

**GAZE-CONTINGENT MULTI-MODAL AND MULTI-SENSORY
APPLICATIONS**

MARGARITA VINNIKOV

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN COMPUTER SCIENCE AND ENGINEERING
YORK UNIVERSITY
TORONTO, ONTARIO

APRIL 2015

© Margarita Vinnikov, 2015

Abstract

Gaze-contingent displays are applications that are driven by the user's gaze. They are an important tool for many multi-modal and multi-sensory domains. They can be used to precisely control the retinal image in real time to study visual control of natural behaviour through experimentation, or to improve user experience in virtual reality applications. In this dissertation, I explored the application of gaze-contingent display technology to different models and senses and evaluate whether such applications can be useful for simulation, psychophysical research and human computer interaction. Specifically, I have looked at visual gaze-contingent display and an audio gaze-contingent display. I examined the effects of simulated visual defects on user's perception and control of self-motion during locomotion. I found that gaze-contingent display simulations of visual defects significantly altered visual patterns and impaired the accuracy and precision of judgement of heading. I also examined the impact of simulating gaze-contingent depth-of-field for monocular and stereoscopic displays. The experimental data showed that the alleviation of

negative effects associated with stereo displays depends on the user's age and the types of scenes that are viewed. Finally, I simulated gaze-contingent audio displays that imitated the cocktail party effect. My audio enhancement techniques turned to be very beneficial for applications that have to deal with user's attention to multiple sources of sounds (speakers) such as tele-conferences and social games. Finally, in this dissertation I demonstrated that gaze-contingent systems can be used in many aspects of virtual system design and if combined together (used for multiple cues and senses) can be a very powerful tool for augmenting and improving the overall user experience.

Acknowledgements

I would like to express the deepest appreciation and thanks to my supervisor Dr. Robert Allison. You have been a tremendous mentor for me. I would like to thank you for your guidance through my academic achievements and career choices. Thank you for your patience and support, for teaching me not to be afraid of new and complex things, for guiding by example, to be a versatile researcher, who always learns and explores new techniques and ideas, for reading and editing my work. Without your kind supervision and help, this dissertation would not have been possible.

My special thanks to thank Dr. Stephen Palmisano, who helped me tremendously with chapter 3 by allowing me to work in his lab and collaborating on my research. I am thankful for words of wisdom and encouragement through the years.

I would also like to thank my committee members, Dr. Wolfgang Stuerzlinger, Dr. Burton Ma, Dr. Melanie Baljko, and Dr. Laurie Wilcox for serving as my committee members, invaluable comments, great discussion during the defence and

being great examples of what true researcher should be.

This work would not be possible without the incredible support from all of my research assistants: Carly Hylton, Suzette Fernandes, Rida Waseem, and Anna Kiseleva and many others. Thank you for spending hours in the dark lab and doing all the boring chores. It would not be possible without support from all my lab mates: Pearl Guterman, Ramy Kirollos, Muna Shabenah and Tracy Hirlihey and every one else who worked along my side over the years.

Big Thank You goes as well to all the administrative staff (Teresa Manini, Susan Cameron and Ouma Jaipaul Gill) and technical staff (Oana Hala, Seela Balkissoon, Paul Griffith, Nam Tran, and Ulya Yigit).

Huge appreciation goes to my friends for support, encouragement and for being my participants (Irina Kapler, Yulia Mintz and Lena Utterman, Sergui Tchertok and Inna Tsirlin).

And now my family:

My grandmother for support and understanding. My dad for making me fall in love with education, being proud of my achievements and teaching me to dream big. My mom for teaching me to never give up and for unconditional love and support. My dear husband for patience, eating frozen dinners and for being proud of me.

Finally, I would like to thank the Government of Ontario for providing me with

the Ontario Graduate Scholarship in Science and Technology (OGSST) and the Government of Australia for the Endeavor Fellowship.

All the work in this dissertation is novel and has never been published except for chapters 4. Specifically, the preliminary results for chapter 4 have been published in the proceedings of ETRA 2014.

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents.....	vii
List of Tables.....	xii
List of Figures.....	xiv
Abbreviations	xviii
1 Introduction	1
1.1 Motivation	4
1.2 Goals	6
1.3 Organization	7
2 Background	9

2.1	Gaze-Driven Applications	9
2.2	Where can we use gaze-contingent display (GCD) applications? . .	9
2.2.1	Vision	10
2.2.2	Tactile and Haptic Displays	22
2.2.3	Audition	23
2.3	Implementations	25
2.3.1	Foveation	25
2.3.2	Perceptual Modelling	28
2.3.3	Rendering Techniques	30
2.3.4	Gaze-Contingent Collision Handling	46
2.3.5	Saccade-Contingent Displays	48
2.3.6	Depth and GCD	49
2.4	Conclusions	53
3	Simulating Visual Defects for Studying Heading and Steering	55
3.1	Introduction	55
3.2	Theories in Locomotion	57
3.2.1	Optic Flow and Heading	58
3.2.2	Egocentric Direction cues and Steering points	59
3.2.3	Gaze and Steering Behaviour	61

3.3	Gaze-Contingent System Implementation	66
3.3.1	Platform	66
3.3.2	Apparatus	66
3.3.3	Defect Simulation	67
3.4	Use of the System in the Study of Steering Behaviour	68
3.5	Heading Experiments	70
3.5.1	General Methodology	73
3.5.2	Experiment 1	74
3.5.3	Experiment 2	89
3.6	Conclusions	117
4	Gaze-Contingent Depth of Field Simulation	118
4.1	Introduction	118
4.2	System Concept	127
4.2.1	Apparatus	130
4.2.2	Calibration	131
4.3	Depth Impressions, Visual Comfort and Image Quality	132
4.3.1	Stimuli	133
4.3.2	Task	135
4.3.3	Participants	136

4.3.4	Results	137
4.3.5	Discussion	147
4.4	Long Term Comfort	152
4.4.1	Evaluation	156
4.4.2	Results	161
4.4.3	Discussion	165
4.5	General Discussion	169
5	Audio Gaze-Contingent Display	175
5.1	Introduction	175
5.2	Previous Studies	179
5.2.1	Cocktail Party Effect	180
5.2.2	Gaze-Contingent or Selective displays	180
5.3	General Methodology and Experimental Approach	183
5.3.1	System Setup	184
5.3.2	Apparatus	186
5.4	Experiment 1	187
5.4.1	Methods	188
5.4.2	Results	197
5.4.3	Discussion	214

5.5	Experiment 2	223
5.5.1	Methods	224
5.5.2	Results	228
5.5.3	Discussion	234
5.6	General Discussion	236
5.7	Conclusions	238
6	Conclusions	240
6.1	Heading and driving	240
6.1.1	Summary	240
6.1.2	Lessons Learned and Future Studies	241
6.2	Depth-of-Field Simulations	242
6.2.1	Summary	242
6.2.2	Lessons Learned and Future Studies	243
6.3	Cocktail Party	244
6.3.1	Summary	244
6.3.2	Lessons Learned and Future Studies	245
6.4	Experimental Sample Size	246
6.5	Final Words	247
	Bibliography	249

