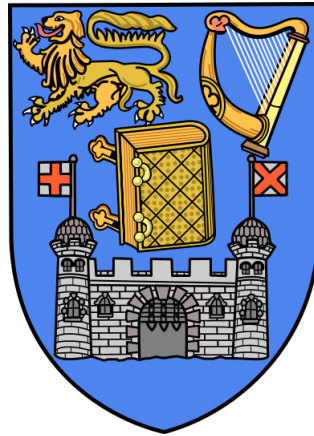


UNIVERSITY OF DUBLIN, TRINITY COLLEGE



Guiding Attention in Immersive 3D Virtual Reality

Author:

Sarah NOLAN

Supervisor:

Dr. John DINGLIANA

*A dissertation submitted in fulfilment of the requirements
for the degree of*

Master in Computer Science

Submitted to the

University of Dublin, Trinity College

May 2015

Declaration of Authorship

I, Sarah NOLAN, declare that the following dissertation, except where otherwise stated, is entirely my own work; that it has not previously been submitted as an exercise for a degree, either in Trinity College Dublin, or in any other University; and that the library may lend or copy it or any part thereof on request.

Signed:

Date:

Summary of Dissertation

This dissertation is concerned with guiding attention in an immersive 3D virtual reality. The improving quality in virtual reality has come with the added challenge of keeping the experience for users as realistic as possible. Attention guidance has been applied in 2D images, video and interactive video games to provide a sense of immersion. As applications become more photorealistic these techniques are required to be more subtle.

We investigated the use of attention guidance in virtual reality using the Oculus Rift™ head-mounted display. The Oculus™ affords a user more freedom through head tracking and isolates their vision from the real environment. The attention guidance technique used in this work is based on the Subtle Gaze Direction (SGD) technique. This technique uses a visual cue in a viewer’s periphery to draw their attention to a location. For our adaptation of SGD we developed two visual cues; modulation and saturation. Modulation is based on the modulation used in the original SGD technique for 2D static images. It was developed as a semi-transparent sphere that interpolates from blue to red (warm-cool modulation). Saturation is defined in this work as an increase in brightness along with saturation of an object in order to emphasise it in a scene. This was developed from techniques used to emphasise locations in illustrations in order to draw attention to them. Our adaptation of SGD places the visual cue in 3D space as opposed to 2D space. The visual cue also has a depth component allowing it to fit more naturally into a 3D virtual environment.

A user study implemented using the Unity® game engine was conducted to investigate the effectiveness of these visual cues at guiding attention. The hypotheses this study was investigating were that “Attention can be guided in a subtle way in 3D virtual reality” and also that “Attention guidance can be used to improve the performance of a visual search task in a 3D virtual reality”. This search task was designed to be purposely difficult to perform accurately in the absence of attention guidance. Participants were asked to press a key if they noticed a gesture being performed by one character out of a crowd surrounding them in a 3D virtual environment. Participants were also asked to let the researcher know if they noticed anything strange in the scene while performing the task.

The results showed a clear increase in performance of the visual search task when attention guidance was applied. On average without any attention guidance participants were only correctly identifying 54% of the gestures. The saturation visual cue increased this to between 62% at its lowest intensity and 89% at its highest. The modulation visual cue increased this to 69% at its lowest intensity and 88% at its highest intensity. From these results we conclude that we succeeded in guiding attention in a 3D virtual reality and increased the performance of a visual search task using attention guidance. However, we were not able to conclusively prove that the technique was subtle as participants noticed the visual cues. Based on our findings we outline a number of areas for further improvement, most notably in determining the location of a viewer's focus before placing a visual cue. In future research eye tracking would be beneficial to improving the subtlety of the guidance techniques developed in this work.

Acknowledgements

I would like to thank my supervisor Dr. John Dingliana for his help throughout the course of this dissertation. His advice and guidance were crucial to the development of this work.

I would also like to thank my friends and family for their continued support throughout the last year.

Sarah Nolan

University of Dublin, Trinity College

May 2015

Contents

Declaration of Authorship	i
Summary	ii
Acknowledgements	iv
Contents	v
List of Figures	vii
1 Introduction	1
1.1 Motivation	2
1.2 Methodology	3
1.3 Contribution	4
1.4 Summary of Chapters	5
2 Background	7
2.1 Virtual Reality	7
2.1.1 Immersion and Presence	8
2.1.2 Oculus Rift™	9
2.1.3 Unity®	10
2.2 Gaze Direction in Video Games	11
2.3 Models of Visual Attention	14
3 Previous Work	18
3.1 Quantifying Immersion in Virtual Reality	18
3.2 Subtle Gaze Direction	21
3.3 Directing Gaze in Narrative Art	27
3.4 Guiding Attention in Controlled Real-World Environments	30
3.5 Directing Gaze in 3D Models with Stylised Focus	32

4	Design & Implementation	35
4.1	Guidance Techniques	36
4.1.1	Modulation	36
4.1.2	Saturation	38
4.2	Design of Experiment	40
4.2.1	Experimental Procedure	40
4.2.2	Implementation	42
4.2.3	Data collection	48
5	Results & Evaluation	49
5.1	Trials with No Guidance	50
5.1.1	Results	50
5.1.2	Evaluation	50
5.2	Trials with Saturation Guidance	51
5.2.1	Results	51
5.2.2	Evaluation	52
5.3	Trials with Modulation Guidance	53
5.3.1	Results	53
5.3.2	Evaluation	54
5.4	Comparison	55
5.5	SaliencyToolbox	58
5.5.1	No guidance	58
5.5.2	Saturation	59
5.5.3	Modulation	61
6	Conclusions	62
6.1	Modulation	63
6.2	Saturation	63
6.3	Overall Conclusions	64
7	Future Work	66
A	Experiment Documents	69
B	Code Description	75
	Bibliography	85
	Glossary	89

List of Figures

2.1	Frames from virtual reality film “Sidra” stretched into a rectangle[Mil]	8
2.2	Oculus Rift™ Development Kit 2 (DK2)	9
2.3	The Unity® Logo[Unib]	10
2.4	Glowing Trail used to guide players in “Fable II” [Fab]	12
2.5	Arrows over characters indicate you can interact with them in “The Legend of Zelda: The Wind Waker”. Link’s eyes also look at important locations. [zel]	12
2.6	Water Flow Direction in “Left 4 Dead 2”[Vla]	13
2.7	Screenshot from “Mass Effect 2” where the radar to the goal is seen in the bottom right corner. Attention is also drawn to enemy in top centre using red icon.[ME2]	14
2.8	General architecture of the model[IKN98]	15
2.9	The normalisation operator[IKN98]	16
3.1	Scene and controller used in experiment[PPW97]	19
3.2	Hypothetical image with current fixation region F and pre-determined region of interest A. Inset illustrates saccadic masking	23
3.3	Plot of all Pearson coefficient of correlation values obtained by comparing the gaze distribution of modulated images with the corresponding natural gaze distribution	25
3.4	Narrative art used the in user study with panels highlighted from [MBS+12]	28
3.5	This figure shows the heat maps and scan paths for one participant from each of the two groups (no modulation and SGD) [MBS+12]	30
3.6	Annotated photograph of the real-world gaze manipulation setup[BSM+13]	31
3.7	From top to bottom: colour effects (desaturation, fade, blur); line effects (opacity, texture, density); all six combined[CDF+06].	33
4.1	Modulating sphere	36

4.2	From top left to bottom Right: No Modulation, Low Modulation, Medium Modulation, High Modulation	37
4.3	From left to right: No saturation, Low Saturation, Medium Saturation, High Saturation	39
4.4	Plan view of scene	45
4.5	Participant's view of scene	46
5.1	Average gestures correctly seen with no attention guidance .	51
5.2	Average gestures correctly seen with saturation guidance cues	53
5.3	Average gestures correctly seen with modulation guidance cues	54
5.4	Average time taken to notice (seconds) gesture after onset of visual cue	56
5.5	Average gestures correctly seen comparison	57
5.6	Histogram of the identified targets distance from the centre of the screen (normalised coordinates)	57
5.7	First location that draws attention in image with no cue . . .	58
5.8	The saliency map produced by figure 5.7 - most salient spots are brightest spots	59
5.9	First location that draws attention in image (using saturation guidance)	60
5.10	The saliency map produced by figure 5.9 - most salient spots are brightest spots	60
5.11	First location that draws attention in image (using modulation guidance)	61
5.12	The saliency map produced by figure 5.11 - most salient spots are brightest spots	61
7.1	Scene used for user study in [BSM ⁺ 13]	67

Chapter 1

Introduction

This work explores the use of gaze direction techniques in an immersive virtual reality. It takes inspiration from a variety of research covering perception and virtual reality. We make use of the Subtle Gaze Direction[\[BMSG09\]](#) technique. This technique works by using subtle colour modulation in a viewer's periphery to draw their gaze. This technique has been used in 2D images, video, paintings and a real world environment. It is adapted here to fit into a 3D virtual environment. A modulation visual cue was developed in 3D based on the visual cue originally used for SGD. A saturation visual cue was also developed from work done to emphasise locations in illustrations by manipulating how saturated/desaturated certain locations were [\[CDF+06\]](#).

The system implemented here is used to emphasise locations in a scene in order to draw a user's attention to them. This emphasis is placed per object/mesh. It

explores the use of three distinct intensities of two separate visual cues for drawing attention. A simple user study to determine the effectiveness of these visual cues has been conducted.

1.1 Motivation

Virtual reality provides the experience of immersion or ‘being there’ in a computer generated virtual environment. With this type of experience comes the challenge of providing an effective method of guiding the user without breaking this sense of immersion. Users can become disorientated and confused without some form of guidance in this type of environment. This is usually due to a lack of experience using virtual reality equipment along with being afforded more freedom in a virtual reality application than in a desktop application.

Applications of virtual reality include therapy, training, simulation, architectural walk through and entertainment. A key motivation in this research is the idea of using virtual reality to tell a story with the user feeling like they are a part of the narrative. Virtual reality has the ability to make a story or game come to life and evoke real emotions and experiences. A narrative is often needed to be told in a chronological order. In a virtual environment there is a need to ensure a user follows this order correctly. Attention guidance can be used for this. Designers of games with a traditional display often use some form of attention guidance to highlight

important objects. With the added degree of freedom afforded in head tracked virtual reality guaranteeing the user sees what the author intends is more difficult.

Attention guidance is a very well explored field. Research has been conducted into guiding attention on a static 2D image, in video and even in a real world environment.

Most research in this area takes advantage of physiological processes in how humans perceive and process visual stimuli. Many of these techniques can be applied in virtual reality.

1.2 Methodology

Two visual cues were developed to explore the problem of attention guidance in an immersive 3D virtual reality environment. The cues are intended to be sufficiently subtle so as not to break this immersion. We determine that a cue is subtle if it can go unnoticed by a viewer or does not degrade their viewing experience. The effectiveness and subtlety of these cues was examined by exposing users to different intensities of two separate visual cues while performing a visual search task. These cues are developed from previous research into attention guidance in both photorealistic and non-photorealistic scenes. A simple user study was designed using the Unity[®] game engine with the Oculus Rift virtual reality headset. The study was designed to examine how the performance of a simple visual search task changes when attention guidance is used. The visual cues are applied to the targets of the search task to

make them more obvious to a viewer. Participants were also encouraged to comment if they noticed any anomalies during the trials.

As an added analysis of the performance of our visual cues, the attention guidance mechanisms are tested using an existing tool for determining how gaze is directed about an image. This architecture is mainly used for 2D images. However, it is still useful for locating areas where attention will be guided in a 3D scene by using a screenshot.

1.3 Contribution

This work presents two visual cues for guiding attention in a virtual reality environment. Techniques have already been developed and successfully applied to guide attention when viewing images, video, 3D scenes and even a real world environment.

The visual cues developed for this work were derived from this research. We call these modulation and saturation. Modulation is based on the visual cue used in the Subtle Gaze Direction (SGD) technique and consists of a semi-transparent sphere that interpolates between red and blue. This visual cue appears as a halo in the location where attention should be drawn. The saturation visual cue is a combination of increased saturation and brightness, not simply conventional saturation.

The performance of these techniques were analysed by conducting a user study. This study had a participant focus on a visual search task that consisted of trials with a mix of no guidance, modulation guidance and saturation guidance. The results

indicate that both techniques were effective at guiding attention. There was an increase in performance of the search task with attention guidance compared to no guidance. However, the attention guidance cues were noticed at higher intensities and therefore did not achieve subtlety. This is due to being unable to determine where a participant's gaze is focused in the environment. If this could be tracked we could ensure a viewer never directly focuses on the visual cue. In previous research an eye tracker provided this information. These results have provided some valuable insights into potential future research.

1.4 Summary of Chapters

Chapter 2

This chapter provides some background on virtual reality and its applications. Focus is given to how attention guidance has been used in computer generated environments previously. Finally there is a summary of a model developed for predicting locations that draw attention based on the visual system of a macaque monkey (very similar to human visual system).

Chapter 3

This chapter details previous research done in the area of attention guidance. An explanation of the SGD technique is given along with its application in different areas. There are also descriptions of previous work conducted in

virtual reality and the use of ‘stylized focus pull’ to guide attention in non-photorealistic illustrations.

Chapter 4

This chapter discusses the design and implementation of the experiment in Unity[®]. The experiment is designed to measure how performance in a visual search task changes with the introduction of gaze direction.

Chapter 5

This chapter presents and evaluates the results obtained from the user study. The results from a saliency-based toolkit (based on the visual model presented in chapter 2) using screenshots from the user study trials are also presented.

Chapter 6

This chapter summarises the project and provides conclusions we have come to from analysing the results.

Chapter 7

This chapter discusses what future work could potentially be done in the area of gaze direction in virtual reality.

Chapter 2

Background

This chapter gives some insight into immersion in virtual reality along with how attention guidance has been previously been used in video games. There is also a description of the model used for predicting locations in a 2D image that draw attention.

2.1 Virtual Reality

Virtual reality (VR) is an area of technology that aims to immerse a user in a computer generated environment. These environments are three-dimensional and can be explored and interacted with by the user. VR achieves this immersion by using a head-mounted display ([HMD](#)) or pair of VR glasses. Along with visual input, sound or touch (haptic) feedback can be provided.



