behavior. Special issue: Attention-aware systems*, Vol. 22, No. 4, pp. 588-602

[Web34] WEBER, E. H. 1834. De pulsu, resorptione, audita et tactu. *Annotationes anatomicae et physiologicae.* (English translation. by H.E. Ross, Academic Press, New York)

[Wed94] WEDELL, D. H. 1994. Contextual contrast in evaluative judgments: A test of pre- versus post-integration models of contrast. *Journal of Personality and Social Psychology*, Vol. 66, pp. 1007-1019

[Wer12] WERTHEIMER. M. 1912. Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie.* pp. 161-265,.. (English translation by T. Shipley. 1961. Classics in Psychology. *New York: Philosophical Library*)

[Wer20] WERTHEIMER. M. 1920. Über Schlußprozesse im produktiven Denken. *Berlin: Weltkreisverlag.* (English translation by W.D. Ellis. 1980. A Source Book of Gestalt Psychology. pp. 274-282)

[Wer22] WERTHEIMER. M. 1922. Untersuchungen zur Lehre von der Gestalt. I. Prinzipielle Bemerkungen. Psychologische Forschung, 1, 47-58. (English translation by W.D. Ellis. *A Source Book of Gestalt Psychology.* 1980, pp. 71-88)

[Wer22b] WERTHEIMER. M. 1922. Untersuchungen zur Lehre von der Gestalt. I. Prinzipielle Bemerkungen. Psychologische Forschung, 1, 47-58. (English translation by W.D. Ellis. *A Source Book of Gestalt Psychology.* 1980, p. 88)

[WF06] Unknown Author. Sample image of a wireframe rendering of a cube, icosahedron, and sphere. Distributed under the GNU Free Documentation License. Online at www.gnu.org (Accessed: 17/12/06).

[WZ98] WATSEN, K., & ZYDA, M. 1998. Bamboo - A Portable System for Dynamically Extensible, Real-time, Networked, Virtual Environments. *IEEE Virtual Reality Annual International Symposium (VRAIS'98), Atlanta, Georgia*

[WZ98b] WATSEN, K., & ZYDA, M. 1998. Bamboo - Supporting Dynamic Protocols For Virtual Environments, *IMAGE 98 Conference, Scottsdale, Arizona*

[WFK06] WALTHER, D., FEI-FEI, L., & KOCH, C. 2006. Measuring the cost of deploying top-down visual attention. *Journal of Vision*, Vol. 6(6), p. 310

[WHK04] WOLFE, J. M., HOROWITZ, T., KENNER, N., HYLE, M. & VASAN, N. 2004. How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, Vol. 44, pp. 1411–1426

[WUS05] University of Washington, St. Louis, *Human Vision*. Online at http://thalamus.wustl.edu/course/basvis.html (Accessed 15/8/06)

[Yar67] YARBUS, D. 1967. Eye motion and vision, *Plenum Press, New York*

[You75] YOUNG. L. R. & SHEENA, D. 1975. Survey of eye movement recording methods. *Behavior research methods and instrumentation*, Vol. 7, No. 5, pp. 397-429.

[YPG01] YEE, H., PATTANAIK, S. AND GREENBERG, D. P. 2001. Spatiotemporal Sensitivity and Visual Attention for Efficient Rendering of Dynamic Environments, ACM Transactions on Graphics, 20, 1, 39-65.

[Zel92] ZELTZER, D. 1992. Autonomy, interaction, and presence. *Presence: Teleoperators and Virtual Environments*, Vol. 1 pp. 127–132

[ZS00] ZYDA, M. AND SINGHAL, S. 2000. Networked Virtual Environments: Design and Implementation *ACM Press, New York*, p. 11

[ZYM02] ZHOU, Q., YU, H., MAKRAKIS, D., GEORGANAS, N., & PETRIU, E. 2002. Quality of Service Support of Distributed Interactive Virtual Environment

Applications in IP-Diffserv Networks, *Proc. HAVE'2002 - IEEE Workshop on Haptic Virtual Environments and their Applications, Ottawa, ON, Canada*, pp. 97-102

APPENDIX I

HANDOUT ISSUED TO SUBJECTS PRIOR TO EXPERIMENT SERIES

Internet-based Virtual Environments

Overview

The environment you will be working in will be shared with the other people within Java lab, and, to better simulate a large-scale environment, 100 more clients with simulated behaviour will be added. This will give some simulation of overcrowding. However, these clients are not considered participants in the tasks.

The session should proceed as follows:

14:00 – 14:10 Introduction

Before starting, it's important you familiarise yourself with the controls and layout of the environment. Refer to the remainder of this handout and feel free to experiment with the controls. A short presentation will aim to explain what you'll be attempting to evaluate during the experiments.

14:10 – 14:30 Grouping Task

In the first session of experiments, you'll be asked to search out other participants within the environment, and form a single, large group. The task will be repeated **six** times, with a 4 minute time limit each time. After each experiment, you'll be asked to make an evaluation of the motion of other users within the environment.

14:35 – 14:55 Tag Experiment

In the second session, you'll be playing a game of tag with the other participants. Again, **six** games will be played within a 4-minute time limit, and you'll be asked to evaluate the motion of the other users.