List of Tables

3.1	Repeated measures ANOVA for viewing conditions	101
3.2	Two-way repeated measures ANOVA for ROI as function of viewing modes	105
3.3	Two-way repeated measures ANOVA for defect conditions . . .	107
3.4	Two-way repeated measures ANOVA for region of interest (ROI) as function of defect conditions	112
4.1	Estimated weights for depth preferences	140
4.2	Estimated weights for comfort preferences	143
4.3	Estimated weights for image quality preferences	146
4.4	Measures of visual and oculomotor functions	163
4.5	Kruskal-Wallis Chi-square for pre-test and post-test responses	166
4.6	Kruskal-Wallis Chi-square for Stereo and DOF+Stereo	168
5.1	Post-block questionnaire	197

5.2	Repeated measures ANOVA for number of avatars and visual behaviour	211
5.3	Repeated measures ANOVA for technique and visual behaviour	211
5.4	Estimated weights for Q1 preferences	232
5.5	Estimated weights for Q2 preferences	234

List of Figures

1.1	Interactive applications	3
2.1	VDR resolution	27
2.2	Gaze-contingent display	31
2.3	Visual effects	35
3.1	Bird's eye view of steering points and visual regions used for steering	60
3.2	Scenes and visual defects	69
3.3	Example of driving and gaze behaviour	70
3.4	Visual displays for experiment 1	77
3.5	Discrimination thresholds for each participant	81
3.6	Bias for each participant	84
3.7	Frequencies of large heading thresholds	87
3.8	Scenes used in Experiment 2	93

3.9	Frequency of large heading thresholds and of large directional bias	98
3.10	Heatmaps for participant 1 for 3 viewing conditions	100
3.11	Gaze position for simulated heading locations during free view- ing conditions	101
3.12	Cluster characteristics for viewing conditions	102
3.13	ROI characteristics for viewing conditions	106
3.14	Heat-map for participant 1 for 4 visual defect conditions	108
3.15	Gaze locations for participant 1 for 4 visual defect conditions .	109
3.16	Cluster characteristics for visual defect conditions	110
3.17	ROI characteristics for visual defect conditions	113
3.18	Percentage of gaze clusters within a radius of 2°	114
3.19	Mean slope value across visual defect conditions	115
4.1	Depth of field	120
4.2	Accommodation-vergence conflict	123
4.3	System architecture	127
4.4	Experimental setup	131
4.5	Experimental scenes	134
4.6	Depth cue administration	135
4.7	Depth preferences	138

4.8	Depth preference for the 3 scenes	140
4.9	Comfort preferences	141
4.10	Comfort preference for the 3 scenes	144
4.11	Image quality preferencse	145
4.12	Image quality preference for the 3 scenes	147
4.13	Chess board	153
4.14	Near point vision test	162
4.15	Randot stereo test	162
4.16	Subjective ratings results	167
5.1	System’s architecture	185
5.2	Experimental Setup	187
5.3	7 speakers example	191
5.4	Keyword detection time	200
5.5	Spatial perception ratings	202
5.6	Visual perception ratings	204
5.7	Audio attention ratings	205
5.8	‘Action-reaction’ ratings	207
5.9	Overall experience ratings	209
5.10	Fixation duration	212
5.11	Total time spent looking at the selected speaker	213

5.12	Fixation was redirected from speaker to another	214
5.13	Longest duration fixation on a speaker	215
5.14	Procedure for experiment 2	225
5.15	Percentage correct for speaker identification	230
5.16	Mean preference	233

Abbreviations

AMD	-	Age related macular degeneration
AOI	-	Area of interest
AR	-	Augment reality
CoC	-	Circle of confusion
CP	-	Cocktail party
DOF	-	Depth-of-field
FBO	-	Frame buffer object
FoE	-	Focus of expansion
FOV	-	Field of view
GCD	-	Gaze-contingent display
GPU	-	Graphics processing unit
HCI	-	Human computer interaction
HRTF	-	Head-related transfer function
HMD	-	Head mounted display
LOD	-	Level of details
POR	-	Point of regard
ROI	-	Region of interest
TP	-	Tangent point
VD	-	Visual direction
VE	-	Virtual environment
VR	-	Virtual reality
VRD	-	Variable resolution display

1 Introduction

Although we rarely notice, our eyes constantly fixating, a different part of a scene several times per second. Despite being one of our most common movements and their important role in information acquisition, eye movements are seldom a consideration in human computer interaction (HCI). High acuity vision, attention and even intent and social interaction depend on, or are signaled by, gaze-behaviour. Nevertheless, there are numerous applications where responding to a user's gaze could significantly enhance user experience. For example, by providing contextual information to what is viewed by additional visual markers or by auditory cues [138]. To achieve these potential benefits, a designer must overcome some major problems, such as latency, inherent in building such as system. To evaluate these potential benefits and obstacles, I believe that it is important to consider the different domains where gaze driven systems have already been used and to identify new avenues for innovation and development.

Before addressing challenges and future avenues for gaze-contingent displays it

is very important to establish a base-line and specifically define gaze-contingent displays in the context of other systems. Gaze-driven displays are applications that alter sensory displays or the feedback provided back to the user based on the user's gaze. There are many applications that rely on the user's gaze as an input device, many falling under the broad classifications of attentive user interfaces or interactive systems. In his book, Duchowski [47] divides such interactive systems into two categories - selective and gaze-contingent displays (Figure 1.1).

The first category is often referred as a *selective system*, which treat the user's gaze as a pointing device. The most obvious example of such systems is to replace the mouse pointer with human gaze direction. In this case, if the user wants to select an object on the screen, she has to direct her gaze at a point of interest. This is the equivalent of a mouse click for selection to occur. For example, she may be required to fixate the location of interest over period of some time so that the system will understand that the intent is to select it. Consequently, such a mode of interaction requires the user to be constantly aware of the location of her gaze, as each eye movement can be interpreted by the system as an action. Eye movements are normally used to look at things and explore a scene, only more rarely to select or indicate. The problem of unintentionally selecting objects when trying to look about a display has been termed the Midas touch problem [83]. Unfortunately, there is no trivial way to separate the intentional and the information gathering

aspects of eye movements.

Gaze-contingent displays (GCD) fall into the second category identified by Duchowski [47]. GCD displays are based on a more natural interaction with the system, where the user does not have to be consciously aware of their gaze. Instead, in many cases, system designers are more interested in changing the rendered content in such a way that the user will not notice the change. For example, such system can continuously change resolution to lower resolution in the periphery. This is possible due to the non-uniform structure of the human eye.