FIGURE 2.1: **Frames from virtual reality film “Sidra” stretched into a rectangle**[Mil]

Chris Milk, an interactive storyteller, has been using virtual reality technology to record life in some of the world’s most dangerous and destitute countries. He gives a TED talk where he discusses the production of a film in Syria which followed a girl named Sidra. He uses virtual reality to portray to people how it feels to live in that environment. A screenshot from this film is shown in figure 2.1. Milk emphasises the benefit of virtual reality in his talk: “When you’re sitting there in her room, watching her, you’re not watching it through a television screen, you’re not watching it through a window, you’re sitting there with her. When you look down, you’re sitting on the same ground that she’s sitting on. And because of that, you feel her humanity in a deeper way. You empathise with her in a deeper way.” [Mil].

2.1.1 Immersion and Presence

In ‘Measuring Presence in Virtual Environments: A Presence Questionnaire’ [WS98] Witmer and Singer attempt to quantify presence in virtual environments. Therein, they discuss immersion and how it contributes to a greater sense of presence and of ‘being there’. They describe presence as “the subjective experience of being in one place or environment, even when one is physically situated in another”. The

necessary conditions for presence seem to be a matter of focus. When a user is focusing their attention on a task they experience involvement. This makes them more involved with the virtual environment experience. Immersion finally is a psychological state characterised by perceiving oneself to be enveloped by, included in and interacting with an environment that provides a continuous stream of stimuli and experiences. A HMD is instrumental to the immersive experience as it isolates the user from their physical environment. When users discover that they can interact naturally in a virtual environment they become more immersed in that environment.

2.1.2 Oculus Rift™

The Rift™ is a virtual reality HMD developed by Oculus VR®. The Oculus™ DK2



FIGURE 2.2: **Oculus Rift™ Development Kit 2 (DK2)**

that was used in this work features several improvements over the first development kit. It has a higher-resolution (960x1080 per eye) low-persistence pentile AMOLED

display, higher refresh rate, head positional tracking, a detachable cable, and the omission of the need for the external control box[ocu]. The Oculus Rift™ was chosen for this project as it is a binocular HMD (camera in front of each eye) and fulfils the need to give users a sense of immersion.

2.1.3 Unity®



FIGURE 2.3: The Unity® Logo[Unib]

Unity®[Unib] is a cross-platform game engine developed by Unity Technologies and used to develop video games for PC, consoles, mobile devices and websites. It is noted for its ability to develop games across platforms. Unity® is one of the game engines that helped democratise game development by providing to developers the tools used by large game companies at little to no cost. This has also resulted in developers being able to focus on game/application design instead of on the underlying technology.

The Unity® editor provides a drag and drop environment for creating scenes. Scripts can be written in JavaScript, C# or Boo (a python-like scripting language). Unity3D is a ‘component-based game object system’[Unia] where every game object can have scripts attached. This became a useful feature when developing the user study as

it allowed greater modularity of the code e.g. code controlling different visual cues could be placed in separate scripts. This also resulted in more maintainable code.

The user study and techniques used in this work were designed and implemented using the Unity game engine. Although this engine is not specifically used for virtual reality environments there is an Oculus™ plugin available for it. The use of a game engine provided better control over the design of the gaze direction techniques and the scene. It also resulted in a more optimised application for the user study detailed in Chapter 4.

2.2 Gaze Direction in Video Games

With a narrative-style application that allows user control comes the challenge of directing the user to the correct parts of a scene in a chronological order. This work proposes the use of SGD to enable this manipulation. Gaze direction has often been employed in video games to draw the user’s attention to important items. An example of this type of user direction would be the ‘Glowing Trail’ or ‘Breadcrumb Trail’ used in “Fable II” and “Fable III”. It is a golden-coloured glowing trail that leads the player to objectives in the quest they are currently undertaking. This can be seen in figure 2.4. This mechanism works well in a non-photorealistic video game but could affect the immersive quality of a photorealistic virtual reality environment due to its unnatural appearance. Another example of some obvious guidance mechanisms employed can be seen in “The Legend of Zelda: The Wind Waker”.



FIGURE 2.4: Glowing Trail used to guide players in “Fable II” [Fab]



FIGURE 2.5: Arrows over characters indicate you can interact with them in “The Legend of Zelda: The Wind Waker”. Link’s eyes also look at important locations. [zel]

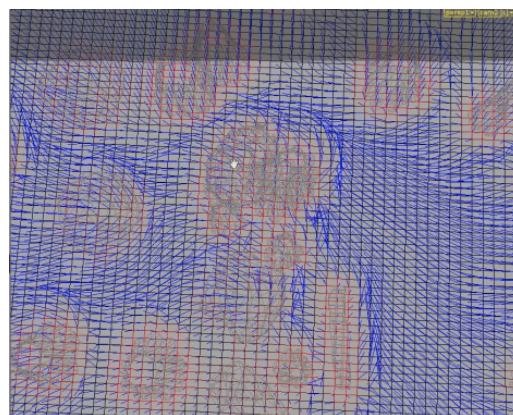
This game is animated in the non-photorealistic rendering technique cel-shading. An example of directing attention in this game can be seen in figure 2.5. Subtlety is clearly not the aim of this method however it fits into the game’s style. A more subtle approach for guiding players in locating important objects or clues was also

implemented in this game. Link, the player’s character, looks towards locations or characters that are important. This is often useful in solving puzzles throughout the game.

Directing users in this way is necessary to prevent them from becoming confused and lost. A better experience can be provided for the user by designing applications in such a way that the user has guidance. A similar motivation exists in directing users in a virtual reality environment. Without some form of guidance users can become disorientated and unsure of how to proceed. This is especially important in applications where the designer wants to convey a narrative to the user. A good example of directing the user in a subtle manner can be found in the design of the bayou level of “Left 4 Dead 2”. In a presentation given by Alex Vlachos (of Valve[®]), he outlines the use of directing water flow to help guide players through the bayou level[Vla]. This was necessary as it was evident from playthroughs that players were becoming confused and lost in the large swamp. They designed the level so that the



(A) Bayou scene



(B) Scene in Houdini with masks for flow direction applied

FIGURE 2.6: **Water Flow Direction in “Left 4 Dead 2”**[Vla]



FIGURE 2.7: Screenshot from “Mass Effect 2” where the radar to the goal is seen in the bottom right corner. Attention is also drawn to enemy in top centre using red icon.[ME2]

water would flow in the direction the user needed to go in order to advance. This resulted in 17% fewer wrong turns and users completing the level in less time. Some games guide players without using rendering techniques and instead use some form of a compass to guide them. Many games employ this mechanism such as “The Elder Scrolls: Skyrim”, “The Witcher” and “Mass Effect”. An example of such a compass can be seen in figure 2.7. The games that guide players in this way are often designed to be photorealistic. It seems to be a design choice to not degrade the quality of this photorealism by using unnatural rendering techniques to highlight important items.

2.3 Models of Visual Attention

‘A Model of Saliency-Based Visual Attention for Rapid Scene Analysis’[IKN98] by Koch et al. presents a visual attention system based on the behaviour and neuronal

architecture of the early primate visual system. This model simulates how attention is directed to and then shifted between salient features. Saliency is defined as “the perceptual quality by which an observable thing stands out relative to its environment”. The model presented in figure 2.8 builds on a model that is related to

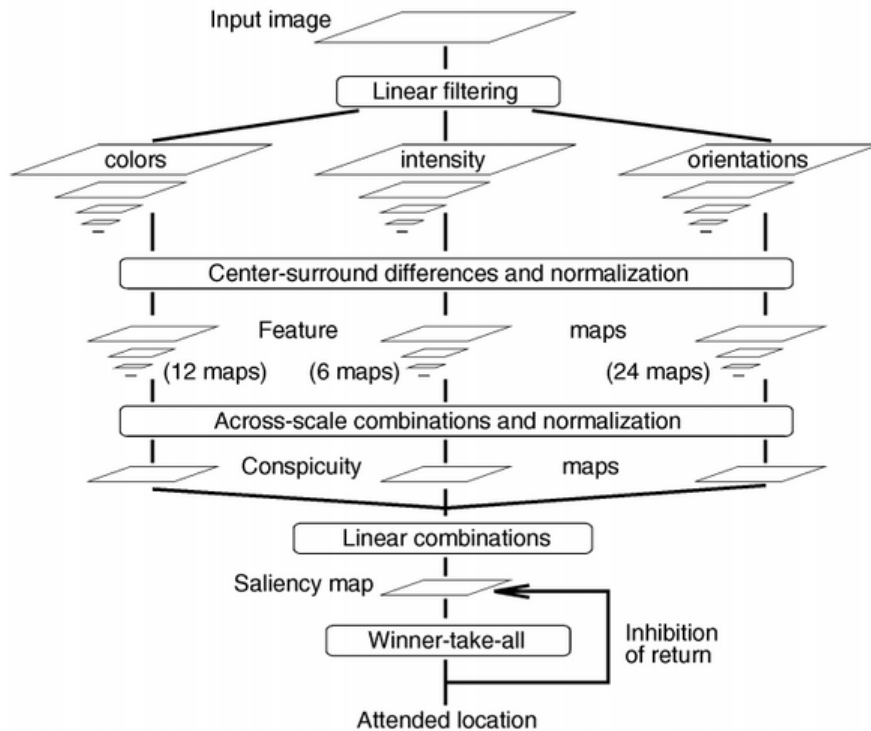


FIGURE 2.8: **General architecture of the model**[IKN98]

‘feature integration theory’ [TG80]. Feature integration theory states that “attention must be directed serially to each stimulus whenever conjunctions of more than one separable feature are needed to characterise or distinguish the possible objects presented”. This model locates salient objects/locations through the following steps:

1. Decompose visual input into set of topographic feature maps

2. Determine and keep locations which locally stand out in each map whilst suppressing non-salient locations
3. Feed these feature maps into the master ‘saliency map’

This model works by selecting salient features in a bottom-up manner. Top-down selection of salient features is dependent on a task or the viewer consciously searching for something. Features are computed by a set of linear ‘center-surround operations’. This is implemented as the difference between fine and coarse scale. The architecture is sensitive to local spatial discontinuities and therefore can detect locations that stand out from their surroundings. The features extracted are in modalities of colour, intensity and orientation. In total 42 feature maps are computed from the input image; six for intensity, 12 for colour, and 24 for orientation. These feature maps provide bottom-up input to a saliency map which is used to represent the saliency at every location. This map is a grid of scalar values to help guide the selection of locations based on the spatial distribution of saliency. The feature maps

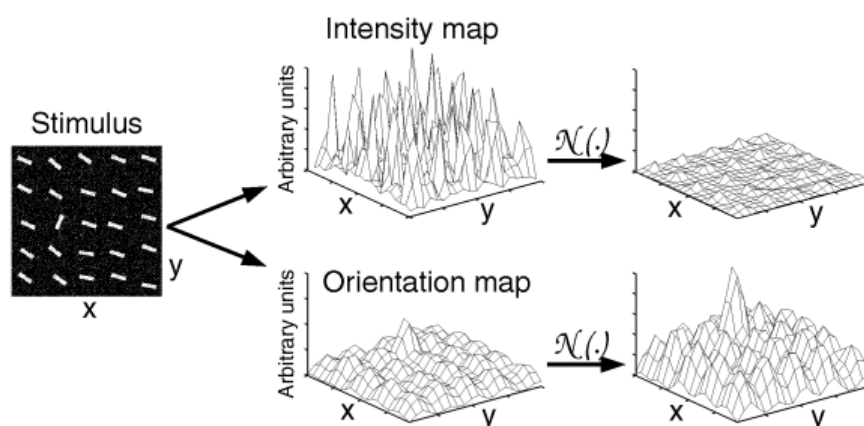


FIGURE 2.9: The normalisation operator [IKN98]

are combined using a normalisation operator to prevent masking of salient features between maps. This operator can be seen in figure 2.9. It works by normalising the values in the map to a fixed range $[0..M]$ to eliminate modality-dependent amplitude differences. M is the global maximum of the map. In this way only local maxima of activity are considered and homogeneous areas are ignored. When an active location differs from the average by a large amount it is strongly promoted. When the active location differs by a small amount it is suppressed. The feature maps are combined into three ‘conspicuity maps’, for intensity, colour and orientation. Three separate channels are used as they contribute independently to the saliency map. These maps are then normalised and summed into the final saliency map.

The maximum of the saliency map defines the most salient location and will be the first location to be fixated on. As the model is a neuronally plausible implementation it uses a similar process for selection as neurons in the brain do. A ‘winner-take-all’ neural network suppresses all other locations except the most active one. This yields dynamic shifts of the focus of attention by allowing the next salient location to become the winner and prevents it from immediately returning to the previous location. This ‘inhibition of return’ has been demonstrated in human psychophysics and is therefore suitable for this model.

Chapter 3

Previous Work

Previous research conducted in the areas of virtual reality and attention guidance are presented in this chapter.

3.1 Quantifying Immersion in Virtual Reality

This work is concerned with directing attention in an immersive and subtle way. Research has been carried out investigating the benefit of using virtual reality to provide training for tasks and whether this would improve on current methods implemented in desktop 3D e.g. 3D on a computer screen. In ‘Quantifying Immersion in Virtual Reality’[[PPW97](#)] Pausch et al. investigated this concept by comparing

the performance of a visual search task on desktop to the same search with a virtual reality headset. The motivation behind this research is to find which tasks and application merit the expense and difficulty of VR interfaces.



(A) Scene used for visual search task in Quantifying Immersion in Virtual Reality

(B) Controller containing 6DOF Tracker

FIGURE 3.1: **Scene and controller used in experiment**[PPW97]

To hold variables in the experiment constant they used the same HMD as both the head-tracked display and the stationary monitor. The HMD was bolted onto a ceiling-mounted post to hold it stationary for the desktop group. For testing the VR aspect participants were allowed to alter the view by moving their heads. The participants with the stationary monitor were given the use of a controller that contained the same electromagnetic tracker from the HMD device. This controller is shown in Figure 3.1b. This was done to reduce lag that was found when using a mouse.

Their pilot experiment involved 28 participants searching for easy-to-find, uncamouflaged targets at random locations in a virtual room. They found that participants using VR located targets 42% faster than those with the stationary setup. They claimed “they can better remember where they have already looked in the scene” and that the stationary users would double check locations.

To examine this further another experiment was conducted with 48 participants; 24 using VR, and 24 using desktop. The search task was performed in the scene shown in figure 3.1a. The task was to locate a target letter in the scene. After this, participants were required to search for a target that was not in the scene requiring a much more systematic search. They recorded how long it took a participant to confidently search the entire scene. They used test results illustrating how long it took a participant to locate a target that was present to predict how long a complete search of the scene should take. If a participant took longer than the predicted search time, they concluded that they were searching locations more than once. They found that the complete search time for VR users was only 1.4% over this predicted time and for desktop users it was 41% over this predicted time. This gave an indication that participants who were using VR to search the scene were better able to remember where they had already searched. This also indicated a much better performance of a task using VR over desktop.