1. Getting Started

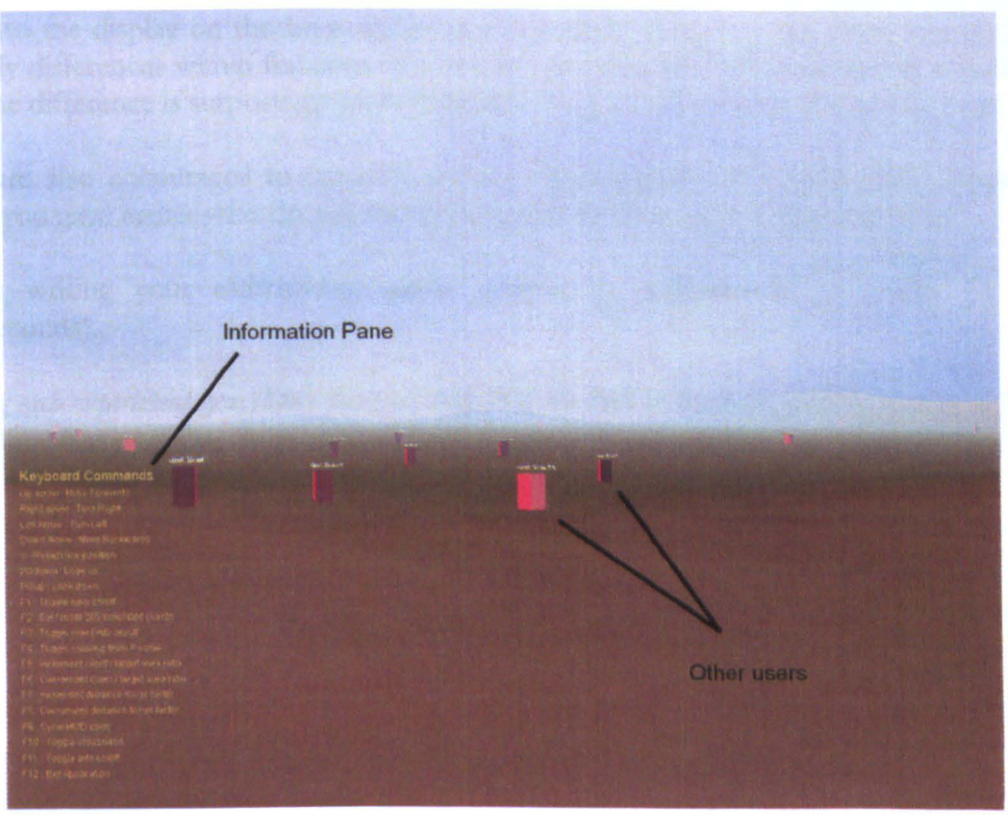
Hopefully, your PC will already have the VE running. In this case, you can proceed to section 2 ***If not***, you can launch a new client at a command prompt from within the virtual environment's directory (if you can't find it, ask) with:

```
java Launcher
```

This will subsequently prompt you for a username and an ID number (please carefully enter the one which should be written at the top of the attached sheet) and then select to run in full screen mode. If you have any problems getting things running, please let someone know.

2. The User Interface

The interface is a very common one to virtual environments, and you may already be familiar with it. By moving the mouse left and right you can turn your point of view from side to side (if this doesn't work, try left clicking once to give the window focus) or you can alternatively use the left and right arrow keys. To move forwards or backwards use the up and down arrow keys respectively. The other keyboard commands are summarised on-screen with an online help overlay, which can be toggled on and off with F1 (shown in Figure 1). If you have trouble reading the overlays, you can adjust their colour to increase the contrast by pressing F9. ***Other users can be identified by the name above their heads; in the case of simulated clients this starts with 'simClient'.***



3. Evaluating Client Motion

During the session, you'll be participating in two series of tasks. After each experiment, you'll be asked to make an estimation of what you feel was the average delay between updates of the positions of the other users within the environment.

This means trying to take into account how smoothly other users seemed to be moving. Refer to the large screen display for an ongoing illustration of the effects of various delays, and how movement seems as delay increases. Whilst a delay of 2 seconds or more results in noticeable behaviour, with users vanishing and reappearing as they move, lower values result in gradually smoother motion. Visual perception is highly sensitive to movement, and you should find that you are able identify differences between movement with even very low time intervals.

Following each experiment, you'll be asked to try to estimate the average delay between client movements. Whilst your priority should be to complete the tasks given within the time limit (i.e. grouping or tag), please try to ensure you complete the questionnaire following each task.

Important Notes:

Whilst you should try to guess the delay as accurately as possible, it is, naturally, extremely hard to guess the correct value to within fractions of a second. *However, please attempt to answer for each experiment, even if you feel your answer may be substantially inaccurate.*

Refer to the display on the large screen as a reference. Whilst it may seem impossible to identify differences within fractions of a second, you may find when observing this display that the difference is surprisingly more noticeable than $1/10^{th}$ of a second would seem.

You are also encouraged to consider previous experiments, and relate your answers to those you gave earlier – i.e. do you feel clients were moving more or less smoothly?