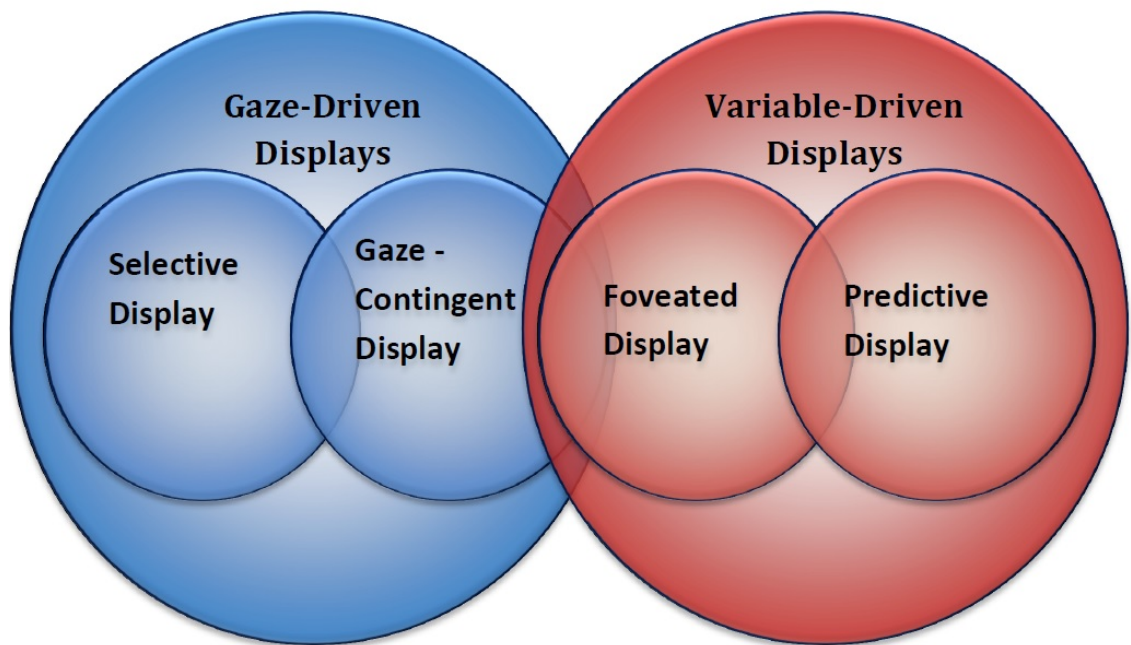


Figure 1.1: Classification of interactive applications.

1.1 Motivation

Although *holodeck* technology is the ideal of fictional *Star Trek*, in the current state of the art, designers and developers of VE often aim to create a realistic virtual environments (VEs) under many restrictions. For example, they need to develop technology that supports an illusion of immersion in an imaginary world while being restricted to a small display that usually accommodates at most two senses (vision and audition). GCD offers the opportunity to alleviate some of these VE restrictions in a natural and transparent manner.

GCD applications can be used to enhance or augment visual feedback to enhance an interface and to also enable computer based simulations to explore different perceptual capabilities, specifically visual impairments. For example, GCD applications can be very useful in the simulation of a user experience with biological constraints such as partial vision lost, commonly seen in age related macular degeneration (AMD) patients. To date, most of this research has focused on simulating scotomas (blind spots), with little considerations of other deficits such as glare or metamorphopsia (visual perturbation of a scene). However, I believe that much can be learned about coping strategies for human locomotion and other behaviours by simulating distortions of different regions of the visual field.

Gaze contingent displays are also potentially useful for HCI. Virtual reality

development often requires computationally demanding simulations that can introduce significant delay that can severely impact the overall performance of the system, human performance and overall user experience. To solve this problem in the graphics domain, a variety of heuristics are used to minimize the computational load. Once again, GCD can be used as a tool to minimize the computation load by only simulating those aspects that the user is currently looking at. Usually, this is done by concentrating the rendering and modelling effort on the areas of attention, while using simplifications in other regions. For example, Sullivan et al. [145], used GCD for collision dynamics simplification in the areas where users' were not looking. For example, Sullivan et al. [145], used a more simplistic algorithms for simulating collisions between objects in the areas where users were not looking by utilizing the capabilities of GCD. Similarly, one can use GCD to substitute large and complex 3D models with their simplified versions [33]. These techniques are not inherently restricted to the graphics domain and conceptually the same techniques could be used in the audio domain. Nevertheless, the benefits of such a a multimodal system have not been researched or developed to my knowledge.

Furthermore, in some cases GCD could be used to simulate a visual cue that is not otherwise provided by the standard visual displays. For example, GCD can be utilized to simulate depth cues such as depth-of-field. By modelling gaze-contingent

depth-of-field (DOF) one can change the clarity of objects at different depths in real-time based on knowing which virtual plane the user is fixating at any give instance. It is believed that by simulating DOF one could improve a user’s comfort with both monocular and stereoscopic displays. However, this topic has not been thoroughly explored in the literature.

1.2 Goals

In my dissertation, I examined a series of challenges and opportunities associated with gaze-contingent multi-modal and multi-sensory displays. I was interested in developing a system that would serve as a test-bed to develop and evaluate gaze-contingent techniques. In addition, I developed a set of exemplar real-life scenarios such as urban road driving, virtual desk and cocktail-party simulations, in order to evaluate the system I developed as well as to perform experiments that would address important psychophysical or HCI issues. Specifically, I addressed the following topics:

- In the context of visual impairment, I investigated virtual-reality simulations that support navigation research. Specifically, I investigated the effects of visual deficiencies on different aspects of locomotion, such as heading judgments and steering, in the context of driving simulation. In this context, real-

time gaze-contingent simulations were required to control the retinal stimulus provided to the participant in order to obtain a controlled simulated visual deficit under natural, free-gaze conditions.

- In the context of display usability, I examined the augmentation of graphic displays for normal users to support the display of three-dimensional imagery. In stereoscopic and other 3D displays, the eye must always focus on the display for clear vision, which leads to a conflict between the eye's focus state and the depth presented on the display. Gaze-contingent depth-of-field blur could be used to simulate the changes in relative focus in the image, resulting from changes in fixation. I examined whether perception of depth and visual comfort were enhanced with simulation of focal-blur in the virtual scene.
- Finally, I investigated the potential of gaze-contingent audio-visual displays and how they improved the auditory experience in a 3D VE.

1.3 Organization

To arrive at a common understanding of what is a GCD, in Chapter 2 I examine different GCD applications, their limitations and possible areas of improvement. I also examine different domains and applications where GCD are used. Current algorithms used to implement GCD are also discussed and then I outline the current

issues with this technology. In Chapter 3, I present a gaze-contingent system for virtual environments, which supports simulation of visual field defects. The chapter also presents a series of heading and steering experiments that examined the impact of visual defects such as occlusions and distortions on heading judgements and steering ability on curvy roads. In Chapter 4, I describe a GCD system that simulated DOF and a set of experiments I have conducted to evaluate the effectiveness of incorporating DOF simulations with stereo displays. In Chapter 5, I develop gaze-contingent audio techniques and assess their impact on user experience. This chapter also presents a series of user studies. Finally in Chapter 6, a summary of all the gaze-contingent techniques implemented here is provided and I present the overall conclusions, contributions and outline possible future extensions.

2 Background

2.1 Gaze-Driven Applications

In this chapter I will discuss the application domains suitable for GCD. I will then talk about some implementation strategies and issues associated with different implementation algorithms.

2.2 Where can we use GCD applications?

The word display can have different meanings in different contexts. In this section, I will treat display as a multi-sensory device that can provide information for different senses. Specifically, I will address three types of display - the most common is the visual or graphical display. However, I will also talk about auditory and tactile displays, since GCD can also enhance our other senses.

2.2.1 Vision

There is always incentive to improve a visual display’s resolution, speed, and dynamic range as these usually fall short of human perceptual capabilities. Designers have worked to mitigate trade-offs between size and resolution [110]. Since the human visual system is “an elegant compromise between the competing goals of maximizing field of view, maximizing spatial resolution, and maximizing neural resources” [64], its simulation can be used to present a large display with effectively high resolution with limited resources.