The use of VR for training would only be useful if it resulted in better performance in real life compared to using desktop for training. They examined the transfer effects of VR versus desktop. They conducted another user study to examine this, where

VR participants performed ten searches initially using VR then ten searches using desktop. The desktop participants initially performed ten searches using desktop and then ten searches using VR. It was found that using VR first had a positive transfer effect; practicing VR improved their performance in the desktop. It was found that using desktop first had a negative transfer effect. The performance of the task in VR after desktop was worse than it was when performed on its own. This indicates that training using a desktop interface could actually degrade real-world performance. The immersion aspect of VR provides a better platform for training in real-world tasks. They conclude that these results convey how VR is beneficial.

3.2 Subtle Gaze Direction

‘Subtle Gaze Direction’[BMSG09] by Bailey et al. presents a novel technique for directing a viewer’s gaze about a digital image. A brief subtle image-space modulation is applied to a location in the viewer’s periphery to draw their attention. Eye tracking is used to determine where the viewer’s current fixation in the image is. The modulation is terminated before the viewer’s foveal vision enters the modulated region to ensure the viewer never directly focuses on the stimuli that drew their attention. This is what provides the subtlety of the method.

The aim of this technique is to guide a viewer’s attention to regions of an image they would otherwise not have attended to. These regions would be uninteresting such as low contrast, low colour saturation and containing uninteresting objects. They

are manually picked out per image beforehand and a modulation is applied to draw attention. By choosing these uninteresting regions the viewer's natural gaze pattern would be violated if they attended to them. Altering this gaze pattern would prove the effectiveness of the technique at directing the viewer's gaze.

This technique takes advantage of some previously studied behaviour of the human vision system. Humans have a very high density of cones at the fovea resulting in a higher visual acuity compared to the periphery i.e. locations viewed with the foveal vision are in focus. It has been shown that the peripheral region responds faster to stimuli than the foveal vision. This is due to fast-conducting optic fibers carrying signals from this region to the primary visual cortex. The optic fibers that carry signals from the fovea are much slower. The peripheral vision therefore locates regions of interest and directs the high acuity foveal vision to move ([saccade](#)) to fixate on these regions. The SGD technique exploits this behaviour by choosing regions in the viewer's periphery to apply modulations to. This technique also takes advantage of a visual phenomenon known as 'saccadic masking' to terminate the modulation before the viewer can scrutinise it. Saccadic masking (or saccadic suppression) is a series of neurological processes which ensure that any resulting gaps in the visual signal are not perceived.

The stimuli used to draw attention in this technique is either a warm-cool modulation or luminance modulation. These modulations alternate interpolations of the pixels in a small circular region with black and white in the case of luminance modulation,

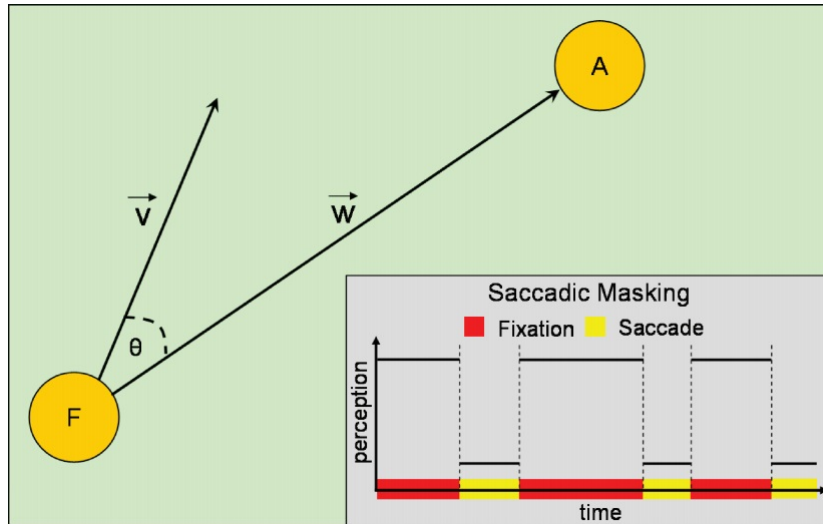


FIGURE 3.2: **Hypothetical image with current fixation region F and predetermined region of interest A . Inset illustrates saccadic masking**

or with a warm colour and a cool colour in the case of warm-cool modulation. These interpolations occur at a rate of 10Hz.

A visual explanation of how the technique is accomplished is presented in figure 3.2. F represents the last recorded fixation of the viewer and A represents the new region to draw attention to. The vector v is the velocity of the current saccade and the vector w is the vector from F to A . ϑ is the angle between v and w and is updated using feedback from the eye tracker after the modulation commences. A small value of ϑ (no greater than 10) indicates that the centre of the gaze is moving towards the modulated region and the modulation can be terminated immediately. This allows for greater subtlety in the technique. In order to discover the most subtle intensity for the modulations a small pilot study was conducted. Three viewers were asked to vary the intensity of the modulations in five random images until it was just intense enough to be noticed.

The effectiveness of this technique was determined by conducting an experiment to measure a viewer's gaze pattern as well as their perception of the image quality. Participants were divided into the control static group with no modulations and the modulated group with modulations. During the experiment the participants rested the head on a chin rest to prevent movement and were presented with images each displayed for 8 seconds with a 2 second break in between. The break allowed participants a break and also to give a quality score to the image out of ten. Quality was not defined by the researchers. Data files compiled for each participant included the following data; time elapsed, filename of the image being displayed, position of modulated region, type of modulation and the location of the current fixation.

The data from the experiments showed how quickly the viewer responded (activation time), how close they got to the target, overall change of the viewing pattern and the perceived quality of the image. It was found that the modulations did draw the viewer's attention with users responding on average 0.5 seconds after the onset of the visual cue. However, their gaze did not always reach the modulated regions. This is due to the termination of the modulation to stop the viewer from being able to focus on it. The viewers reported not seeing anything unusual but the images with modulation were consistently found to be of lower quality. They hypothesised that this was due to violating the viewer's natural gaze pattern and giving them less time to view more salient regions of the image.

Gaze patterns were recorded for each image by dividing the image into a 5x5 grid and counting the number of fixations in each region. They compared the gaze patterns

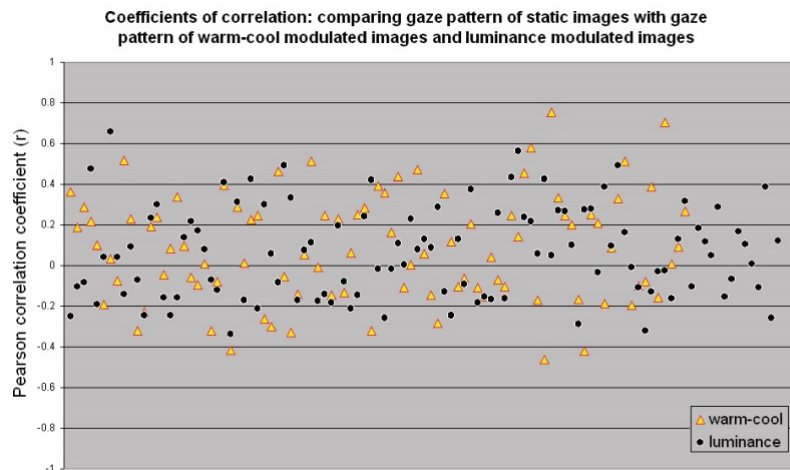


FIGURE 3.3: Plot of all Pearson coefficient of correlation values obtained by comparing the gaze distribution of modulated images with the corresponding natural gaze distribution

from the control group against the modulated group and found a poor correlation, indicating that the modulations were successful in altering the natural gaze pattern. They also found that the luminance modulation was more successful at shifting attention than the warm-cool modulation. This can be seen in Figure 3.3 where the coefficients of correlation for the luminance modulated images occupy a narrower range than the warm-cool modulated images. This was considered to be due to the luminance modulation being more noticeable across all images when compared to the warm-cool modulation that can sometimes be difficult to detect.

The SGD technique is built upon in [MBG08]. Bailey et al outline the application of SGD in guiding participants while performing a simple visual search task; counting targets in an image. They first investigated performing a search task using three different conditions; Group 1: no modulation, Group 2: subtle modulation, Group

3: obvious modulation. For this set of experiments 24 images were used - six environments with four different target counts. The targets were small transparent spheres evenly distributed within the scene. To prove the efficiency of SGD at locating targets some spheres were placed in locations that made them hard to see. Participants were asked to count the targets in the scene and report it verbally.

It was found that SGD did slightly improve the results of the search task even though the modulations were noticed in the obvious modulation group. This was found by examining the average response times across all three conditions as well as comparing the reported count of the targets to the actual count to define a correlation. Higher correlation corresponded to more accurate task performance. They found higher correlation for the modulated groups than for the static group, with correlation values of 0.80, 0.89 and 0.90 for Groups 1, 2, and 3, respectively. This indicated an improvement in search task performance in the presence of SGD. The obvious modulations were noticed but were reported to have helped viewers identify targets they would have otherwise missed.

Another set of experiments was conducted with the same conditions but with distractors added into the modulated group. These were to analyse how search task performance changes in the presence of false positives. The aim was to see how SGD would perform if modulations were driven by an algorithm that produces false positives. Three false positives are put in random locations away from the targets in each image. The correlations produced from these experiments were much higher at a 0.93 average over all the images. The results indicated that the presence of

distractors increased the overall performance of the visual search task. They hypothesised that this may be due to distractors drawing a viewer's gaze over a larger portion of the image, and thus allowing them to spot more targets.

3.3 Directing Gaze in Narrative Art

McNamara et al. investigated whether the SGD technique could be used to guide a user in viewing locations in a piece of narrative art in the correct order in “Directing Gaze in Narrative Art” [MBS⁺12]. Narrative art was commonly used during the Renaissance period to depict stories from the Bible. In that time period audiences were conditioned to recognise repeated elements in a frame and identify panels. This helped them understand the intended order. Today people are not able to accurately ‘read’ the paintings in the same way. They propose using SGD to guide viewers to the correct order of panels in the painting without disrupting their visual experience.

SGD is implemented here in the same manner as it was in section 3.2, however they only use luminance modulation. In their they investigated if SGD was effective at guiding a viewer to locations in a specific sequence. This property would be very useful when guiding users in a narrative in a 3D virtual reality as it would require them to be guided to events in a sequence (chronological order). They conducted a user experiment to determine the effectiveness of SGD at guiding users when viewing narrative art. They separated each art piece into rectangular panels as can be seen in figure 3.4, each of which enclosed a sequence in the story. Participants were



FIGURE 3.4: Narrative art used the in user study with panels highlighted from [MBS⁺12]

split into a control group with no attention guidance and a non-control group where attention guidance was applied. They were presented with a short tutorial to allow them to become familiar with the experimental procedure and user interface. Images were presented for a short period of time which was proportional to the number of panels in the piece. Modulations were applied to the relevant panel in the sequence but never when the participant was directly looking at them. The modulations were terminated when the participant's gaze moved towards the correct panel. Using the previous finding from 'Subtle Gaze Direction' [BMSG09] that viewers typically responded to the visual cue in 0.5 seconds, a 0.5 second delay was added between successive modulations. After viewing the piece for the predetermined duration the panels were outlined for the user. They were instructed to click the panels using a

mouse in the order that they thought was accurate.

To analyse the results of this data they used Levenshtein distance, a string metric to compute differences between sequences. They used this metric to compare the participant's chosen sequence with the correct sequence, with lower numbers indicating a better result. This provides a value that indicated how close to the correct order a sequence is. They found that participants in the control group had a mean distance of 57.32 and the non-control group had a mean distance of 34.79. This indicated that the participants who had attention guidance were consistently more accurate at predicting the correct sequence. They also compared heat maps that displayed where participants were frequently focusing their attention. These heat maps revealed that there was an increase in gaze coverage and attention to the relevant locations in the image in the non-control group. They concluded that SGD was effective at improving the performance on a within-image panel ordering set without noticeably disrupting the visual experience of the image.

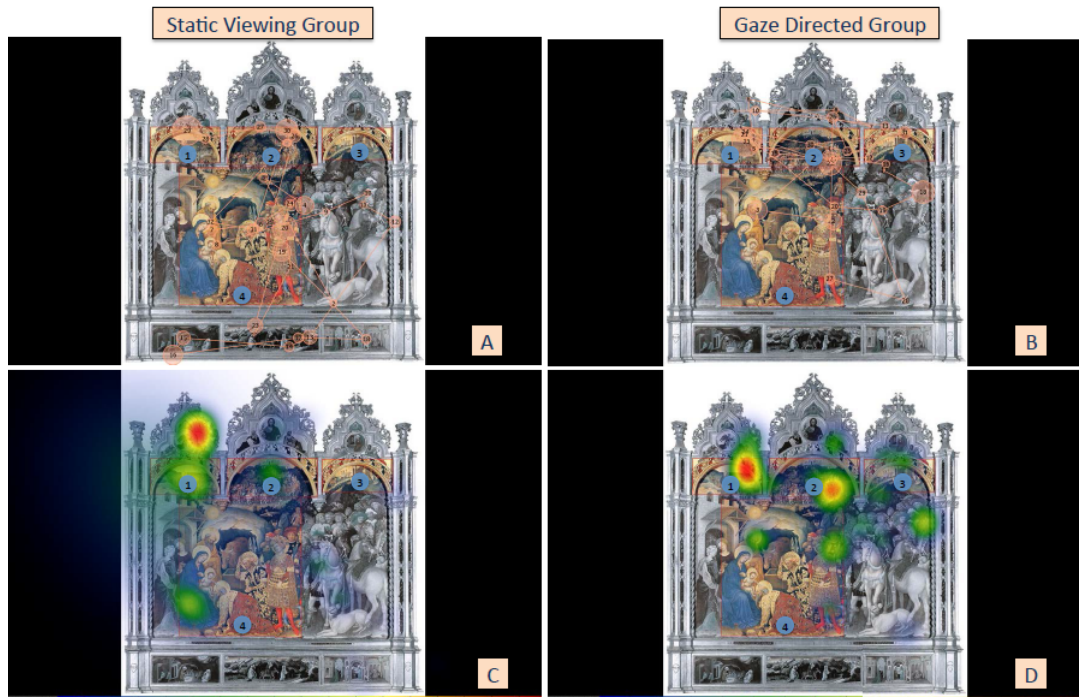


FIGURE 3.5: This figure shows the heat maps and scan paths for one participant from each of the two groups (no modulation and SGD) [MBS⁺12]

3.4 Guiding Attention in Controlled Real-World Environments

‘Guiding Attention in Controlled Real-World Environments’ [BSM⁺13] by Bailey et al. uses the SGD technique to guide attention in a controlled, real world environment. The two main challenges of adapting the technique to a real world environment are: (1) determining what the viewer is currently paying attention to, and (2) projecting a visual cue onto other parts of the scene to draw attention there. These are similar challenges that are faced when adapting the same method to virtual reality.