When writing your estimations, please answer in milliseconds (1 second = 1000 milliseconds).

During each experiment, you should focus on completing the task at hand (i.e. forming a group or playing tag). Please leave completing the questionnaire until the end of the experiment.

3. Experiment 1 - Grouping

The first experiment challenges you to group up with the other real-world users as quickly as possible, trying to co-ordinate yourselves to form up in a single area. To do this, you'll need to identify other users by name and move in a group to find other clients. To keep the experiment fair, no verbal communication about your whereabouts is permitted. The experiment will start once all users are ready and familiar with the user interface, and your position will be randomly reset automatically as it begins. You then have 4 minutes to form a group with other users. This will be repeated six times.

Following each experiment refer to the demonstration on the overhead display and attempt to approximate what you believe was the average latency in the environment during the test. Your answer should be filled in on the included questionnaire form.

4. Experiment 2 – Playing Tag

The next experiment encourages you to avoid, rather than group with other users. When it begins, players will be randomly positioned again, and then one player will be randomly selected as being 'tagged'. They can then move 20% faster, and have to chase down other real world clients within the environment. The simulated clients remain present but outside of the game, so they can't be tagged.

You can identify players who have been tagged by the colour of their avatar (it changes from blue to white). Once tagged, you can also move faster and have to chase down other clients. You tag other players simply by moving near to them. The winner is the last player left untagged.

Once again, following each experiment refer to the demonstration on latency on the overhead display and attempt to approximate what you believe was the average latency in the environment during the test.

5. Questionnaire

Before leaving, please ensure all information on the questionnaire is completed, including the brief section at the end regarding how easy you found the experiment. If you have any questions, please feel free to ask.

APPENDIX II

QUESTIONNAIRE COMPLETED BY SUBJECTS DURING EXPERIMENT SERIES

Questionnaire

Username	
User ID#	

1. Grouping Exercise

Referring to the display on the large screen at the front of the room, what do you feel was the approximate average update delay of clients within the simulation? Please give your answers in milliseconds.

Experiment 1:

Experiment 2:

Experiment 3:

Experiment 4:

Experiment 5:

Experiment 6:

2. Tag Game

Referring to the display on the large screen at the front of the room, what do you feel was the approximate average update delay of clients within the simulation? Please give your answers in milliseconds.

Experiment 1: [ ]

Experiment 2: [ ]

Experiment 3: [ ]

Experiment 4: [ ]

Experiment 5: [ ]

Experiment 6: [ ]

3. General Questions and Comments

a. Do you think there was a noticeable difference in the update delay between the various experiments?

[] Yes, significant difference between all experiments
[] Difference between some experiments
[] No difference between experiments
[] Unsure

b. Have you ever used a virtual environment or played a virtual-environment based game with similar controls before?

[] Yes
[] No

c. Do you feel any of the following may have impacted your behaviour or ability to participate during the tests (check all that apply):

[] Low framerate
[] Unfamiliarity with the controls / trouble moving
[] Crashes / bugs
[] Colour blindness
[] Other

If other please specify:

____

____

d. Finally, do you have any suggestions for improvements to any aspect of the environment or the experiments? Continue overleaf if necessary.

____

APPENDIX III

NOTES ON PSUEDOCODE FORMATTING

Psuedocode Structure

Where possible, pseudocode is adapted directly from the source code within the system, and adjusted for clarity with regards to the concepts encapsulated within the code. An object-oriented structure is used, wherein methods are defined as:

```
method <methodname>(Parameters)
{
< method content >
}
```

With brackets used to encapsulate method content. Calls to methods are expressed as:

```
<methodname>(Parameters)
```

Where parameters are any number of values matching the format of the method declaration.

Methods may not be explicitly defined prior to being called – where this is the case the method name or included description should clearly detail the function of the method.

Variables

Where possible, the data type is defined clearly and the variable declared prior to use, using the format

```
define <variable name> as an <variable type>
```

In the case of arrays, the dimensions and elements are defined and referenced by the use of square brackets, e.g.

```
define myArray as an array[][]
```

Implies future references to the two-dimensional myArray should suggest an element, e.g. myArray[30][2]. Note that variable definitions are only used where necessary for clarity – in cases where simple variables are used to store interim results the declaration may be omitted.

Logic

Logic statements again mirror source code, with if…else statements retaining a basic format of:

```
If <condition> then
{
<statements to execute if condition is true>
}
else
{
<statements to execute if condition is false>
}
```

Conditions are formatted using standard logical operators, with == being used as "is equal to", != as "is not equal to", ">" as greater than, and so forth. In the case of algebra standard mathematical notations are adopted, with "/" signifying "divided by", and "*" signifying "multiplied by".

Iteration

To aid clarity, only two condition-based loops common to programming are used within the psuedocode: FOR… NEXT, and DO…WHILE

They adopt a syntax common to most programming languages, in the more common case of the FOR loop this follows the format:

```
for ( <initial condition>, <break condition>, <increment> )

{

< code to be iterated >

}
```

To induce an immediate exit from a loop, the keyword "break" is used.