2.2.1.1 Bandwidth Reduction

One of the most common motivations to use GCD is to reduce transmission bandwidth through image compression[110]. To achieve image compression with this method, one reduces a transmitted image resolution to match the fall off in acuity in peripheral vision [63, 23]. In the ideal case, if the local image resolution matched the resolution distribution of the retina, the effect would not be noticeable to the observer, but the resulting images could be compressed efficiently since they would contain less high-frequency content [23]. Many system designers quantify their system performance by computing a compression ratio. For example, Geisler and Perry [63] demonstrated a three times greater compression ratio in videos with

multi-resolution images. The main commercial applications that are interested in bandwidth reductions are in broadcasting, tele-operation, tele-presence and video conferencing [45, 93, 178, 195]. One of the advantages of broadcasting applications is that most of the time the transmitted content is not in real-time. Even “live” television can tolerate considerable latency. Consequently, processing can be performed off-line, and any delay in post-processing will not affect the user. In gaze-contingent compression this advantage is eliminated as it is hard to predict where the viewers will look on a given frame, and therefore real-time GCD is not normally feasible in these scenarios. Thus, the processing typically uses an automatic prediction or eye-movement recordings from a single user or set of users to represent the point of regard (POR) of a typical viewer.

Tele-operation, on the other hand, happens in real-time and can greatly benefit from incorporating GCD [47]. This domain is often very resource constrained, especially in terms of communication links, storage or computational power. On the other hand, such a system still needs to provide reliable, interruption free and fast data transmission. Failure to do so can lead to a rise in the user’s frustration with the system and/or failure to complete important transactions. The solution lies in understanding the specifics of the tele-operation application and in simplification of the problem. In many cases there is no need for transmitting details of an entire image, instead it is often sufficient to provide details in the area of the user’s in-

terest. For example, when tele-operating a remote vehicle, high resolution is used to recognize landmarks, estimate the heading direction for path planning and for obstacle avoidance. However, for judgments of heading from optical flow, and for detection of incoming objects and/or vehicles, low resolution in the periphery is sufficient. Thus, it has been claimed that using foveated imaging driven by an eye-tracker allows bandwidth to be significantly reduced while still providing adequate information to the user [63]. Finding better algorithms for improving bandwidth will continue to be an active research question as long as the hardware cannot handle the load of increasingly complex applications. In tele-operation, it is important to note that the delays associated with such system increase due to the need to transmit gaze to the sender and then return the processed content back to the user. Yet, in the context of GCD and gaze-contingent foveation techniques, one of the more interesting aspects is in determining the appropriate thresholds of simulation integrity required to accommodate both the user's experience and bandwidth improvement. Furthermore, although gaze-tracking has been used in telecommunications, the mechanisms have been very simple and have often incorporated only visual or auditory simplifications. Very often, the information provided distracts the user instead of improving performance or facilitating better collaboration. It is also important to incorporate temporal gaze-contingent techniques to improve these systems.

2.2.1.2 Detail reduction

When very complex models with a high degree of “details” are rendered, it can be demanding of a system’s resources and time. Some costly rendering operations can be of little benefit as the user won’t actually see many of the model’s details. For example, in standard 3D graphics when the application tries to render a complex model that is located behind the camera’s position, such content is typically culled and not rendered. Similarly one can cull rendered details in modelled objects located in regions of the visual field where such detail could not be resolved. One solution to this problem is to use a model-based GCD level of detail approach [34]. Along with reducing rendering effort, using a model-based approach (section 2.3.3.2) can also significantly contribute to on-line storage reduction. Once again, the system designer has to look at how many computational resources they have when using a GCD and what is the best technique would allow them to achieve their goals.

2.2.1.3 Reading

Since the mid 1970’s, eye tracking technology has enabled and promoted reading research and conversely reading research has been a key driver for eye-tracking development. Early reading studies simply used eye-tracking for passive collection and

analysis of eye movement data, but have evolved into more complex gaze-contingent systems known as eye-contingent display change paradigm systems [156]. These were among the first GCD systems and their primary purpose was to investigate people's perceptual span during reading.

2.2.1.4 Virtual Reality and Simulators

The virtual reality (VR) interaction metaphor implies that the user interacts with virtual objects in a natural way with a high degree of freedom. With the help of VR simulations, researchers can conduct controlled experiments to study activities people perform in everyday life. VR simulation can also provide training, rehabilitation or educational environments, where people can be exposed to experiences that they otherwise would not be able to go through in real life. Although the technology today does not allow building a holodeck, system designers try very hard to make the virtual experience as realistic as possible. For this purpose, large displays are used and usually tracking systems are incorporated. The presence of eye and head-trackers make it easier to incorporate GCD in the VR system. Yet there are still several challenges any VR system must address. For instance, it has to deal with a heavy rendering load across a large field of view with real-time response requirements. In many cases to solve this problem, system designers have to compromise on the display quality. GCD system can be a good solution as dis-

cussed earlier. Another problem associated with VR is simulator sickness usually linked to sensory distortions [125, 89]. The cause of many of these distortions can be attributed to two factors - long delays and lags in feedback (due to complexity of the rendering content or other computationally intensive algorithms such as physical simulations)[151], and inconsistent sensory cues[125], for example, from convergence and accommodation. Hence, there is a long standing interest in using GCD to minimize the effects of simulator sickness. For instance, Parkhurst and Niebur [150] explored the claim that GCD is beneficial in reducing the likelihood of simulator sickness. Although they did not find that their system reduced simulator sickness, GCD indeed reduced the computational complexity of a rendering as discussed earlier, while maintaining perceived graphics quality. Schumacher et al. [171] explored a different strategy for latency amelioration by hiding graphic updates during saccades so that they were not perceptible to the users.

Navigation: One of the primary interests in this type of research is to use virtual environments to understand people’s navigational mechanisms. For example, goal-directed walking can be very difficult for some individuals, some can be slower than their peers, and others can hit obstacles on their way. The consequences of these undesirable behaviour patterns can be even more significant when trying to steer a vehicle moving at higher speed. Many researchers have tried to explain this type of

behaviour [196, 161, 162, 224] by looking at gaze positions during different phases of steering (Chapter 3). While, it is possible to perform experiments with people who have an impaired visual system, the results can be inconsistent and hard to repeat. Simulation of visual impairments combined with GCD, on the other hand, can be very effective. For instance, Fortenbaugh et al. [57] combined a goal-directed walking task and GCD. Their results supported the claim that good navigators rely on internal spatial representations to a greater extent than poor navigators do.

Flight Simulation: One of the first implementations of GCD was in flight simulation. In these early applications, developers used gaze-contingent systems to reduce level of details (LOD) and to blur the edges between the foreground and background [1]. Some simulations used gaze-contingent foveation techniques with only two or three levels of resolution and a very large central region [93]. Mostly, GCDs were used to reduce the computational complexity of a rendered 360° visual environment [48]. For example, the ESPRIT system used an eye-slaved projected raster inset to simulate a detailed conical volume from the pilot’s eye and less detailed peripheral images [56]. Although, there is a lot of potential for GCD in visual simulations, it has not been widely exploited by the industry in either medical or entertainment domains. One of the questions that GCD systems have not fully explored and evaluated is simulation of depth of field for virtual reality (VR)

applications in general (section 2.3.6).