As can be seen in figure 3.6 the viewer is equipped with eye tracking glasses. These can be used to track the viewer's gaze and monitor what they are currently paying attention to. The glasses are also equipped with a front facing camera that captures the scene that the viewer is looking at. A projector is used for projecting the visual cues to guide attention. Features are extracted from around the fixation point in the scene using the scale-invariant feature transform image processing technique(SIFT). These features are then compared against precomputed features from the projectors point of view to determine what the viewer is focusing on. This allows them to monitor features on the source image where the viewer is looking and to record it. Once they know where the viewer is attending to they can subsequently redirect the viewer's attention to other parts of the scene.

They performed a user study to determine the effectiveness of their system. The participant viewed a simple scene consisting of eight objects. They attempted to guide the viewer through a sequence of objects. These objects were briefly highlighted by

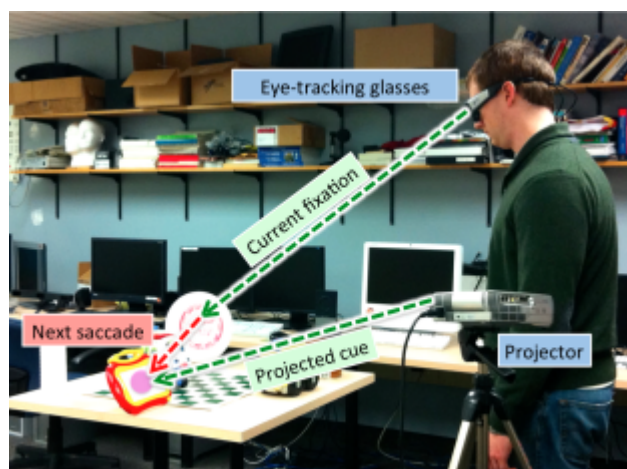


FIGURE 3.6: Annotated photograph of the real-world gaze manipulation setup[BSM⁺13]

projecting a brief luminance from the projector. They extract the viewing sequence from the eye tracking data by identifying the first fixation that occurs after the onset of the visual cue. They observed that their technique was effective and had a similar response time (0.5 seconds after the onset of the visual cue) as the original use of SGD in 2D images.

3.5 Directing Gaze in 3D Models with Stylised Focus

Research has been conducted into gaze direction in non-photorealistic scenes. These techniques are not completely adaptable to photorealism but they do provide a good background for attention guidance in 3D. Examples of this are the techniques developed by Cole et al.[CDF⁺06]. They developed an effect called ‘stylized focus pull’ where the system automatically renders the scene with emphasis in the area of interest. Many of the non-photorealistic techniques that are used in this paper such as control of line density are not transferable to a photorealistic rendering approach. However, certain techniques such as controlling colour saturation and contrast can be used for photorealism as well as non-photorealism.

The research conducted in this paper is largely motivated by architectural rendering. Four qualities were controlled to emphasise or de-emphasise part of an image:

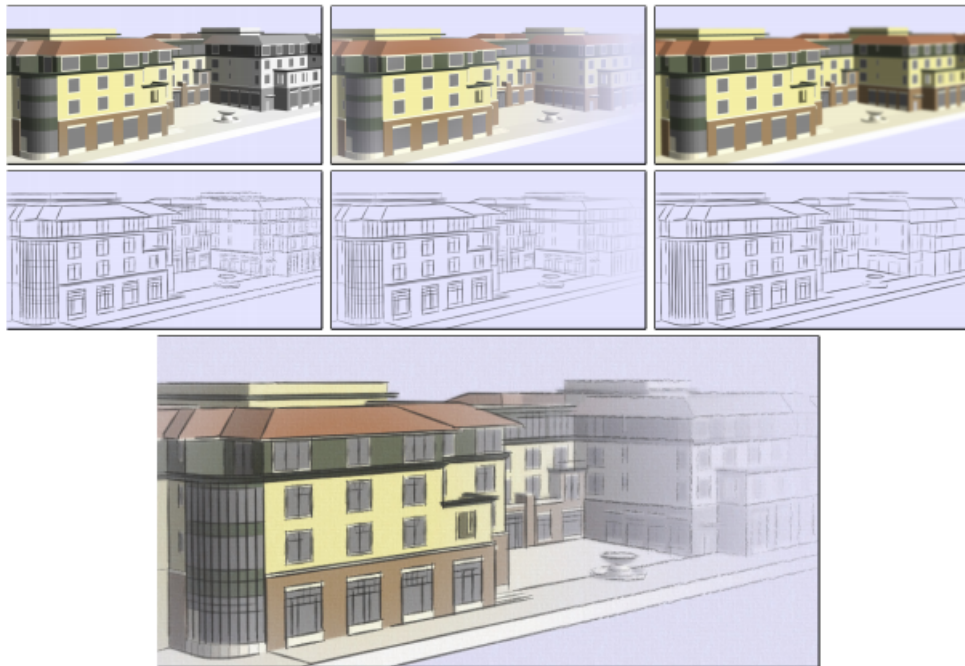


FIGURE 3.7: **From top to bottom: colour effects (desaturation, fade, blur); line effects (opacity, texture, density); all six combined**[CDF⁺06].

contrast, colour saturation, line density and line sharpness. An image of normalised scalar values was used to indicate how much focus to place at each point.

The system described in this work provides four intuitive focus models that allow the artist to express this emphasis. Segmentation is the simplest model and works by assigning focus to a labelled set of discrete components. The Focal Plane model is inspired by a real-world camera lens and expresses focus as the distance from the ‘focal plane’. The 2D Focal Point model allows the artist to pick a 2D point and the focus is taken from the 2D distance. This achieves a foveal effect that can feel quite natural. The 3D Focal Point model allows the artist to pick a 3D point where the focus will fall off radially in 3D. This is the most intuitive for 3D scenes and distinguishes this work from previous work. They based their ‘stylized focus pull’

technique off of the ‘focus pull’ device used in live action films to shift the focus of the camera lens.

They implemented eight effects that responded locally to the emphasis in the image - three colour effects and five adjust line qualities. The colour effects were desaturation, fade and blur. Each of the effects had a transfer function. The desaturation transfer function had four parameters for adjusting the saturation per point. This function was dependent on the emphasis at that point. An example of the result of this emphasis can be seen in figure 3.7. As you can see from the figure the desaturation, fading and blurring are used to de-emphasise part of the scene so the building on the right stands out.

They found their techniques for emphasising parts of an image were effective for drawing the eye to these locations. Eye tracking was used to monitor viewings of the images that were emphasised and ones that were not emphasised. The results indicated that viewers examined the emphasised regions more than they examined the same location in the non-emphasised image of the same scene.

Chapter 4

Design & Implementation

The goal of this work is to investigate whether techniques can be developed to guide attention in an immersive 3D virtual reality. This work is based on the hypothesis that “Attention can be guided in a subtle way in 3D virtual reality” and also that “Attention guidance can be used to improve the performance of a visual search task in a 3D virtual reality”.

To determine whether these hypotheses are plausible, a simple user study was conducted. The aim of this user study was to investigate the effect these guidance techniques have on the performance of a visual search task. We also investigate whether the use of these techniques is subtle enough to go unnoticed by asking participants in the study to provide feedback on their observations.

4.1 Guidance Techniques

4.1.1 Modulation

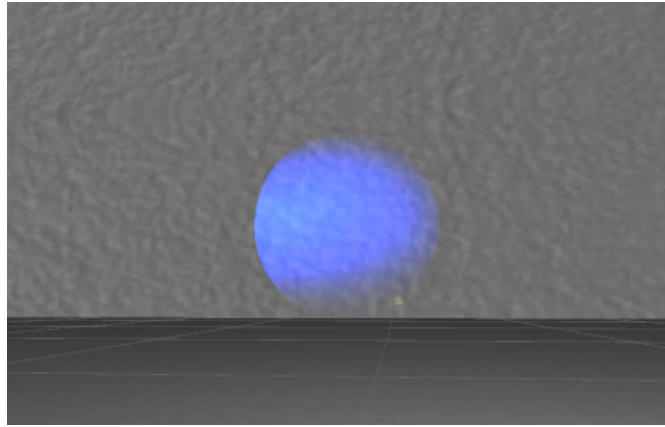


FIGURE 4.1: **Modulating sphere**

The modulation technique used here builds on the use of warm-cool [modulation](#) described in section [3.2](#). In previous implementations of SGD the modulation is rendered in 2D image-space. 2D image-space is only concerned with emphasising the location in pixels on the final image rendered. The cue is placed using a 2D coordinate that takes no account of the depth of the location in a scene that attention should be drawn to. The approach used in section [3.4](#) performs SGD by using a projector to project a brief luminance on the object/location where they want to direct attention. Instead of being rendered on the 2D [focal plane](#), the modulation occurs in 3D space. Using 3D space in this way allows depth to be taken into account and is more compatible with how humans perceive their environment. Similarly, the visual cues in this work are implemented in 3D world space for a virtual reality 3D scene. This is achieved by using a semi-transparent sphere that appears as a

halo around an object/location; where we want to draw attention to. This sphere is the 3D version of previous visual cues developed for image space where a circular region was modulated to draw attention. Unlike the 2D cue, this 3D cue has a depth position and thus is compatible with the 3D binocular stereo of the Oculus Rift™. The colour of this sphere interpolates between red and blue (warm-cool). The sphere is semi-transparent so that it adds its colour to (or modulates) the background. In



FIGURE 4.2: **From top left to bottom Right: No Modulation, Low Modulation, Medium Modulation, High Modulation**

order to create this modulating sphere the built-in Unity® particle [shader](#) was used as a base. This shader was used to render the sphere. An intensity value input was added to the shader to allow control over the intensity of the sphere. The intensity defines how opaque and saturated the sphere is. The shader already contained a tint-colour input which was used to interpolate between red and blue. This colour is multiplied by the intensity value to increase its saturation. The intensity value

is also used as the maximum alpha value of the sphere i.e. how opaque the sphere becomes. The different levels of intensity can be seen in figure 4.2.

The sphere was blended between the blue seen in the figure and red creating a purple hue in between. A script written in C# was used to interpolate between red and blue, and also to fade the sphere in and out at a target location. The interpolation between red and blue was achieved by constantly updating the sphere's tint-colour. The sphere initialised with a tint colour of blue. A predetermined duration was given as the total time to interpolate over with red being the colour to interpolate to. The intensity value was also faded in and out to allow the sphere to appear and disappear seamlessly. This fading in/out is achieved by using linear interpolation of the alpha value of the colour for the duration of the cue. The alpha value is the fourth component of the colour of the sphere and it controls the opacity of the object. The intensity values used for low, medium and high intensities respectively were 0.05, 0.1, 0.15. These values were chosen by observation.

4.1.2 Saturation

For this visual cue the intensity of the character's colour was increased. For convenience we refer to the cue as saturation although it is in fact an increase in brightness along with saturation. This design was based on the techniques described in section 3.5. These techniques use saturation and colour to emphasise certain parts of an illustration. Similarly, the increase of saturation and brightness was used here to

emphasise the target object and draw attention to it. This manipulation of colour in certain locations is based on the findings described in section 2.3. They found that local changes in colour contribute to the saliency of a location, allowing it to locally stand out from its surroundings.



FIGURE 4.3: **From left to right: No saturation, Low Saturation, Medium Saturation, High Saturation**

Saturation was fully applied to a target object/mesh through the use of a saturation shader. This shader is a bump-diffuse shader that has an added ‘saturation factor’ input. This factor was used in equation 4.1 to saturate and brighten the colour of the character.

$$(4.1) \quad pCol.rgb = \text{lerp}(\text{luminance}(c.rgb), c.rgb, sf) + c.rgb * (sf - 1)$$

where $pCol$ is the final colour of pixel, $c.rgb$ is the calculated red-green-blue(RGB) value from the object’s texture and sf is the saturation factor. This equation is not solely altering saturation as it also effects the luminosity of the object. The saturation is achieved by blending the original colour with the luminosity/grayscale

using the saturation factor as the blend value. The second part of the equation adds a fraction of the colour to the saturated colour to brighten it. Increasing only saturation resulted in the character used in the user study looking very unnatural e.g. orange hair. By increasing brightness and saturation together the unnatural colour is ‘washed out’. Normally saturation is altered by converting the RGB colour into its hue-luminance-saturation(HLS) counterpart. The saturation channel is then increased and the colour is converted back to RGB. This method was avoided as it resulted in too many calculations being executed in the shader when the shader should be kept simple.

In the user study, the character’s saturation was faded in by interpolating from the character’s normal saturation factor of 1.0 to the intended saturation factor. The interpolation is done in the opposite direction for fading out. This was accomplished by using simple linear interpolation over the duration of the cue, the same as was done in modulation. The saturation factors used for low, medium and high saturation respectively were 1.05, 1.15, 1.25. These values were chosen by observation.

4.2 Design of Experiment

4.2.1 Experimental Procedure

This section describes the user study conducted for examining the effectiveness and subtlety of the visual cues described in the above section. The user study consisted

of a simple visual search task where users were shown a 3D VR scene in which they were surrounded by identical characters. The participants were tasked with noting when a gesture (nodding or shaking of the character's head) was performed by the characters, one at a time. Attention guidance cues were applied in some of the trials to investigate how they affected the participant's performance. The participants were also asked to let the researcher know if they noticed any anomalies during each trial.

The experiment consisted of 21 one-minute trials. Three trials had no attention guidance (the control trials), nine had modulation at three intensities applied and nine had saturation at three increasing intensities applied. The same scene was used for each trial. The trials were randomly ordered with no two consecutive trials using the same intensity for the guidance cue being tested. There were approximately ten gestures per trial that were randomly triggered on characters in the participant's field of view. No more than one character performed a gesture at one time.