APPENDIX IV

PERCEPTUALLY-ORIENTED INTEREST MANAGEMENT IN LARGE-SCALE NETWORKED VIRTUAL ENVIRONMENTS

AS PUBLISHED IN EG UK THEORY AND PRACTICE OF COMPUTER GRAPHICS, 2005.

EG UK Theory and Practice of Computer Graphics (2005)
L. Lever, M. McDerby (Editors)

Perceptually-Oriented Interest Management In Large-Scale Networked Virtual Environments

I. Dunwell & Dr. J. C. Whelan

Center for Internet Computing, The University of Hull, UK

Abstract

Amongst the most significant challenges in developing large-scale multi-user virtual environments is the efficient filtering of data to each user - a process commonly described as "interest management". This work-in-progress paper presents a broad summary of existing approaches, placing an emphasis upon the relationship between interest management and human perception. Subsequently, an introduction to the challenges in evaluating the success of interest management, given such a relationship to perception, is presented. The initial development of a test environment aimed at overcoming some of the challenges in providing a platform for such evaluation is then described, together with discussion of a perceptually-oriented approach to interest management which relies on the description of perception as a dynamic field, formed by analysis of the user's focus.

1. Introduction

As virtual environments become more sophisticated in terms of both content and interactivity, the potential for a single, extensible virtual environment (VE) capable of supporting massive numbers of users has frequently captured both the attention of researchers and the imagination of the general public alike. The implementation and advancement of such massively-multi-user networked virtual environments (MMVEs) presents many challenges in a wide range of research areas.

Amongst these challenges, one of the most significant and complex is the filtering of network traffic to ensure an optimal use of available bandwidth. This filtering, or "interest management" (IM), involves ensuring only content relevant to each user is transmitted via the network. Current approaches aiming to provide IM are wide-ranging, and typically integrated into existing simulations. Such integration often comes at the cost of a certain degree of re-usability and makes evaluation of the generic potential of an approach difficult. Following a limited summary of existing approaches and their history, the challenges in such evaluation are discussed in Section 3. This discussion is followed by an outline of a test system intended to provide the grounds for such an evaluation, and, given such facilities, the concept of providing a perceptually driven approach centered around a user's focus is presented.

2. Background

Some of the earliest taxonomies of IM approaches [Mor96] identified the concept of intrinsic filtering - filtering based upon that attributes of entities within an environment. Further classification of these techniques leads to a subsequent distinction emerging between grid and aura-based approaches. Grid techniques, pioneered largely by SIMNET [MT95], revolve around partitioning virtual environments into regions of interest (commonly referred to as locales [BWA96]), with clients being considered relevant to all other clients within the same locale. By comparison, aura-based approaches such as those used in early systems such as MASSIVE [Gre98], define a region centered upon each client. If the interest regions of two clients intersect then they are considered to be mutually visible.

These two techniques have been refined and combined over recent years, in response to the rapidly emerging need for efficient IM within large-scale virtual environments. Commercial endeavors (primarily within the gaming industry) have brought large-scale multi-user virtual environments to the general public, creating online economies with substantial turnovers, recently described as being in excess of the economies of a significant number of real-world countries [Cas01]. As hardware continues to advance, significant potential for further applications remains. Hence the commercial need for efficient IM is becoming increasingly apparent.

One of the most obvious, and successful, IM techniques is the combination of multiple layers of interest management within a single system. Zyda et al. [AWZ98] demonstrate a three-tiered approach within the Bamboo toolkit for distributed environments, illustrating the effectiveness of combined IM paradigms. The combination of grid and aura-based approaches is commonly used - at their most fundamental level, grid and aura-based techniques both provide a simplistic and efficient means for filtering data within a large-scale VE. More recent refinements have sought to add additional aspects to the filtering process, such as considering visibility more closely [HPG02]. Whereas more basic aura and grid based systems filter spatially, irrespective of surrounding geometry, further culling based on visibility presents a clear example of the potential for refinements based on a better understanding of a users perceptual needs. The system is successful, noting a significant performance gain with no perceived loss in quality. Hence the potential for such visibility-oriented interest management in a large-scale virtual environment centered around a replicated-database model is clear. However, the potential for perceptual consideration is broader than visibility-based culling would suggest, since an interest manager for a truly extensible environment must take into account other aspects, such as extensibility, sound, and interactivity.

Thus whilst visibility alone is unquestionably a valuable criterion by which to perform filtering, the notion of perceptual relevance offers far more significant scope for refinement. Much as Reddy [Red01] applies perceptual techniques to rendering, with a significant performance gain, so may perception be considered in IM. Beharee et al. successfully demonstrate the potential for the use of perceptually oriented IM in a variety of simulations, ultimately concluding with a simulation of an 800-client virtual city [BWH]. The perceptual culling is based around the notion of change blindness; in effect aiming to filter out items of data in which the user is unlikely to identify a change. Change blindness itself is a significant field of research, and a large volume of work offers a variety of further possibilities for enhancement [Tse04].