2.2.1.5 Augmented Reality

Wearable computers such as Google Glass have become more prominent and we are not far away from wearable eye-trackers that can augment our interactions with the real world. Specifically, gaze-guided object recognition can provide additional information about objects the user is attending to [193, 53]; this has been applied to applications such as information kiosks [53] or museum guides [193]. There are several issues that need to be resolved in order for GCD to become an integral part of such an augmented reality (AR) application. Calibration in difficult and varying outdoor environments present a challenge. Also eye-tracking should not restrict vision and allow easy maneuvering through large spaces. I believe this will be soon resolved with wearable computing systems and advances in mobile eye tracking.

2.2.1.6 Low-Vision Enhancements

Another area where GCD can be applied is in low-vision enhancement. In many cases, as a result of retinal pathologies, such as macular degeneration or glaucoma, patients can lose portions of their visual field. In such cases, GCD can be used to map the visual input projected to the damaged fovea to other nearby functioning retinal locations [150]. For example, in the case of macular degeneration, the in-

formation from the central visual region can be moved to a different region that has not been affected by the disease. Such an application can be very beneficial for reading. Tlapale et al. [191] built a SOLAIR system that modified text location so that it would be visible outside of regions of user’s blind spot (scotoma). Similarly, Duchowski and Eaddy [49] presented an application to improve the visual experience of macular degeneration patients. Instead of shifting the visual information, the system magnified the image outside of the scotoma area. In their review, Dagnelie and Massof [40], proposed other visual enchantments such as local contrast enhancement or image stabilization. Despite this interest, there have been no empirical studies that show that GCD enchantments would actually improve the visual experience of visually impaired individuals. One of the reasons might be that is very hard to track the eye movements of these patients. The most crucial factor is the calibration process that usually requires a stable fixation at the calibration points, while macular degeneration patients typically cannot perform this task well [55, 130]. For example, Model and Eizenman [130] have built such a system, however, it is not as accurate as other eye-tracking systems and is not readily commercially available. Furthermore, such a system should be adapted to the scotoma regions, which requires individualized and precise visual field testing. Furthermore, when simulating visual experiences of different people, it is very important to convey the actual patient experience. This can be very hard as any

experience is subjective. Hence a person who constructs such a system should be able to answer questions such as how realistic their simulations are. Do people get the full understanding of what it is like to have a scotoma in their central vision, despite viewing a short second-person presentation? How does it affect their walking or reading strategies compared to people who actually have to live with permanent visual damage?

2.2.1.7 Education and Social Interaction

For Perry and Geisler [65], one of the main motivations for developing a gaze-contingent system was for educational purposes. A system that can imitate the visual experience of visually impaired individuals can be very helpful for the patient, their families and the health care staff.

Visual simulations can help to educate people about the difficulties each visual disease can entail and can alert people to warning signs of serious disease. In addition, the gaze-contingent simulations can assist in developing techniques that can help in coping with visual disabilities. In particular, many visual impairments are very hard to understand without visualizations. For example, Vinnikov et al. [210] created a system that could simulate different impairments such as glare or metamorphopsia (visual distortion, one of the side effects of macular degeneration disease) that are commonly reported by patients and are not fully explored. Vi-

sual education could also increase sensitivity of patients to the progression of their condition if it goes untreated and promote empathy in families and caregivers. A different take on use GCD systems in education was presented by Dorr et al. [42, 84]. Dorr et al. [42] proposed to present foveated-footage to biology students, so that, by observation, they could learn proper gaze-patterns used by experts to classify and identify fish. Their system made the region of an expert’s fixation very detailed, while the rest of the image was gradually degraded. The results showed that such an approach can be very efficient for learning the expert’s scanning patterns. Due to the fact that this system is off-line, it avoids many challenges that real-time systems might face.

It is also important to mention that some of the aspects of education are social. A social interaction with a virtual avatar can be significantly improved by having a realistic gaze behavior [10, 15, 189, 99]. However, pairing the avatar’s gaze with the user’s tracked gaze could lead to a more realistic user experience. For example, this is what Wilms et al. [225] found in their paper, where they created a gaze-contingent stimuli for a neuroimaging study of social cognition.

2.2.1.8 Visualization

The term visualization is very broadly used throughout computer graphics. In many cases in the GCD literature the term refers to an attempt to display a richly

detailed image or scene with limited resources [30, 34]. In contrast, visualization in a computer graphics context refers to any technique that conveys information through images, diagrams, or animations. GCD for visualization has been explored only in a very limited context. Yet by applying gaze-contingent mechanisms, one can provide a much better visual representation of complex models at different levels of simplification [138]. In addition, most research has focused on the visual domain, while new techniques that involve auditory or tactile ‘visualization’ may prove to be very useful.

2.2.1.9 Attentional Guidance

A very common scenario is in the visual arts, such as in movie production or magic shows where techniques are introduced to guide the user’s attention and gaze. In some cases, it can be very beneficial if a system can actually help users to develop those patterns or compensate, when a required visual behavior was not achieved. For instance, this can be used to bring a driver’s attention towards obstacles on the road that would not be detected without the guidance [11]. Barth et al. [11] first identified the most salient locations and then changed the probability that one of these regions would be looked at by briefly (and nearly imperceptibly) flashing red dots in the region or conversely, by locally reducing the temporal frequency content of regions whose saliency should be decreased. Barth et al. [11] used

GCD techniques to ensure that their manipulations would be unnoticed by the user. Dorr et al. [43] used gaze-contingent temporal processing to filter motion in the periphery, to reduce the likelihood of saccades to visual regions that did not include objects of interest. Indeed, the experimental results showed that such filtering was very successful in controlling eye movements. Dorr et al. [43] reported that, with thier GCD system the number of saccades to the peripheral region was reduced by 20%. Overall this type of research demonstrates that GCD techniques can be efficient when there is a need to modify eye movement patterns. One of the problems with gaze-contingent attentional guidance is the latency of the system. These systems need to predict where people will be looking next, and guide them to a different location. In such scenarios, people become very sensitive to delay in the system’s response. As a result, it is very important to use gaze guiding techniques that do not distract users from the actual task at hand.

2.2.2 Tactile and Haptic Displays

The desirability of multi-modality displays grows with the amount of the transmitted information required by a single user. This is due to the fact that sometimes it is not enough to use only one channel of communication. Particularly in aviation, when the visual channel is overloaded, tactile feedback becomes important. The tactile and haptic information can be presented to any part of the user’s body.

Although most commonly the information is transmitted through the user's hands [25], it can be also be a Tactile Torso Display [204, 203] and/or a tactile seat. These displays could be greatly enhanced with gaze-contingent information. For example, if there is some important visual event occurring in the user's peripheral vision that would not be detected by the user, the warning could be executed through the tactile feedback. Furthermore, tactile displays can be used as visual guidance. In a more traditional setting, tactile and haptic displays have been used when visual feedback is not available. Similarly when combined with GCD, it is possible to share what is seen by one person with another person who is blind or who cannot see from their point of view. Unfortunately, there is almost no research known to me in this direction. For now, with the wearable devices such as with smart watches and Google Glass, even small gaze actions that could help to activate and deactivate the device might be a very useful tool to the users.