The participants were asked to mention between trials if they noticed anything strange during the previous trial. Before the trials began they were presented an empty scene for displaying the health and safety warnings from the Oculus Rift™ [SDK](#). After this scene they were shown a scene with a single character standing in front of them. By pressing a key the character would cycle through the two gestures the participant would be looking for. Once the participant was satisfied that they could recognise the gestures they were moved onto the trials. During the trials the participants were instructed to press a key if and when they noticed one of the

behaviours. At the end of the trial a menu popped up asking them to move onto the next trial. This served as a break between trials. At this point the participants were instructed to let the researcher know what trial they were moving on to. This was so the researcher could keep track of what trial they were on and associate any feedback at that point with the previous trial. The participants were also given breaks every few trials as required. This break consisted of the experiment being paused and the participants removing the headset. Once they felt comfortable to continue, the experiment resumed. The entire experiment took approximately 30 minutes. The information sheet and informed consent form associated with the user study can be found in appendix A.

4.2.2 Implementation

The following experiment was designed to determine the effectiveness of the two visual cues at guiding attention. This experiment was implemented using the Unity[®] game engine (see section 2.1.3). The experiment was adapted from previous experiments that have been conducted in perception research. The use of virtual reality technology such as the Oculus Rift[™] comes with the risk of cyber sickness. This risk was considered in the design of the experiment. The trials were limited to one minute in duration with small breaks between each trial due to this risk. The longer breaks afforded to participants every three to four trials were used to minimise this risk. This risk of cyber sickness needed to be fully outlined when applying for ethical approval. This also resulted in ethical approval not being granted for a significant

length of time while all scenarios were considered. As a result there was only sufficient time to conduct one experiment as more experiments would require the same process. The documents associated with the ethical approval process can be found in the appendix B.

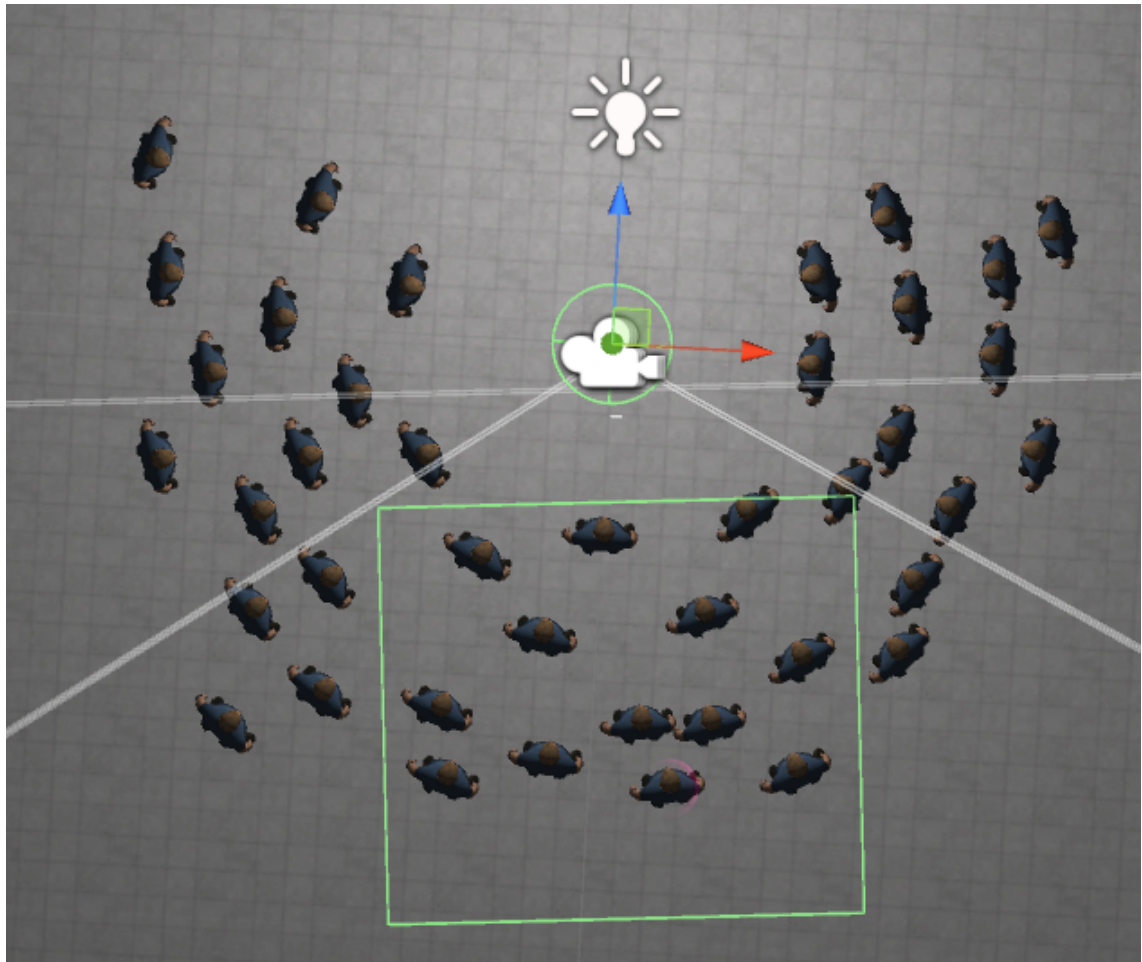
In order to collect as much useful data as possible the scene for the user study was purposely designed to be simple. The scene was designed this way to remove other unknown factors in order to better compare the effect of the different cues and triggers. The background scene needed to be plain to remove any accidental bias that could be caused by it. This allowed the viewer to maintain a better focus on the characters. The dull grey of the scene also worked as a good background for the warm-cool modulations to stand out against. The gestures were designed to be very subtle in order to make the task more difficult. This was implemented using the animation system in Unity[®] called Mecanim to restrict the characters head movement. This resulted in the nodding/shaking of the head being very slight and therefore much less likely to be noticed. For this task, learning how to use the Mecanim system was required along with importing characters along with their animations into Unity[®]. The animations were developed in Blender[®], an open-source 3D content-creation program. Characters were exported from Blender in [FBX](#) format and then imported into Unity[®].

The viewer was placed in the centre of the scene and surrounded by an arc of characters facing them. A script needed to be written to ensure every character was facing the user's direction at all times. This was achieved for each character by calculating

the vector from the character to the user by subtracting the user's position from the target's position. The resultant vector defined the direction the character needed to face. A quaternion was then calculated from this using the function 'LookRotation' defined on quaternions in Unity[®]. A quaternion has four components, the first three define an axis and the fourth defines an angle to rotate about that axis. This quaternion rotation was then applied to the characters transform rotation. Each character was identical. The mesh for the character was provided by the Graphics, Vision and Visualisation (GV2) group in Trinity College Dublin for the purpose of this experiment. The character can be seen in figure 4.3. This character was suitable for the experiment due to his plain clothing and features. Using the same mesh in this way guaranteed that no character was more salient or distracting than another character.

Figure 4.4 provides a view of the scene from above. The participant was placed where the camera is located with the characters placed in an arc in front of them. This was to give the participant the opportunity to turn their head in the virtual reality environment and still have characters in view. Only characters within the field of view could perform a gesture. This ensured that no gesture was missed by appearing off screen where the participant could not see it. In figure 4.5 the participant's view of the scene is shown. The characters were placed in such a way that no character was occluded by another character.

In Unity[®] a bounding box can be added to a game object. Normally this is used for collision detection e.g. to prevent an object falling through the floor. Bounding

FIGURE 4.4: **Plan view of scene**

boxes can also be used to trigger events in scripts. A bounding box can be flagged as a trigger (to prevent it from being associated with collision detection) and a script with functions ‘OnTriggerEnter’ and ‘OnTriggerExit’ can be used to define what happens when another game object enters/exits the bounding box. For this experiment only characters in the field of view were to be triggered to perform a gesture. A bounding box was attached to the front of the camera to capture possible targets. This can be seen in figure 4.4. The bounding box rotated along with the HMD allowing it to constantly cover the area in front of the camera. When characters entered the bounding box a check was performed to ensure they would



FIGURE 4.5: **Participant's view of scene**

appear on the screen using the screen coordinates. In Unity[®] a camera object has a function that converts a game objects position to normalised screen coordinates. If these coordinates are outside of a $\{0.0-1.0\}$ range the object will not appear on screen. Along with this check there was a check to ensure the character's face was fully visible. Two raycasts were used to determine this. Ray casting is the use of ray-surface intersection tests to solve a variety of problems such as collision detection [Rot82]. A ray is cast into the scene to determine what object it hits first from a specified location, often the camera position. Raycasts were used to determine that both the left and right half of the character's face were visible on screen. If either of the raycasts did not hit the target then the character's face was not fully visible. All these checks were coded in C# in the 'Trigger' file that was attached to the player game object. The bounding box notably did not encompass all characters that could appear on the screen. The characters that it did not include appeared far back in the periphery of the screen and could be very difficult to see in the Rift[™]. Characters that pass the previously described checks were added to a queue. When

a character exited the bounding box they were removed from this queue. A target was then randomly chosen from the queue to perform a gesture.

The gesture type was randomly chosen per character. Both gestures were purposely very subtle to ensure they were difficult to notice under normal conditions. Each gesture lasted between 3-5 seconds and gestures had 2-5 seconds between them. The duration for each was chosen randomly per gesture. This was to prevent a pattern forming that the participant would recognise. All random numbers were generated by seeding a Random with the trial number. This ensured very similar trials per participant by using pseudo-random numbers. The visual cue was triggered on the character at the same time the gesture was. It was also terminated at the same point the gesture ends. This is a departure from the use of SGD in previous implementations where the visual cue was terminated once the viewer's gaze moved towards the target location. This was not possible to implement due to the lack of eye tracking in the Rift™. When the visual cue was triggered it faded in until half of the predetermined duration of the gesture had elapsed and then it began to fade out. This was to allow it to blend more seamlessly into the environment and prevent it from abruptly appearing. If a gesture ended without being noticed (the key being pressed) the data of that gesture was logged as not seen. When the trials time runtime had exceeded its intended duration it was flagged that the trial ended. The current gesture was allowed to finish before the trial was paused and ended. This was done to allow the participant time to notice the gesture without the trial ending before the gesture did.

A description of each of the scripts used in the above user study can be found in the Appendix.

4.2.3 Data collection

The following data was collected during the experiments per gesture: the gesture type (Nod/Shake Head), the target position in world coordinates, the distance the target was from the viewer in the scene, the target position in normalised screen coordinates, whether the gesture was seen (1 or 0), the time in seconds it took the participant to press the key after the gesture started (activation time), the visual cue being tested, the visual cue intensity and the target's distance from the centre of the screen in screen space. Data was recorded for a gesture either when it was seen (the participant presses the key) or when it was fully terminated (in the case of no key press). The gesture was fully terminated one second after it ended. The reason for this was to allow for the participant to register it as seen if they noticed it just before it ended. If the participant saw the gesture and pressed the key the gesture is recorded as seen along with the activation time. If the participant did not see the gesture and it was fully terminated it was then recorded as not seen with the activation time set to 0. The rest of the data was recorded in both cases. Data was written to a CSV file to allow for importing it into statistical software such as Microsoft[®] Excel[®] and Minitab[®].

Chapter 5

Results & Evaluation

This chapter presents the results from the user study as well as an evaluation of those results. Eleven people volunteered to participate in the user study described in section 4.2.1. All participants were between the ages of 22 and 34. All participants had normal or corrected-to-normal vision.

The results of using the SaliencyToolbox[WK06] application are also presented. This application is based on the architecture discussed in section 2.3. The results for the percentage of gestures seen is achieved by finding the mean of ‘seen’ gestures which is 1 for seen or 0 for not seen. Time noticed is defined as the time taken to press the key after the gesture started. We acknowledge that there is some time elapsed between noticing the gesture and pressing the key.

5.1 Trials with No Guidance

5.1.1 Results

On average participants correctly identified 54% of the cues when there was no attention guidance applied. Results also show that the time taken to notice the gestures once they began was on average 1.75 seconds (see figure 5.4). Participants pointed out the difficulty of correctly identifying the very subtle gestures.

5.1.2 Evaluation

Poor results were expected with no attention guidance due to how subtle the actual gestures were. The intention was to have a task that needed and would benefit from attention guidance. The increase in performance throughout the trials can be seen in figure 5.1. This indicates that there was some learning as the average gestures noticed rose starkly between trial 1 and trial 11. With training it appears that participants were able to correctly identify 65% of the gestures.

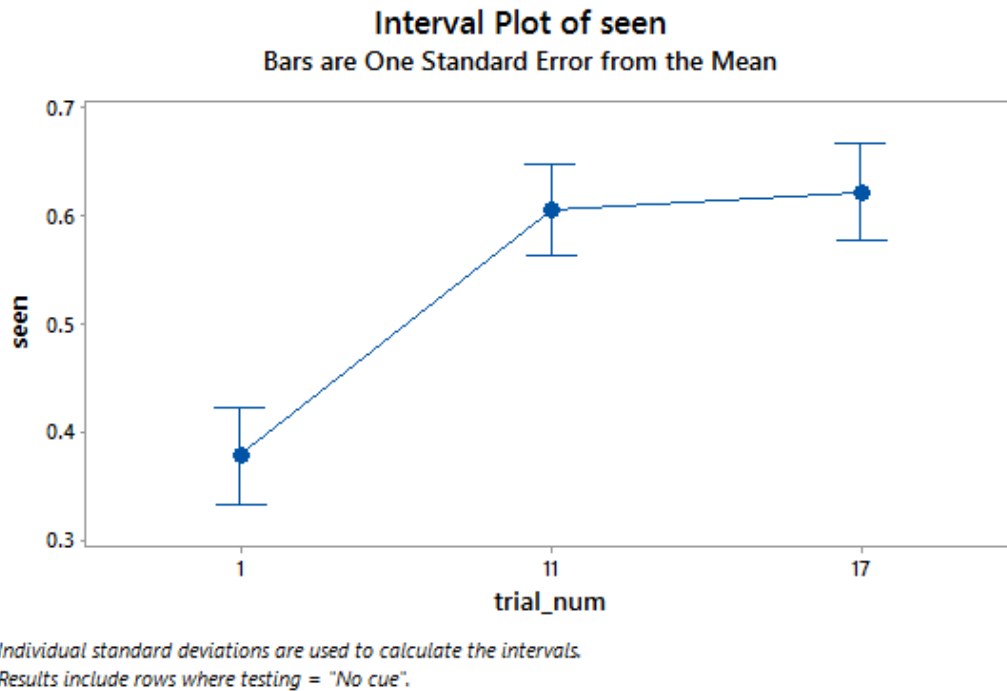


FIGURE 5.1: Average gestures correctly seen with no attention guidance

5.2 Trials with Saturation Guidance

5.2.1 Results

With saturation guidance the average percentage of gestures correctly identified rose to 77% (averaged across all intensities). For the low intensity this average was raised to 62%, medium intensity raised it to 78% and high intensity raised it to 89% (see figure 5.5). Users noticed gestures on average 1.5 seconds after the onset of a saturation cue.