Beharee et al. focus heavily on obtaining results with no noticeable decrease in quality of simulation. In a massive-scale virtual environment, situations in which available bandwidth is insufficient to offer such a result can prove commonplace, and hence analysis of which perceptual qualities offer the greatest scope for 'graceful degradation' is also of considerable importance. With regards to extensibility, as environments become more sophisticated - particularly in terms of object interactivity - the replicated data model common to the majority of existing systems (whereby simulation content is installed or downloaded to each client prior to run-time), prevents expansive and highly interactive environments. Subsequently an increasing level of importance can be placed upon creating IM capable of filtering textures, geometry and other world content to each user. Furthermore, the nature of input devices to virtual environments allows for a unique opportunity to analyse the behaviour of each client, and tailor the filtering process in response. Such behavioural analysis can range from following user behaviour and mouse input to more sophisticated data capture techniques such as eye-tracking [Jac95]. Since this effectively permits a users behaviour to be analysed at run-time, it becomes possible to consider filtering approaches which are capable of tailoring themselves more closely to behavioural input.

Whilst each of these aspects presents individual potential for refinement of IM, a recurring theme when examining existing approaches is a significant difficulty in comparison and evaluation, due to both the qualitative nature of perceptually oriented results and the application dependence of many IM approaches. The subsequent section aims to define this challenge in more detail, and goes on to describe a test system developed to address it.

3. Evaluating Interest Management

3.1. Generic IM

Singhal (a pioneer of the aforementioned Bamboo toolkit) stated confidently as early as 1998 that "Eventually there will be a persistent VE shared simultaneously by billions of participants" [ZS00]. This vision remains a long-term goal of MMVE developers, due in large part to the commercial potential for having such an immense audience co-existing within a single environment. Providing such a system implies two main factors need to be addressed; firstly the technical barriers in place must be overcome (such as bandwidth and rendering limitations), and secondly, on a more conceptual level, the environment must be sufficiently ubiquitous and extensible to fulfill the needs of all users. Given such needs may range from simple social interaction, through simulation and collaborative working, such fulfillment is no simple task.

Clearly, the role of interest management will be pivotal in providing such an environment; without the means to efficiently and rapidly control communication between clients such an application would prove impossible. Primarily, extensibility in such an environment implies a degree of dynamism within an interest management routine, due to the differing needs of users and expanding environment. However, the wealth of existing environments with broad ranging applications and near-identical interest management approaches strongly suggests that despite a need to interest management to suit a specific environment, generic principles do exist. The approach to interest management within the e-Agora system [Mr02] shows one such approach to generic interest management, in this case attempting to utilise 'general variables' to store different types of data.

Whilst there are many different aspects that may be considered in establishing a generic 'theme' of virtual environment, some of the most fundamental include:

1) The concept of each client as a virtual embodiment of a human. This brings with it a variety of notions, such as the centering of the region of interest around the client, and a relationship between the scene rendered to the client and the information required.

2) The finite amount of bandwidth available, implying a need for graceful degradation and, more fundamentally, scalability, in areas of high traffic (such as a crowded virtual room).

3) The implicit relationship between the software and hardware underlying the simulation and the simulation itself. Multiple levels of interest, for example, offer diminishing returns with increasing numbers of levels due to the fact they require formation of additional multicast groups [PB95].

4) A basis for the simulation in the real world. Whilst virtual environments exist which do not aim to provide such simulation, for the purposes of this context we choose to focus on those aiming to provide a 'realistic' simulation on a human avatar within a virtual world. This has broad implications regarding the relationship between the positions and properties of objects relative to the client and their perceptual relevance.

We may build upon these concepts to create interest management paradigms that, whilst not truly application-independent, lend themselves to a significantly broad range of applications. In particular we choose to focus on environments where the client is considered to be a virtual human, and thus their relationship with the VE considered parallel to the relationship of a human within the real world. As an important side-note, whilst the employment of any interest management technique has obvious overheads in terms of processing, as hardware advances and the depth of interactivity within virtual environments increases, bandwidth can be seen to be fast emerging as the true bottleneck for such environments. Hence despite a clear need for processor efficiency in any interest management approach, its application may not necessarily imply any tangible negative effects upon the simulation. It may, however, prove beneficial to accept some loss of simulation quality if a corresponding gain in bandwidth reduction is possible. Balancing - or providing some degree of scalability towards the two distinct criteria of bandwidth reduction and impact upon quality of simulation is hence important to any IM approach. Further defining these criteria is the focus of the next section.

3.2. Bandwidth vs. Quality of Simulation

Of crucial importance to the development of any interest management approach for such a general application is the provision of means to test and evaluate differing approaches against one another. Due to the breadth of existing large-scale VE applications, and the scope for future development, evaluating the success of differing approaches in a generic fashion proves a complex task.

One clear criteria for evaluation is the decrease in bandwidth consumption achieved. However, this must be weighed against the impact upon the quality of simulation. This is a broad-ranging concept, which must take into account a large number of qualitative variables in order to attempt to establish how a simulation 'feels'. Obviously, some of the common effects of interest management, such as entities 'popping up' near the user have a strong negative impact on this quality. Similarly, the effects on frame rate of a highly CPU-intensive interest manager may be factored into this broad notion. Thus we can loosely define the success of any interest management approach in a generic context to be the trade off between the reduction in bandwidth consumption compared to the perceived decrease in quality of simulation.