2.2.3 Audition

In his visionary paper, Bolt [24] looked at the problems of multiplicity of “windows” for large format displays. Although at the time there were not that many users, who would face such a problem or use large displays for their daily activities, today this problem is very relevant. People use large displays or even a few displays with multiple windows and applications running simultaneously. Hence creating a

dynamic, gaze-interactive interface, is much more desirable today, especially when dealing with large amount of distinct video and audio information. Bolt's main interest was to control sound through gaze tracking [24]. He proposed several approaches:

- The loudest sound track comes from the looked-at window, while the other windows are attenuated to a background volume.
- The user's auditory locus is created by providing stereophonic output from the window that user is attending to and the windows nearby, while windows that are far from POR provide monaural sound to only one of the ears.
- Spatial-temporal sound effects were based on amplifying sound relative to its proximity to the viewer's gaze location and the time passed since the observer last interacted with this window.

Although, some research was done in this area, it was limited by the technological state-of-the-art at the time. It would be interesting to see what results could be achieved if a similar system were revisited today. Another interesting use of GCD in the audio domain was reported by Starker and Bolt [177]. In their paper, the authors presented a system that inferred areas of interest from the user's gaze and then performed zooming and change of audio content. In particular, the narration was driven by the user's area of interest (AOI). Such application can be very useful

in public domains such as at exhibits and other educational outlets.

Similar to Starker and Bolt [177], Eaddy et al. [53] proposed a system that augmented the user’s visual search with audio based on their gaze location. Unfortunately, the research was limited by slow tracking technology; hence the results were not conclusive. In summary, this area is not very well researched or exploited. Using gaze driven audio augmentation could greatly benefit areas such as visual aids for the visually impaired, educational settings that support self-exploration and multi-modal complex systems’ operation.

2.3 Implementations

2.3.1 Foveation

The name for the foveation method comes from the word fovea, which is the area of highest visual acuity. The fovea lies at the center of the visual field. The density of cone receptors is highest here and then declines with eccentricity, causing the visual acuity to degrade towards the periphery. This feature of the human visual system can be exploited for many different purposes. In particular, images or video can be modified to simulate the variable resolution across the visual field. This is often referred to as the foveation method [23].

The foveation technique can be used in many different systems grouped under

the label of *variable resolution display (VRD)* [150]. The primary characteristic of a VRD is the emphasis of the output (the end result) of the processing rather the input. A VRD can be a GCD systems but a VRD can also be implemented with systems that predict, rather than measure gaze [52, 12, 68, 60] (Figure 1.1). The major disadvantage of these predictive systems are that they are not robust enough to predict the exact location of the user’s gaze location[149]. In order to stay within the scope of this chapter, I will only concentrate on the problems of VRD within the context of GCD.

One ways to classify a VRD is based on the level of resolution presented to the user (figure 2.1). Usually it directly correlates with the application domain. Overall, there is a wide spectrum of applications that use different levels of resolution as shown in figure 2.1. For instance, some GCDs were developed to improve refresh rate and/or to reduce bandwidth. As the result, developers identified and processed the displayed image at two levels of resolution: high resolution in the central region and low resolution in the periphery [221, 158]. These are known as bi-resolution displays. Bi-resolution display implementation was one of the most common approaches for early GCD development and still remains a popular approach in certain domains, especially where there are limited number of resources [150]. At the same time, this decrease does not reflect the natural decrease in resolution across the human retina. Hence, unless the size of the high resolution region is chosen in such

a way that it matches or exceeds the resolutional threshold of the user, the two regions will be very apparent to the user. This is known as the window effect can be very noticeable and detrimental to effective task performance [59]. In rare cases, when the two regions are extremely different, such an implementation can actually be beneficial for highlighting information difference between the regions [139, 138].

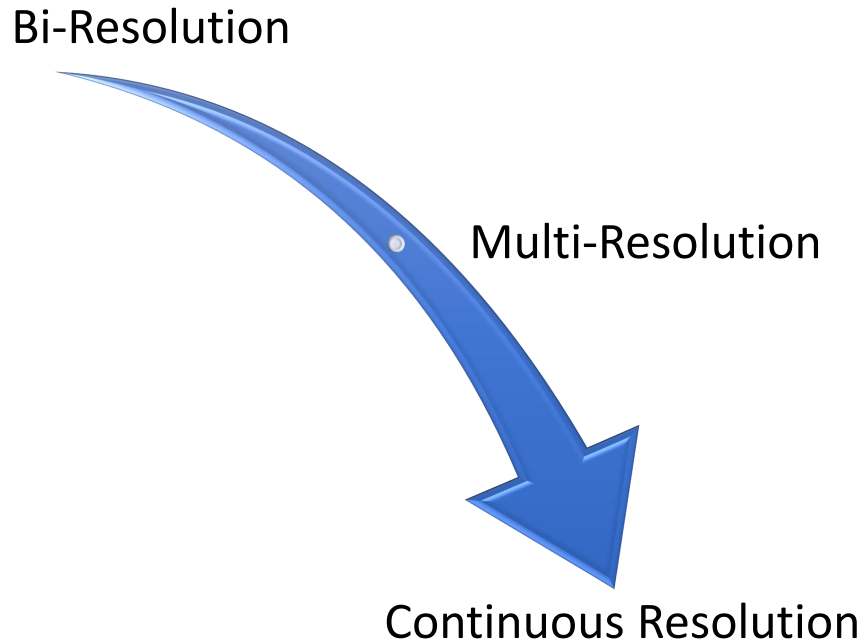


Figure 2.1: Different levels of resolution for VDR.

Following the bi-resolutional displays on the GCD continuum are the multi-resolution displays. These displays have several levels of resolution [68]. If the number of levels is high enough, the user will experience continuous resolution

degradation. In this case, GCD will be referred as a continuous display. One of the first continuously foveated displays was presented by Perry and Geisler [64]. The problem that the GCD developer has to solve is how many levels should be displayed and how they should differ from each other. In many cases, developers will try model their resolution as closely to the human visual system as possible. In order to achieve a close match, a human perceptual model is required. Nevertheless, one has to be aware of the implementation constraints and benefits in order to choose an appropriate type of foveation to develop.

Previous techniques have typically been implemented in monocular or bioptic (same display to both eyes) displays. Yet, it is possible to accommodate binocular vision through GCDs. Hence, one can use stereoscopic foveation. This technique manages the AOI for stereoscopic displays, by generating content for both the left and right eyes. The benefit of this technique is that it models the natural perception more accurately by simulating the binocular visual field.

2.3.2 Perceptual Modelling

Luebke et al. [113] stressed that the GCD should be able to translate an observer’s visual limitations into spatial frequency representation on the visual display. As a matter of fact, GCD displays should also be able to model temporal and motion limitations. Therefore, the GCD system designers have to create a perceptual model

of how the viewer sees through the use of GCD. The most common perceptual aspect modelled is contrast sensitivity. Kelly [86] introduced a general formula to express contrast sensitivity, as follow:

$$F(\alpha) = \alpha^2 e^{-\alpha} \quad (2.1)$$

where, α is the spatial frequency in cycles per degree and e is a constant base of the natural logarithm. This formula can be scaled to model different conditions. For instance, Luebke et al. [113] identified three properties of the human visual system that a GCD system can be based on.

1. Velocity: Modelling kinetic vision is not trivial. First, the perception of motion can depend on many different factors. For example, if the user tracks a moving object, the retinal image of the background would have less detail and if the user is fixating the background the opposite should be true. In addition, the contrast sensitivity varies with the velocity of a moving object [33]. In other words, the perceived detail of a moving object will depend upon its velocity and spatial position relative to an observed object.
2. Eccentricity: The contrast sensitivity varies with angular distance from the centre of user's gaze [33, 19]. As many foveated applications use only the head's position to estimate gaze [143, 220] they do not take this into account.
3. Depth of Field Modelling: Modelling DOF would help to determine which

objects will appear sharp and which will be blurred.