Along with this data participants provided some feedback on how noticeable the saturation guidance cues were. 30% of participants found the saturation cue very

obvious at higher levels and even distracting. At least one participant admitted to looking for the saturation cue instead of the gesture.

5.2.2 Evaluation

Saturation increased performance of the visual search task with on average 22% more cues being seen. At the highest intensity there was an increase of 34% of the gestures correctly identified. The results again indicate some learning between the earlier and later trials. This can be seen between trials 7 and 13 and also between trials 4 and 12 in figure 5.2. However the later trials still show an improvement in performance when compared to the later trials with no cue. Users were 14% faster at noticing the gestures with saturation guidance. These improvements in performance indicate that the participants were noticing the gestures with more ease. This indicates that the saturation was effective at guiding their gaze to the correct gesture.

From the feedback given it would appear that saturation does not fit naturally into the scene. This may be due to the entire characters colour altering while surrounded by identical characters with no colour change. This colour change would appear very obvious with surrounding characters being there to provide comparison.

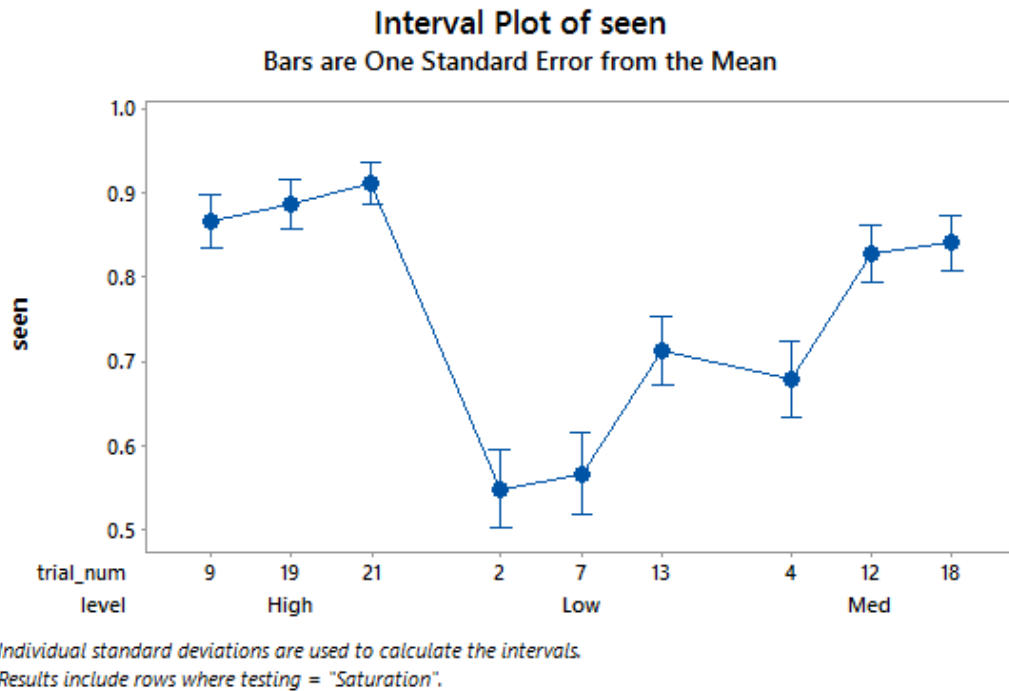


FIGURE 5.2: Average gestures correctly seen with saturation guidance cues

5.3 Trials with Modulation Guidance

5.3.1 Results

The results from the trials with modulation cues applied also show a clear increase in performance of the visual search task. This is shown with on average 77% of gestures being correctly noticed (averaged across all intensities). This figure is approximately the same as the trials with saturation guidance. For low, medium and high intensity, 69%, 77% and 88% of gestures were correctly identified respectively (see figure 5.5). Users noticed the gesture on average 1.57 seconds after the onset of the visual cue. Participants described the modulating sphere as a purple hue or halo when they

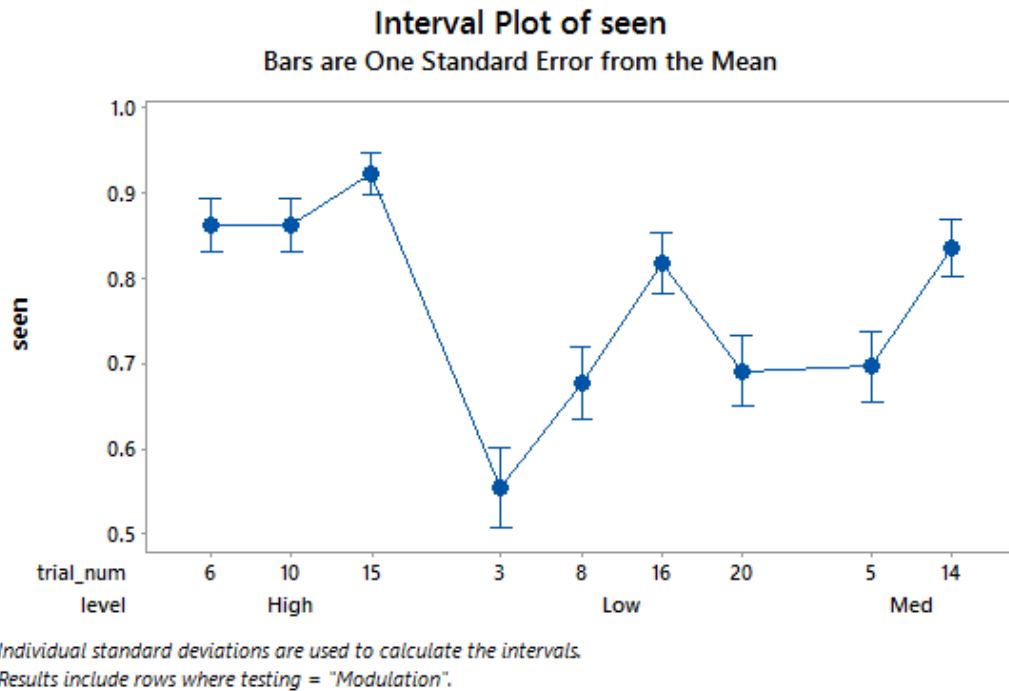


FIGURE 5.3: **Average gestures correctly seen with modulation guidance cues**

noticed it. Another comment was that it was most noticeable at a distance from the camera. 30% of participants also commented that the modulation was more subtle and less distracting than saturation.

5.3.2 Evaluation

The modulation cue resulted in the same performance of the visual search task on average as the saturation cue. The modulation cue appears to have a slower reaction time. This is only by 0.07 seconds which is not significant in terms of cognitive reaction time. As was seen in the figures 5.1 and 5.2 there is an indication

of learning towards later trials with a large increase from trial 3 to trial 16 on low intensity(see figure 5.3).

5.4 Comparison

Figure 5.5 and figure 5.4 indicate very similar performance for modulation and saturation at guiding attention. 30% of participants claimed the modulation cue was more subtle than the saturation cue. The same number claimed saturation was very obvious and found themselves looking for the cue rather than the gesture. Only one participant found saturation to be more subtle than modulation. From this user feedback, modulation was the more subtle of the two techniques. For this reason it appears that modulation is the better cue for guiding attention in a subtle manner. This is expected as it is based on the visual cue from Subtle Gaze Direction which was developed for subtlety.

The location of the target on screen was recorded when the participant noticed them. This was stored in the form of 2D normalised coordinates of the screen space. From this we calculated the distance from the target to the centre of the screen when the character pressed the key to indicate they noticed a gesture. A histogram of these distances is shown in figure 5.6. As can be seen in the figure the majority of the targets were identified close to the centre of the screen. This could be due to users turning their head towards the target and to place them in the centre of their view.

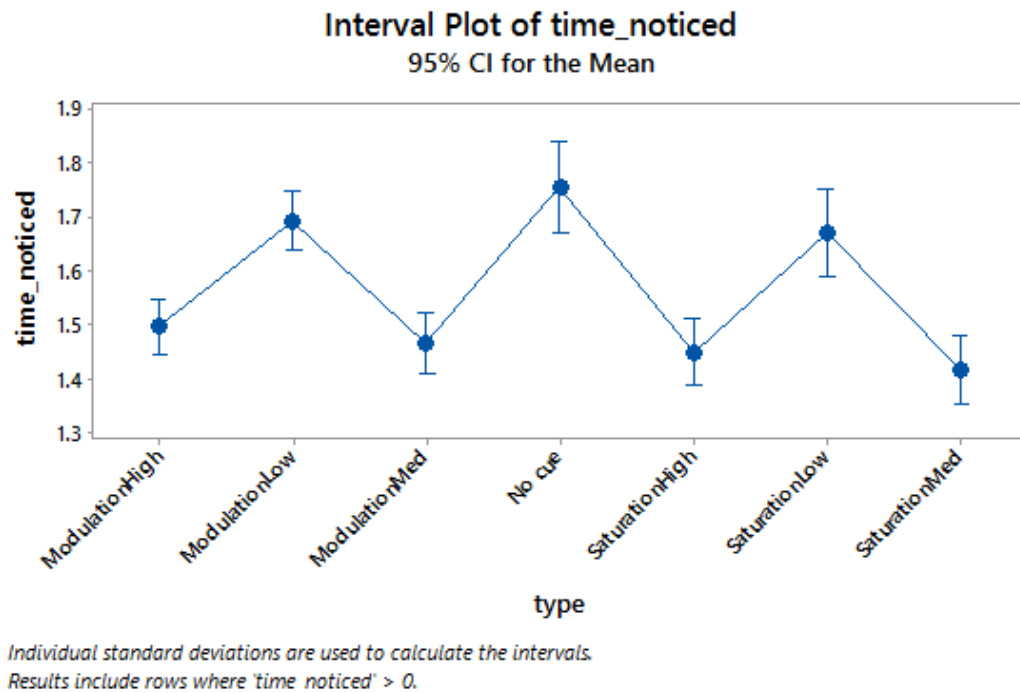
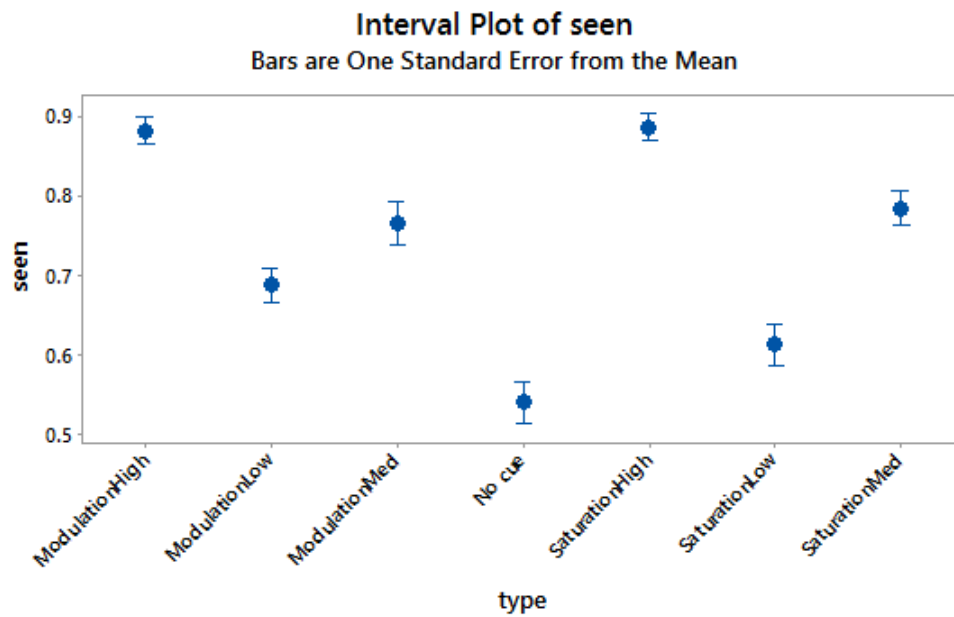


FIGURE 5.4: **Average time taken to notice (seconds) gesture after onset of visual cue**

This could merit further investigation as it could be a means of predicting where a user is focusing their gaze in a virtual reality environment without eye tracking.



Individual standard deviations are used to calculate the intervals.

FIGURE 5.5: Average gestures correctly seen comparison

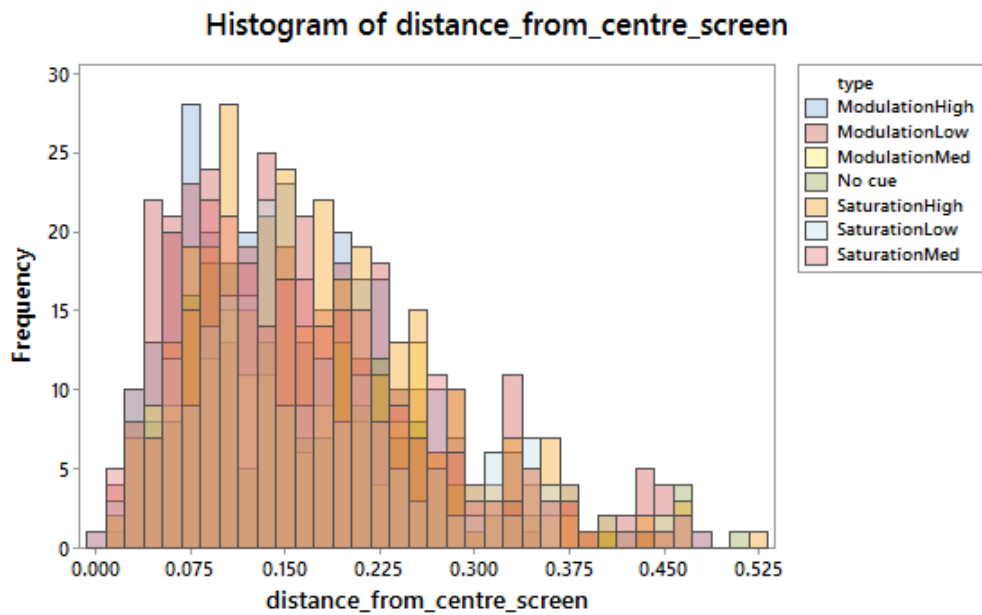


FIGURE 5.6: Histogram of the identified targets distance from the centre of the screen (normalised coordinates)

5.5 SaliencyToolbox

The SaliencyToolbox is based on the work outlined in section 2.3. It works by locating the most salient locations in an image and visiting them in decreasing order. Screenshots of the scene from the trials with visual cues were used as input for this application. We can see in the following figures that the most salient location chosen in the images is the location where attention guidance has been applied. The other locations in the image that stand out can also be seen from the saliency map.