In an attempt to focus on such interest management issues generic to large scale VEs, and to better measure this balance between bandwidth and quality of simulation, a system is currently under development intended to simulate limited aspects of the data model of a peer-to-peer, multicast-based, massive-scale virtual environment. This simulation also includes behavioral simulations of large numbers of clients. This system is intended to allow for the evaluation of these two main concepts of quality of simulation and bandwidth consumption via the following approaches:

1) Qualitative evaluation of the perceptual effects and hence quality of simulation by means of a rendering engine. Users can be placed within the virtual environment and navigate through it freely, observing the surrounding scenery and clients. It is hoped by analyzing the response of a suitably large sample of users, some conclusions can be reached regarding the success of differing techniques in this respect. Of particular emphasis is the success of techniques as bandwidth constraints force quality of simulation to deteriorate.

2) Quantitative prediction of the bandwidth consumption for a variety of client behavioral models. This output may be compared alongside the qualities of simulation in order to better appreciate the effects of differing interest management approaches in as quantitative a method as possible.

4. Developing a Test Simulation

4.1. Overview

The test system developed (via a combination of Java and Java3D) bases itself around the notion of a massive scale, peer-to-peer, virtual environment, such as that of DIVE [FS98]. The system is highly modular, with the emphasis being placed on allowing components to be interchanged and modified as easily as possible.

Each client behaves as an autonomous entity running in its own thread, and, as a consequence of the modular nature, may have its motion governed either by user input or an AI routine. Current AI routines are simplistic, ranging from purely random motion to a 'point of interest' approach whereby points may be inserted into the virtual environment which attract clients, hence allowing for a more realistic simulation of client distribution within a virtual space (since clients would seldom distribute as evenly as a random motion algorithm would suggest).

Output on total traffic between clients is streamed to files, in order to assess client load distributions and bandwidth consumption. The environment is also rendered in real-time 3D (as shown in Figure 1) for purposes of evaluating quality of simulation. The rendering engine is currently primitive; capable only of showing both overviews of the system and a view from the perspective of a single client. Future work aims to extend this into providing a user-controlled client within a more realistic environment to better assess the quality of the simulation for a given interest management approach.

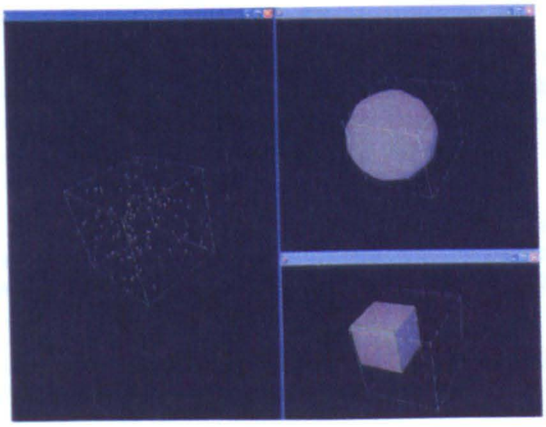


Figure 1: *Universe overview (left) alongside visualizations of both aura (top right) and grid based (bottom right) interest management implementations, focused on a single client*

Quantitative evaluation of bandwidth consumption is performed by keeping track of the number of total update packets sent through the simulation as it runs, which would in a real-world application correspond directly to packets sent via the network (each containing an update on nearby client attributes). By directly keeping track of the total number of packets sent over time by means of a data log, the effects of events within the simulation can clearly be observed.

4.2. Comparison of Existing Approaches

Figure 2 shows output for two basic grid and aura-based interest management implementations, alongside output with no interest management. In all cases clients were initialized at the center of the virtual universe and allowed to randomly distribute outwards, hence providing an illustration of the effects of a shift from an area of high client density to a lower one.

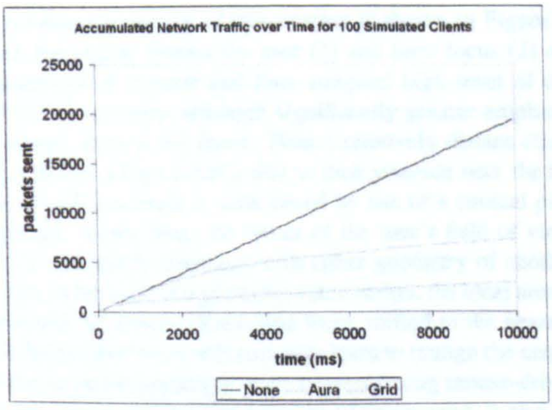


Figure 2: *Packets sent / time for 3 interest management simulations*

As can be observed, initially with all clients at the center of the universe and hence within each other's respective areas of interest, bandwidth use increases linearly with time in all 3 simulations. As clients distribute, the simulations running interest management deviate from this linear relationship, using less bandwidth as distant clients cease requesting updates. Whilst the illustration shown is for the case of 100 clients, identical behavior can be observed in simulations containing both 10 and upwards of 1000 clients. Whilst in this case the results may be similarly proven mathematically, the system is ultimately intended to allow for the interaction of human users, where such proof would be impossible.