Each of the components above can be then combined together for a composite model. The model can then be applied to produce a visualization of any given scene.

2.3.3 Rendering Techniques

Once the perceptual model is chosen, GCD systems developers have to address how to incorporate the visual perception model into the rendering. Duchowski [46] identified two classes of rendering techniques for visual GCD - the screen-based approach (section 2.3.3.1) and the model-based approach (section 2.3.3.2). Later, Murphy et al. [135] developed a hybrid system that combines the two approaches together in attempt to build a more efficient system. In addition, GCD techniques can have an indirect effect on rendering, for example, by changing the dynamics of the manipulated world [146].

2.3.3.1 Screen-Based Approach

Screen-based applications treat the display as a 2D image and all the processing operations are performed on an array of pixels (figure 2.2). The most popular work among screen-based algorithms is by Geisler and Perry [63, 64, 23]. Their implementation was one of the first to support arbitrary shaped high and low

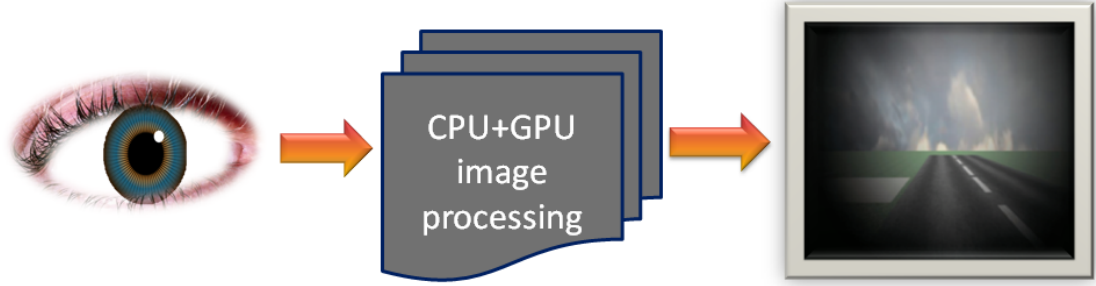


Figure 2.2: Screen-based GCD.

resolution maps, thus, enabling simulation of the visual field of visually impaired as well as non-visual impaired individuals. Böhme et al. [23] claimed that the gaze-contingent spatial filtering algorithm introduced by Perry and Geisler [64] was an improvement since it produced smooth, artefact-free results, while the previous algorithms introduced many visual artefacts. Their algorithm creates a 3D surface from a 2D image of the scene, with the new dimension being the spatial resolution of the image. Thus, the level of detail in the rendered image depends on the z-coordinate of the spatial resolution image. Thus, their algorithms require two inputs: (1) a sequence of images with the direction of gaze of each image; (2) An arbitrary resolution map that specifies the spatial resolution across the visual field. The process is based on multi-resolution pyramid construction. One of the drawbacks of such an algorithm is that it introduces latency for processing all the blur levels. Fortunately, many hardware accelerators can overcome this limitation

General Screen-based Algorithm
(adapted from Perry and Geisler [64])

1. Compute blending functions from the visual field map
2. Obtain the input image
3. Obtain the current gaze direction
4. Compute the multi-resolution pyramid for the input images
5. Shift the resolution map and blending function to align with current gaze direction
6. Up-sample, interpolate and blend the multi-resolution pyramid using the shifted blending functions
7. Display the output image
8. Go to step 2.

by significantly improving processing time.

Visual Field Map (Step 1) Although visual perceptual models could be computed for each pixel, it can be very computationally expensive to an extent that such a solution is not possible for real-time operation. A different solution is to pre-compute the model once and store it for further use. Therefore, the visual field map can be defined as a two-dimensional function that describes the perceptual model relative to the visual field. The range of the function is usually in a range of 0.0 to 1.0. For convenience, it can be stored as a gray scale image. Since, most of the algorithms in the literature deal with visual resolution such a map is often referred as a resolution map. In the early years the resolution map was heavily associated with bandwidth reduction. For example, Perry and Geisler [64] define spatial resolution for each pixel location as “a half height bandwidth of the transfer function associated with the low-pass filter applied at a given pixel location”. It is

important to note that there are many ways to model visual perception and system designers may choose not to incorporate every aspect of the model but rather concentrate on a single characteristic. Furthermore, a system can have multiple visual field maps where each describes a different characteristic of the visual model. For example, Vinnikov [206] described visual acuity, change of contrast, colour, noise and metamorphopsia through different visual field maps. Due to the fact that each visual field map can be encoded into one colour channel, multiple visual field maps can be combined in to the four colour channels of a single image. Nevertheless, the most common way to use a visual field map is to describe the level of blur that will be applied to the image.

The visual field can be either mathematically modelled or can be measured based on different visual tests, for instance, the Goldmann perimetry. Usually the mathematically modelled visual field is taken from normative data, while the empirically measured resolution maps are important when an individual visual map has to be modelled, for instance, when particular regions of the visual field are important either for calibration purposes [206] or for visual enhancement [49]. For example, measured visual field maps can be used for modelling the visual fields of glaucoma or AMD patients. In other cases, they can be used to model an individual's blind spot or other scotomas [207]. Although, the image processing associated with a visual field map happens in 2D, it does not imply that it is

mapped one-to-one to the rendered display area. Typically, the centre of the of the map is the centre of the image representing it. Since the user’s gaze location (unless purposely restricted) can be at any location of the displayed scene, the size of the visual field map has to be at least twice as big as the rendered display to make sure that the resolution map can cover the whole display regardless of gaze location. Additionally, in order to be expressed in terms of visual direction, the visual field should be mapped onto a sphere [206, 57].

Input Image (step 2) Intuitively, screen-based applications should use 2D images, such as movie clips, specially recorded video or photographs as input images, while the virtually generated scenes with 3D graphics should be processed with the model-based approach. Nevertheless, with the help of the graphics processing unit (GPU) and shading languages, it is possible to work with 3D models in the context of a screen-based approach. This can be achieved by taking screen-shots from the environment with the help of a *frame buffer object (FBO)* and using them at the image processing step [206, 57].

Gaze Direction (step 3) Retrieving a user’s gaze location is one of the most crucial functions of a GCD system. In order to avoid lag-related artefacts, the data must be as current as possible. For that purpose many systems use multi-threading algorithms to separate the process of gaze retrieval and image processing [44].

Image Processing (Steps 4, 5, 6, 7, 8) The rendering techniques vary based on the intended visual effect and implementation constraints (figure 2.3). Some approaches can work with a single input image. Nowadays, the GCD changes are usually done through hardware and known as fragment programming [48]. Other techniques work by combining several layers of images with some sort of masking reference. These techniques are called multi-texturing techniques.