5.5.1 No guidance

These results show the location that is considered most salient with no attention guidance. This serves as a control for the other images.



FIGURE 5.7: **First location that draws attention in image with no cue**

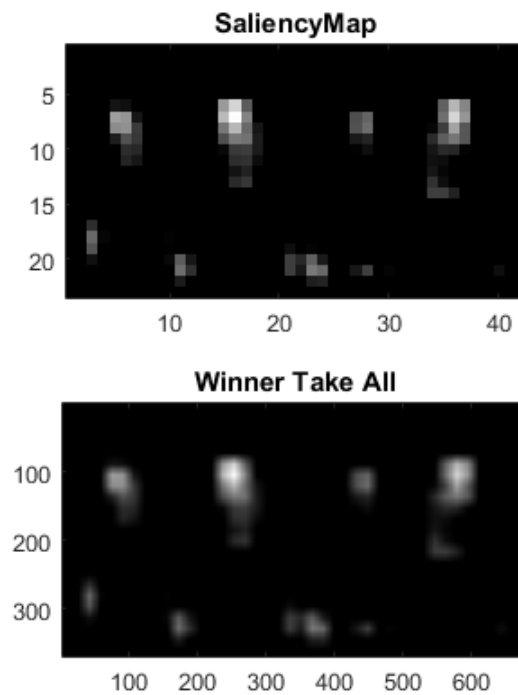


FIGURE 5.8: The saliency map produced by figure 5.7 - most salient spots are brightest spots

5.5.2 Saturation

The following figures are the results of a screenshot from one of the high saturation trials. We can see attention has been shifted to the saturated character. The saliency maps show that the most salient areas are locations where skin or a face is showing. The brightest areas are the ones that gaze will be directed to initially with the next brightest location being chosen next.



FIGURE 5.9: First location that draws attention in image (using saturation guidance)

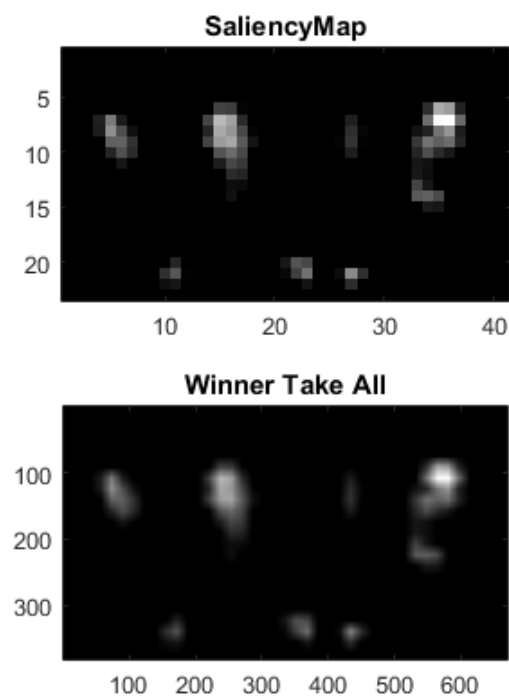


FIGURE 5.10: The saliency map produced by figure 5.9 - most salient spots are brightest spots

5.5.3 Modulation

The following figures are the results of a screenshot from one of the medium modulation trials. Again we can see that attention is shifted to the character with the visual cue.



FIGURE 5.11: First location that draws attention in image (using modulation guidance)

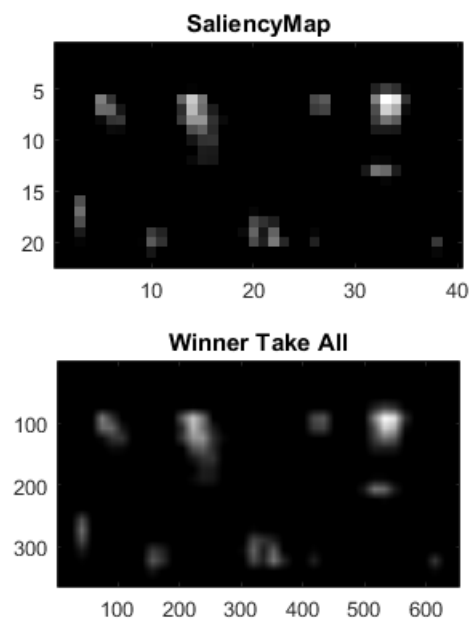


FIGURE 5.12: The saliency map produced by figure 5.11 - most salient spots are brightest spots

Chapter 6

Conclusions

This dissertation introduced attention guidance in an immersive 3D virtual reality. It has detailed how the Subtle Gaze Direction technique was adapted from previous work (see section 3.2). The modulation visual cue was developed based on the visual cue used in that work. A saturation visual cue was also developed for emphasising parts of the scene as had been previously done for emphasising locations in illustrations (see section 3.5).

A user study for measuring performance in a simple visual search task was conducted to determine how effective the guidance cues were. The task in this user study was purposely designed to be very difficult to perform without attention guidance. This user study also served as a way to obtain user feedback on how subtle they found these cues. A cue is considered subtle if it can go unnoticed by the user or does not degrade their viewing experience. The results show that the attention guidance

improved the performance of the task in both modulation guidance and saturation guidance. The participants also noticed the cues meaning subtlety was not achieved.

6.1 Modulation

The modulation technique based on SGD (see section 3.2) was implemented as one of our attention guidance cues. In order to adapt this into a 3D virtual reality environment a transparent sphere was used that appeared as a halo around locations where attention was needed to be drawn.

From the results of the study we can conclude that this technique was effective at guiding attention. It improved the overall performance of a visual search task. Participants also rated this as more subtle than the saturation technique.

6.2 Saturation

The saturation technique was based on emphasising locations using colour as described in section 3.5. Saturation used in this work was not conventional saturation but an increase in saturation and brightness to allow it to more naturally fit into the scene.

A similar increase in performance was found when using the saturation cue as was seen with the modulation cue. Participants commented that this cue appeared more

unnatural and distracting. We surmise that this is due to the character standing out unnaturally from the other identical characters. It was more obvious that there is a change in the appearance of the character in this case. It would be interesting to investigate if the same observations were made if the scene consisted of distinct characters, as opposed to identical characters.

6.3 Overall Conclusions

We conclude that both these visual cues are effective at guiding attention. The increase in performance of the visual search task with attention guidance indicate that the cues allowed participants to find the targets with more ease. The results from the SaliencyToolbox also indicate that the guidance cues can alter the natural gaze pattern and draw attention to the location of our visual cue.

We conclude that although both techniques provide similar results for a visual search task, the modulation technique is more suitable for guiding attention in a virtual environment. We come to this conclusion due to the user feedback that was received. There is also an indication of a learning effect indicated by the user study results that could be mitigated against with training in future iterations. This training could include exposing a user to the search task (with no data collection) until they are comfortable with their performance. The study was designed to allow a participant to be exposed to each gesture at all intensities. Users initially did not notice cues at the lower intensity. It appeared once they were exposed to the higher

intensities they were more aware of the visual cues. To investigate this further a similar set of experiments could be conducted where participants are separated into groups ensuring they are only exposed to certain intensities, a between-subjects experiment.

Chapter 7

Future Work

This section presents some insights into what future work could potentially be completed.

Previous implementations of Subtle Gaze Direction have relied on eye tracking for determining where the viewer is currently focused. This allowed them to take advantage of the way humans process visual stimulus (periphery has low visual acuity, fovea has high visual acuity). With this system it was easy to start their visual cue in a location in the viewer's periphery vision and terminate it before they could properly focus on it. This allowed their gaze direction to be subtle and the viewer to not realise what had caught their attention.

With the techniques developed in this work we were unable to use eye tracking due to the constraints of the Oculus Rift™. The Oculus Rift™ fully covers a user's face which prevents conventional eye tracking from being used. If in future any HMDs



FIGURE 7.1: Scene used for user study in [BSM⁺13]

include eye tracking it would be interesting to implement a similar experiment where cues would only be triggered in the viewer's periphery. There could also be useful data to observe from an eye tracker in this kind of environment. Currently the Oculus Rift™ has no eye tracker. Based on observations made from the results there are other methods for determining where a user is looking that may prove useful in the absence of eye tracking. The results outlined in this work indicate that most participants rotated their heads so that the target was in the centre of the screen before pressing their key. At this point it would appear the user's gaze is in the centre of the screen and possible cues could be triggered towards the edges. This would not be 100% accurate but could produce some useful results.

With more time it would be interesting to conduct another user study with a different setup. Consider figure 7.1. This scene was used in the user study as described in section 3.4 to guide the viewers along a sequence of objects. A similar experiment could be created in a virtual environment for the Oculus™. Without eye tracking participants could be asked to centre the object they are looking at in the screen.

A crosshair could be used for this purpose although this would take away from the photorealism of the scene.

Although these methods could be tested for effectiveness at determining gaze in virtual reality the optimal solution would be to use an eye tracker. Currently Senno-Motoric Instruments (SMI) is offering an eye tracking package for the Oculus™[\[Eye\]](#). For future work the most efficient method for implementing this would either use this package or another HMD with eye tracking capabilities.

Appendix A

Experiment Documents

This appendix consists of the documents provided to the experiment along with the documents required for ethical approval. They are as follows:

1. Informed Consent Form
2. Information Sheet for Prospective Participants
3. Research Proposal

TRINITY COLLEGE DUBLIN

INFORMED CONSENT FORM

LEAD RESEARCHERS: SARAH NOLAN

BACKGROUND OF RESEARCH: This research is being conducted in order to develop methods for guiding attention in virtual reality in a subtle way. The data from these experiments will be used in order to determine how strong or subtle the attention guidance cues are inside a virtual reality scene.

PROCEDURES OF THIS STUDY: Participants will be asked to wear an Oculus Rift, a virtual reality headset, for the experiment. During the immersion the participant will sit through consecutive trials each 1 minute long. During these trials they are asked to press a key when they notice a certain visual event (gestures by virtual characters in the scene). They will also be advised at the start of the experiment to let the researcher know if they notice any anomalies after each trial. At the end of the experiment is encouraged to give open feedback about the experience. The experiment will not exceed 30 minutes.

CYBER SICKNESS: Cyber sickness can occur when using Virtual Reality simulators for prolonged periods. Over 80% of individuals exposed to VR simulations of 20 minutes reported increases in sickness symptoms which are similar in nature to motion sickness. Factors that influence this sickness include stress on the visual system. Poor illumination, poor control and an unusually close working distance can cause cyber sickness symptoms in the user. Overexposure to VR could result in a decrease in visual acuity following immersion in a virtual world. There have been studies done exposing subjects to immersion in a virtual environment resulting in 60% reporting symptoms of eyestrain, headache and nausea and 20% reporting a reduction in binocular visual acuity. The presence of perceived control can reduce the averseness of a situation. Participants will have full control of the view through their head movements which are captured by the Oculus Rift and no other motions of the view are created. They will only be exposed to short doses of the virtual environment to reduce any symptoms. The device will be calibrated per subject to reduce the effects of cyber sickness, their height and visual acuity will determine the best setup for them. Due to the risk of cyber sickness the experiments will be limited to 5-10 minutes per virtual immersion with appropriate rest times of 2 minutes in between. If at any point the participant experiences any symptoms of cyber sickness they should let the researcher know and remove the headset. The experiment will be paused until the symptoms ease or stopped if they participant wishes to withdraw without penalty as state above. In the unlikely case that it is required medical assistance will be put in place.

PUBLICATION: All data will be held in accordance with the Data Protection Act. Data will be anonymized and the aggregated data will be analyzed using statistical methods to prove or disprove my hypothesis for attention guidance.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that my participation is fully anonymous and that no personal details about me will be recorded.
- I do not have epilepsy.
- I have no objection data is published in scientific publications in a way that does not reveal my identity.
- I have received a copy of this agreement.

PARTICIPANT'S NAME:

PARTICIPANT'S SIGNATURE:

Date:

Statement of investigator's responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully

answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

RESEARCHERCONTACTDETAILS: Sarah Nolan (nolans4@tcd.ie)

INVESTIGATOR'S SIGNATURE:

Date:

TRINITY COLLEGE DUBLIN

INFORMATION SHEET FOR PROSPECTIVE PARTICIPANTS

- The below sheet outlines the procedure of the experiment and risks to the participant.
- This research is being conducted to measure attention guidance and in a virtual 3D scene. Previous research has been done into guiding attention in a 2D space (images) and a 3D scene. This experiment builds on this research.
- As part of this study the participant will be asked to wear a virtual reality headset, the Oculus Rift. They will be immersed in a virtual environment and asked to press a key when they notice certain gestures of characters in the scene. This will be repeated for multiple trials of one minute each. The participant is also asked to note after any trial any anomalies they may have noticed. At the end of the experiment the participant is encouraged to ask any questions or give open feedback about the experience.
- All data collected is anonymized and used for research purposes only.
- No audio or video recordings will be taken during the course of the experiment.
- Participation is completely voluntary and a participant can withdraw at any time without penalty.
- People with epilepsy should not take part in this experiment.
- In the unlikely event that illicit activities are reported during the study, the researcher will be obliged to report this to the relevant authorities
- **Cyber Sickness:** Cyber sickness can occur when using Virtual Reality simulators for prolonged periods. Over 80% of individuals exposed to VR simulations of 20 minutes reported increases in sickness symptoms which are similar in nature to motion sickness. Factors that influence this sickness include stress on the visual system. Poor illumination, poor control and an unusually close working distance can cause cyber sickness symptoms in the user. Overexposure to VR could result in a decrease in visual acuity following immersion in a virtual world. There have been studies done exposing subjects to immersion in a virtual environment resulting in 60% reporting symptoms of eyestrain, headache and nausea and 20% reporting a reduction in binocular visual acuity. The presence of perceived control can reduce the averseness of a situation. Participants will have full control of the view through their head movements which are captured by the Oculus Rift and no other motions of the view are created. They will only be exposed to short doses of the virtual environment to reduce any symptoms. The device will be calibrated per subject to reduce the effects of cyber sickness, their height and visual acuity will determine the best setup for them. In the unlikely case that it is required medical assistance will be put in place.
- If at any point the participant experiences any symptoms of cyber sickness they should let the researcher know and remove the headset. The experiment will be paused until the symptoms ease or stopped if they participant wishes to withdraw without penalty as state above. Rest periods of 2 minutes every 5-10 minutes will be given to limit the length of immersion. Thus, we do not expect subjects to experience any negative

symptoms for the short duration.