Using such output, it becomes possible to compare bandwidth usage characteristics of differing interest management techniques. However, there is currently an extremely limited facility for evaluating the quality of simulation, and hence it proves difficult to draw any conclusions from bandwidth analysis alone. Establishing such facilities is a primary development goal, and hence as mentioned in the previous section bandwidth will ultimately be examined directly alongside quality of simulation. Future work intends to allow for contrasting techniques to be evaluated by studying the rendered output of the system and noting the change in quality of simulation whilst total bandwidth use remains constant. This differs from many approaches, which seek to provide evidence for bandwidth reduction while maintaining a constant quality of simulation. In practical applications the available bandwidth is often the constant factor, and thus placing emphasis on providing scalability such that quality of simulation may be optimised for any given bandwidth should prove advantageous.

In order to further develop techniques a consideration needs to be made of the nature of quality of simulation (as discussed in section 3), and ways in which its attributes may be exploited to provide such improvement. A fundamental aspect of quality of simulation of particular interest is its close relationship to the perception of the user. If a user simply 'feels' a simulation is more accurate, then an improvement can be assumed to have been achieved. Existing interest management techniques often provide some appreciation of perception as a simple spatial attribute - if a object is within a certain distance of the user, then it is perceptually important. More refined approaches, such as that of Beharee et al., take fuller advantage of the wealth of research regarding perception, and consider its implication to interest management. The next section introduces an alternative approach to the implementation of such perceptual considerations.

5. A Perceptually-Oriented Interest Management Approach

Currently, a method is being developed which bases itself on the aura concept of interest management. However, the shape and description of this aura is modified to accommodate a series of perceptual concepts:

1) Spatial proximity alone is largely irrelevant as a measure of relevance in the situation of a client observing a distant object. Many IM approaches offer poor performance with regards to distant objects, as they focus primarily on the space around the user.

2) It is possible to provide a good approximation of a users interests by studying their interaction within the environment - specifically, how they orient their viewpoint using mouse or head-mounted input devices.

3) Such interests may shift rapidly, particularly in the case of military simulations or games.

Existing perceptually-oriented IM approaches base themselves largely around the model of attention as a multimodal process [PJR01]. In this scenario, objects are filtered based on intrinsic low-level properties such as their colour, or higher-level simulation dependant attributes. However, a wealth of alternative models for attention exist which offer scope for continued research. In particular, the spotlight model of attention [Tre86] may be seen to offer the potential to provide additional filtering based on user viewpoint orientation. It is this model which forms the basis of the technique currently being developed to work alongside existing methods and offer further refinement.

In order to implement such a model, a system is developed wherein each client is assigned a dynamic aura, a concept not new in itself (Velvet [OG02], for example, implements such a system successfully). However, the shape of the aura is derived by attempting to determine the visual focus of the client, based upon their orientation and the surrounding geometry. This method provides a unique emphasis on the focus of the user, rather than their entire field of vision or immediate aura, and is of particular relevance to large-scale military simulations wherein the target of the user is frequently of more importance than other nearby entities.

A basic illustration of this concept is shown in Figure 3. Both the region around the user (1) and their focus (2) are considered of interest and thus assigned high level of detail (LOD) priority, although significantly greater emphasis is placed around the focus. Thus a relatively distant client (3) receives a high priority due to their position near the focus. The focus itself is determined by use of a conical pick segment, drawn from the center of the user's field of view, which ultimately intersects with either geometry of another client. In the case of a geometry intersection, the local area is analysed for nearby clients and focus shifted to the nearest. It is hoped that users will naturally learn to realign the center of the screen according to their interest (using mouse-driven input), hence making this measure of focus valid. It also integrates well with military simulations and first-person style games, which tend to place the user's weapon crosshair central on the screen.

Subsequent to the establishment of focus, a field of values representing perceptual relevance (termed P-values) is established, based upon field equations arising from the concepts discussed above and generally obeying computationally-simple $1/r^2$ behaviour. The system thus can be described as taking advantage of 'inattentional blindness' [MR98], as objects in the middle-ground generate lower P-values to those near the user or point of focus. These values may be subsequently quantised to provide support for multiple levels of detail (as illustrated in Figure 3), and thus translated into multicast group subscriptions. The generic nature of the P-values allows a separation between providing a measure of perceptual relevance and performing the information culling central to interest management, which is useful for adding additional weighting or filtering prior to data transmission. Current work is considering more closely other perceptually important aspects of behaviour, such as sudden movement, in an attempt to predict focus shifts prior to user interaction - in effect building upon existing methods which seek to predict interactions based purely on object motion [ML03]. Further research intends to consider allowing event-triggered focus shifts (such as a sound near the user) to combat 'degree of blindness' effects (scenarios where clients are not mutually visible) as noted by the Velvet system.