- **Relevant Legislation:** This study complies with the Data Protection Act.
- **Declaration of Conflict of Interest:** It is acknowledged that some participants are being recruited for this study from amongst the researcher's university colleagues and acquaintances.

Research Project Proposal - "Directing Attention in an Immersive Virtual Reality"

To whom it may concern,

I am applying for permission to run a user study as part of my ICS masters dissertation. My project is the investigation of techniques for subtly guiding attention in a photorealistic 3D virtual reality scene.

The study will involve the participant wearing a virtual reality headset and pressing a key when they notice behaviours (nods or a mixture of gestures) of characters in a scene for a series of one minute trials. They will also be asked at the beginning of the experiment to note after each trial if they noticed anything out of place. Cues that will be used to help guide the participant to notice these behaviours are a halo around a character that is modulated to draw attention or by increasing the saturation of the colours on the character. These are altered for each trial. At the end of the experiment the participant will be invited to provide open feedback on their experience - how subtle the guidance was, or how easy it was to notice the gestures. A full experiment should last no longer than 30 minutes including rest periods.

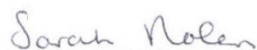
The data from these experiments will be used to prove how strong or subtle the attention guidance cues I have developed are. The aim is to find a subtle cue that still guides the user's attention effectively.

The participants will be invited through class list(s) with previous permission of the list owner. They may also include acquaintances or classmates in which case I will note a conflict of interest. It is envisaged that 10-15 participants will be sufficient for this study based on similar experiments conducted in previous published work by McNamara et al [1]. No participant will be under 18 years of age. Participants can opt out at any time without penalty.

There is sometimes a risk of cyber sickness with virtual immersion for prolonged periods. The participants will be informed of this risk along with any symptoms that it causes. These symptoms along with the precautions that will be taken are outlined in a separate document provided with this one. Due to this risk the participant will be given a rest of 1-2 minutes every 5-10 minutes which should eliminate the chances of these symptoms. Participants can opt out at any time without penalty and their responses deleted at any time if they so request. Due to the nature of this experiment it is advised that subjects with epilepsy do not participate.

Data collected from these trials will be anonymised and only trends will be reported in the dissertation. All data will be held in accordance with the Data Protection Act. Participants will not be recorded in any way as part of the experiment.

Thanks for your time,



Student

[1] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm, "Subtle Gaze Direction", ACM Transactions on Graphics, Vol. 28, No. 4, August 2009.

Appendix B

Code Description

This appendix gives an overview of the different scripts used for the user study as well as a description of the functions they contained.

B.1 Trigger Script

This script is attached to the player game object. It is the central script as it triggers animations and visual cues on characters in the scene. This script consists of the following functions:

B.1.1 Start

Initialises variables such as the local reference to the TrialHandler to get the trial number and what visual cue and intensity is being tested for that trial. Seeds the

Random for the pseudo random values for the trial with this trial number.

B.1.2 isVisible

The screen-space normalised position is calculated using the camera function `WorldToViewportPoint` of a target passed in. If this value is between `[0,0]` and `[1,1]` it is considered visible and `true` is returned. Otherwise `false` is returned.

B.1.3 OnTriggerEnter

Handles the event of an object entering the bounding box. This function will return immediately if the trial is flagged as paused. Checks if the object is tagged as 'man'. If so the function `isVisible` is used to determine if on screen. If not returns. If so performs ray intersection tests for both sides of face. If both on screen the reference to the object to a queue of objects. If the cue was empty a start time to trigger a gesture on a character is initialised.

B.1.4 pauseTrigger

Stops the current gesture if it exists. Flags the trial to be paused.

B.1.5 OnTriggerExit

When an object exits the bounding box this function is called. The function will return immediately if the trial is flagged as paused. If the object is tagged as “man” and not currently performing a gesture it is removed from the queue. If it is performing a gesture (the current target is referencing this object) then stopCue is called with the object. If the object was not seen logSeen is called with the object. A time to start the next gesture is set and the object is removed from the queue.

B.1.6 spawnSphere

This function is used to render a new modulating sphere at the location passed. The sphere is instantiated and it’s duration and intensity values are set.

B.1.7 logSeen

This function writes the data for a gesture to the user study data file. The screen-space normalised position is calculated using the camera function WorldToViewportPoint. This position is then used to calculate the distance from the center of the screen in normalised coordinates (center of the screen is at [0.5,0.5]). If seen has been set to 1 then the time noticed is set to the time elapsed since the gesture was started. The distance of the target from the player is also calculated. Other data to be written out includes the type of gesture, the visual cue and the cue intensity. The TrialHandler function printOut is called with this data.

B.1.8 saturate

This function calls the `updateSaturation` function on the `ChangeSaturation` script attached to a character.

B.1.9 startCue

This function starts a character animation and the visual cue if attention guidance is applied. If attention guidance is applied the `saturate` function is called in the case of saturation and the `spawnSphere` function is called in the case of modulation. The `currentCue` is set to a reference of the object passed to the function.

B.1.10 Update

This function is called every frame. If the trial is flagged as paused it is returned from immediately. Most of the functionality is handled in this function. It functions as a state machine. If the duration of the gesture has elapsed the `stopCue` function is called on the current cue and a flag is set to allow the participant an extra second to press the key if they noticed the gesture. If there are objects in the queue and the time for another cue to begin has elapsed a random character in the queue is chosen to perform the next gesture. The `currentCue` value is set to a reference of this character and it is removed from the queue. The start cue is called on the object the `startNextCue` value is reset. If a key has been pressed while a gesture is

being performed the seen flag is set to 1 and the logSeen function is called with the currentCue. If no key was pressed and the extra second time has elapsed the seen flag is set to 0 and logSeen is called on the currentCue. The currentCue is then set to null. If the trial is flagged as paused the trial handler is called to end the trial. This allows time for the gesture to finish before ending a trial.

B.2 FaceUser

This script is attached to each character to allow them to rotate to face the player at all times.

B.2.1 Start

Initialises variables such as a reference to the player game object. It also gets a reference to the character's animator and uses a random generated value to start each characters idle animation at a different point. This made the character's movements more distinct.

B.2.2 FixedUpdate

This function gets the direction of the target to the player by subtracting the character's position from the user's position. The quaternion of this direction is calculated using Quaternion.LookRotation on it. Using simple linear interpolation between

the characters current rotation and this target rotation the character is rotated over time to face the user.

B.3 TrialHandler

This script is attached to the player class to provide the main trial values such as the visual cue and intensity for each trial hardcoded as a string. This string is 42 characters long with each 2 characters belonging to one trial. The first character encodes the saturation intensity and the second character encodes the modulation intensity. These values go from 0 (no intensity, no cue) to 3 (high intensity).

B.3.1 Awake

The awake function allows the TrialHandler to be initialised with certain values immediately. The trial name is retrieved to determine if it is a trial or one of the two start scenes. No data is recorded during the start scenes. The trial number is also retrieved from the trial name if it is a trial. The start time variable is then set to the current time at the end of the function.

B.3.2 getTrialNum

This function returns the trial number. The trigger calls this function.

B.3.3 getSaturationLevel

Returns the saturation intensity level by using the trial number to index into the string.

B.3.4 getModulationLevel

Returns the modulation intensity level by using the trial number to index into the string.

B.3.5 endTrial

Calls the LevelEnd function on the OVRMainMenu (a script provided with the Oculus Plugin).

B.3.6 Update

If the trial duration has elapsed and it has not already been flagged to end the end flag is set to true and the pauseTrigger function on the Trigger is called.

B.3.7 printOut

This function writes all the data passed from the Trigger to the user study data file. The data is parsed into a string with each entry separated by semi-colons and terminated with an endline.

B.4 ChangeSaturation

This script is attached to each character to control their saturation intensity.

B.4.1 Start

Initialises variables such as the characters render and number of materials.

B.4.2 updateSaturation

Sets the new saturation, start time of the saturation cue and end time of the saturation cue.

B.4.3 Update

Uses the start time and end time of the saturation to get the half way time. This time is used to interpolate the saturation from normal saturation to the new saturation.

Once it has passed the halfway point the opposite interpolation is performed. If the full duration has elapsed the values are reset. Each of the materials have their saturate factor input updated in their saturation shader to the new interpolated value each frame.

B.4.4 Modulation

This script is attached to the modulating sphere to control the warm-cool modulation and fading in/fading out.

B.4.5 Start

Initialises variables such as `colorTo`, `colorFrom`, `transparency` and `startTime`.

B.4.6 setTimes

Allows start time and end time values to be set publicly. The midway point is calculated from these times.

B.4.7 Update

This function performs the fading in and out of the spheres intensity and transparency. While the time is less than the midway point the transparency is interpolated to an opacity determined by the spheres intensity. The saturation of the

sphere is also determined from this value. If the time has passed this midway point the sphere opacity is interpolated back from this value to 0. This function also interpolates from the current colour to the toColor. Finally the alpha value of this colour is set to the current transparency determined earlier in the function. This colour is then set as the TintColor input along with the saturation value to the Modulation shader.

Bibliography

- [BMSG09] Reynold J. Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm, *Subtle gaze direction*, ACM Trans. Graph. **28** (2009), no. 4.
- [BSM⁺13] Thomas Booth, Srinivas Sridharan, Ann McNamara, Cindy Grimm, and Reynold J. Bailey, *Guiding attention in controlled real-world environments*, ACM Symposium on Applied Perception 2013, SAP' 13, Dublin, Ireland, August 22-23, 2013, 2013, pp. 75–82.
- [CDF⁺06] Forrester Cole, Douglas DeCarlo, Adam Finkelstein, Kenrick Kin, R. Keith Morley, and Anthony Santella, *Directing gaze in 3d models with stylized focus*, Proceedings of the Eurographics Symposium on Rendering Techniques, Nicosia, Cyprus, 2006, 2006, pp. 377–387.
- [Eye] *Eye tracking hmd upgrade package for oculus rift dk2*, <http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/eye-tracking-hmd-upgrade.html>, Accessed: 09/05/15.

- [Fab] *Three hours with the finished fable 2*, <http://www.engadget.com/2008/09/15/three-hours-with-the-finished-fable-2/>, Accessed: 01/05/15.
- [IKN98] Laurent Itti, Christof Koch, and Ernst Niebur, *A model of saliency-based visual attention for rapid scene analysis*, IEEE Trans. Pattern Anal. Mach. Intell. **20** (1998), no. 11, 1254–1259.
- [MBG08] Ann McNamara, Reynold J. Bailey, and Cindy Grimm, *Improving search task performance using subtle gaze direction*, Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization, APGV 2008, Los Angeles, California, USA, August 9-10, 2008, 2008, pp. 51–56.
- [MBS⁺12] Ann McNamara, Thomas Booth, Srinivas Sridharan, Stephen Caffey, Cindy Grimm, and Reynold J. Bailey, *Directing gaze in narrative art*, ACM Symposium on Applied Perception 2012, SAP '12, Los Angeles, CA, USA - August 03 - 04, 2012, 2012, pp. 63–70.
- [ME2] *Combat (mass effect 2)*, [http://masseffect.wikia.com/wiki/Combat_\(Mass_Effect_2\)](http://masseffect.wikia.com/wiki/Combat_(Mass_Effect_2)), Accessed: 15/05/15.
- [Mil] *Chris milk ted talk: How virtual reality can create the ultimate empathy machine*, https://www.ted.com/talks/chris_milk_how_virtual_reality_can_create_the_ultimate_empathy_machine, Accessed: 01/05/15.

- [ocu] *Oculus rift development kit 2*, <https://www.oculus.com/dk2/>, Accessed: 01/05/15.
- [PPW97] Randy F. Pausch, Dennis Proffitt, and George H. Williams, *Quantifying immersion in virtual reality*, SIGGRAPH, 1997, pp. 13–18.
- [Rot82] S. D. Roth, *Raycasting for modeling solids*.
- [TG80] Anne M. Treisman and Garry Gelade, *A feature-integration theory of attention*, COGNITIVE PSYCHOLOGY 12, 1980, pp. 97–136.
- [Unia] *How unity3d became a game-development beast*, <http://insights.dice.com/2013/06/03/how-unity3d-become-a-game-development-beast/>, Accessed: 10/05/15.
- [Unib] *Unity*, <https://unity3d.com/>, Accessed: 10/05/15.
- [Vla] *Water flow in portal 2*, http://www.valvesoftware.com/publications/2010/siggraph2010_vlachos_waterflow.pdf, Accessed: 01/05/15.
- [WK06] Dirk Walther and Christof Koch, *Modeling attention to salient proto-objects*, Neural Networks 19 (2006), no. 9, 1395–1407.
- [WS98] Bob G. Witmer and Michael J. Singer, *Measuring presence in virtual environments: A presence questionnaire*, Presence 7 (1998), no. 3, 225–240.

[zel] *The legend of zelda: Wind waker*, http://www.nintendolife.com/games/gamecube/legend_of_zelda_the_wind_waker/screenshots,

Accessed: 15/05/15.

Glossary

FBX FBX (Filmbox) is a proprietary file format (.fbx) developed by Kaydara and owned by Autodesk since 2006. It is used to provide interoperability between digital content creation applications. [43](#)

focal plane a plane that is perpendicular to the axis of a lens or mirror and passes through the focus. [36](#)

HLS a color model that defines colors by the three parameters hue (H), lightness (L), and saturation (S). [40](#)

HMD a head-mounted display is a display device, worn on the head, that has a small display optic in front of one (monocular HMD) or each eye (binocular HMD). [7](#)

modulation the addition of information (or the signal) to an electronic or optical signal carrier. [36](#)

RGB the RGB color model is an additive color model in which red, green, and blue light are added together in various ways to reproduce a broad array of colors.

The name of the model comes from the initials of the three additive primary colors, red, green, and blue. [39](#)

saccade a rapid movement of the eye between fixation points. [22](#)

SDK Software development kit. A software development kit (SDK or "devkit") is typically a set of software development tools that allows the creation of applications for a certain software package, software framework, hardware platform, computer system, video game console, operating system, or similar development platform. [41](#)

shader a program that tells a computer how to draw something in a specific and unique way. [37](#)

SIFT Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. [31](#)