A further advantage of the P-value field description is an ability to downsize in areas of high network traffic not only by a reduction in the volume of each aura, but also by directly changing the thresholds for the translation between values and their corresponding multicast groups. This permits emphasis to shift from displaying large numbers of clients at low resolution or smaller numbers with increased quality without necessitating any recalculations of aura structure or definition. The approach is generally de-

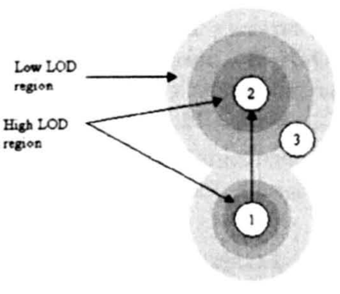


Figure 3: *Perceptual Relevance Model*

signed for scenarios where clients are viewing or interacting with large numbers of other clients in open environments, providing filtering for large scale scenarios such as virtual cities or combat simulations. Thus such techniques which offer a low processing overhead alongside scalability, and remain capable of increasing CPU demand to filter more accurately (by increasing the rate at which the aura is recalculated), are advantageous.

Preliminary implementations within the test system described in Section 4 indicate a sufficiently low CPU load to be viable; however, to accurately evaluate the success of the method a measure of quality of simulation is required (as noted previously). This is the focus of future work.

6. Summary

This paper has presented a brief review of the state-of-the-art regarding interest management in MMVEs, alongside providing a discussion of the potential for further analysis of perceptual relevance, and the shortcomings of existing perceptual techniques in a generic context. It has also provided some discussion into the criteria that need to be established for the comparative evaluation of interest management techniques, and outlined the development of a test system intended to provide facilities for such evaluation of performance is a wide range of scenarios. Finally, an approach has been described which seeks to provide both flexibility and scalability alongside a perceptually-oriented IM paradigm.

It is hoped such refinements to interest management will allow not only for more efficient use of bandwidth alongside improved quality of simulation, but will also allow for far greater extensibility, as content may be streamed at run time rather than downloaded in advance. Such environments would allow for far greater creativity and interactivity than existing environments, and subsequently offer far more potential.

References

[AWZ98] ABRAMS H., WATSEN K., ZYDA M.: Three-tiered interest management for large-scale virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology* (1998), pp. 125 – 129.

[BWA96] BARRUS J. W., WATERS R. C., ANDERSON D. B.: Locales: supporting large multi-user virtual environments. *IEEE Computer Graphics and Applications 16*, 6 (Nov. 1996).

[BWH03] BEEHAREE A. K., WEST A. J., HUBBOLD R.: Visual attention based information culling for distributed virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology* (2003), pp. 213 – 222.

[Cas01] CASTRANOVA E.: Virtual worlds: A first-hand account of market and society on the cyberian frontier. In *CESifo Working Paper Series* (Dec. 2001), no. 618.

[FS98] FRECON E., STENIUS M.: DIVE: a scalable network architecture for networked virtual environments. *Distributed Systems Engineering 5* (Sept. 1998), 91–100.

[Gre98] GREENHALGH C.: Awareness management in the MASSIVE system. *Distributed Systems Engineering 5*, 3 (Sept. 1998), 129–137.

[HPG02] HOSSEINI M., PETTIFER S., GEORGANAS N.: Visibility-based interest management in collaborative virtual environments. In *Proceedings of the 4th international conference on Collaborative virtual environments* (2002), pp. 143–144.

[Jac95] JACOB R. J. K.: Eye tracking in advanced interface design, virtual environments and advanced interface design. *Oxford University Press, Inc., New York* (1995).

[ML03] MORGAN G., LU F.: Predictive interest management: An approach to managing message dissemination for distributed virtual environments. In *Proceedings of the First International Workshop on Interactive Rich Media Content Production: Architectures, Technologies, Applications, Tools* (2003).

[Mor96] MORSE K. L.: *Interest Management in Large-Scale Distributed Simulations*. Tech. rep., University of California, Irvine, 1996.

[MR98] MACK A., ROCK I.: *Inattentional Blindness*. MIT Press, 1998.

[Mr02] MÁSA M., ŐÁRA J.: Generalized interest management in virtual environments. In *Proceedings of the 4th International Conference on Collaborative Virtual Environments* (2002), pp. 149–150.

[MT95] MILLER D. C., THORPE J. A.: SIMNET: The advent of simulator networking. In *Proceedings of the IEEE* (1995), pp. 83(8):1114–1123.

[OG02] OLIVEIRA J., GEORGANAS N.: Velvet: An adaptive hybrid architecture for very large virtual environments. *IEEE International Conference of Communications 1*, 2 (2002), 491–495.

[PB95] PRATT D. R., BRUTZMAN D. P.: Exploiting reality with multicast groups: a network architecture for large-scale virtual environments. In *VRAIS '95: Proceedings of the Virtual Reality Annual International Symposium (VRAIS'95)* (1995), IEEE Computer Society, pp. 38–45.

[PJR01] PASHLER H., JOHNSTONE J. C., RUTHRUFF E.: Attention and performance. *Annual Review of Psychology 52* (2001), 629–51.

[Red01] REDDY M.: Perceptually optimized 3d graphics. *IEEE Computer Graphics and Applications 21*, 5 (2001), 68–75.

[Tre86] TREISMAN A.: Features and objects in visual processing. *Scientific American 254* (1986), 114–124.

[Tse04] TSE P. U.: Mapping visual attention with change blindness: New directions for a new method. *Cognitive Science 28*, 1 (2004), 241–258.

[ZS00] ZYDA M., SINGHAL S.: *Networked Virtual Environments: Design and Implementation*. ACM Press, 2000.