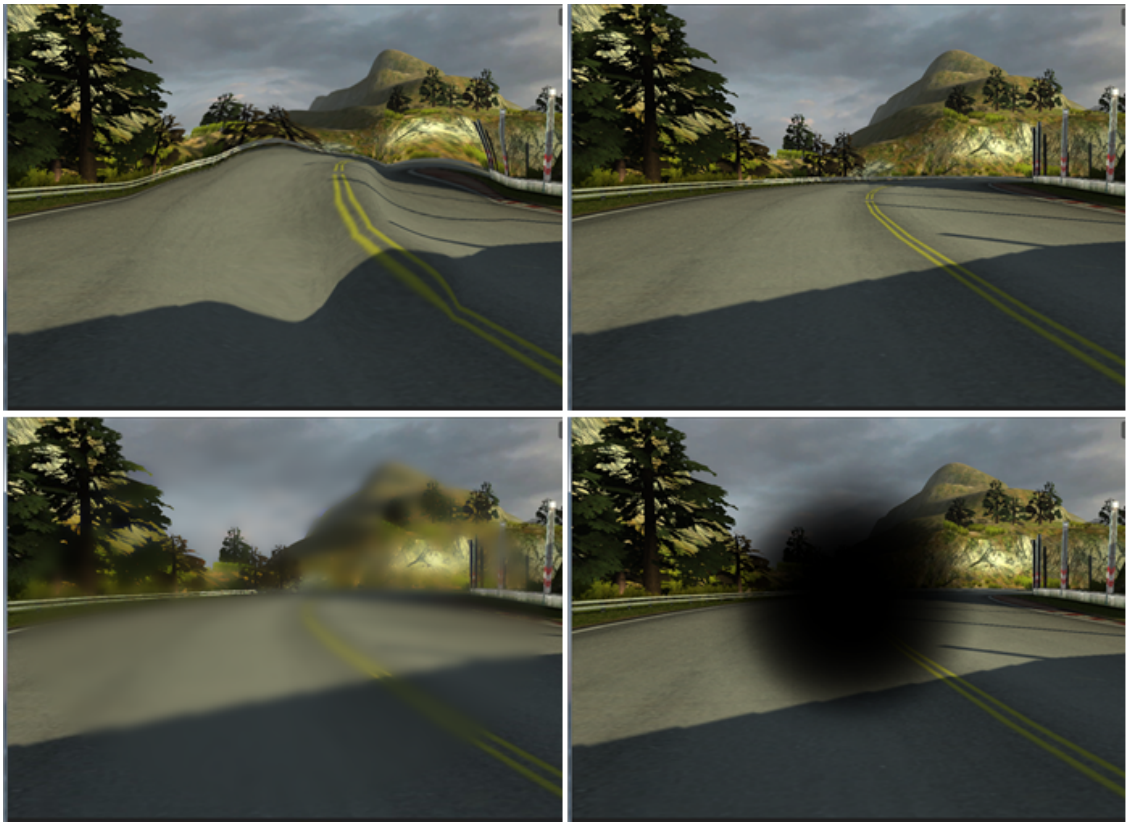


Figure 2.3: Examples of visual effects: (1) No defect (Top-Right); Visual perturbation (Top-Left); Blur (Bottom-Left); Occlusion + Blur (Bottom-Right).

Software Implementation In their paper, Duchowski and Çöltekin [48] differentiate a screen-based approach from shader programming. However, the theoretical aspect of the two approaches is the same and they only differ in their implementation. Hence in this chapter, a screen-based approach is identified as an image processing approach that can be implemented either through direct image processing or through texture-mapping, typically with pixel programming or fragment programming. During direct image processing, the rendered image is a picture or a video stream that was originally pre-processed and rendered directly onto the screen. Conversely, for texture mapping techniques, the image is viewed as a texture map and it has to be mapped onto a graphical object, such as a polygon or a hemisphere. Usually the size and shape of such an object is congruent to the display's shape and size. For both implementations, it is important to ensure that the viewing window, the display and the image are properly scaled with respect to each other. Ideally with 1:1 pixel mapping. Also, one has to ensure that the desired hardware operations are supported by the graphics card being used. In addition, if multiple textures are involved, one has to decide on how to merge them together. The direct approach usually employs a channel for blending, while texture-mapping can work simultaneously with multi-dimensional texture arrays and apply complex blending operations at very fast speed and at a much lower cost.

Drawbacks Although screen-based techniques are very powerful and very commonly used, one should be aware of their limitations. For instance, it is still unclear in the literature what functional correlation exists between the simulated blur and the spatial contrast sensitivity of the human participant. In other words, what physiological measurement should guide developers to decide how to determine a Gaussian or other function for a particular blur level or deciding upon the number of levels in the pyramid?

2.3.3.2 Model-Based Approach

Model-based applications treat the rendered content as a 3D model and all the processing operations are performed on edges and vertices of the polygonal models. Model-based techniques are often referred to as LOD techniques. As with the scene-based techniques, LOD techniques are not always gaze-contingent. They might use other strategies, such as distance prediction or head tracking to determine the part of the scene to be simplified.

LOD Luebke et al. [113] defined LOD as a process of reducing models in a rendered scene by replacing them with models that contain less detail to achieve fast rendering. This process usually applies to small, distant, or unimportant portions of the scene. The basic principles of LOD are attributed to Clark [33]. In his

work, Clark [33] described a hierarchical scene graph that could be used to quickly determine which model level should be rendered. Luebke et al. [113] note that the early scene graphs were constructed by hand and only in the 1990's was an automated process introduced. There are three frameworks for managing LOD:

- Discrete LOD: The first and the most traditional approach is the discrete LOD management approach. This is based on Clark's [33] technique. The processing is done off-line and can support real-time rendering, since it is very simple and efficient. As a result of the off-line processing, this method cannot accommodate the actual viewing distance when the observer is watching the scene. As a result, models are uniform in their level of detail. Hence, this technique is also known as an isotropic technique. Unfortunately, the switch between models can be noticeable causing perceivable flicker referred to as the pop-up effect.
- Continuous LOD: The second framework is the continuous LOD or the progressive LOD framework. This type of technique creates a continuous spectrum of detail for one model during a pre-processing stage. This requires LOD to be extracted during run-time. As a result, the rendering is more precise and supports a better granularity, better resource allocation and higher fidelity for a given polygon count. Furthermore, these techniques help in

progressive rendering and interruptible loading when dealing with retrieving large models.

- View-dependent LOD: The last set of techniques, which are most relevant to this chapter, fall under the view-dependent LOD framework, also known as anisotropic. These techniques use view-dependent simplifications in real-time in order to dynamically select the most accurate LOD needed for each rendered model [77]. Consequently, each scene and even each individual object can have multiple levels of details in its different regions. Consequently, it shares the same benefits as the continuous LOD techniques.

A good implementation can enhance an efficient LOD selection that, in return, can significantly improve a generated frame rate. This is particularly true for terrain rendering. Nevertheless, these techniques require extra processing for evaluating, simplifying and refining the models at run-time; as well, they require extra memory allocation. Hence, in some cases, they can be very demanding on resources.

When working with LOD techniques, the process can be broken into three parts - criteria, mechanism and error metric.

Criteria refers to a set of functions that determines which area of an object or which objects should be drawn at each LOD. There are multiple possible criteria that help to determine what area of a model to simplify. For example, the choices

can be based on properties of the surface, such as curvature, colour or texture [146]. By choosing the right criterion, one can create an illusion of a very detailed mesh. Other factors that can be considered when presenting a non-uniform LOD are as follows:

- The Euclidean (absolute) distance between the current viewpoint and an object in virtual space.
- The size of the objects, since smaller objects would require less details.
- The projected image size - the bigger the displayed model the higher the level of detail it should be presented with.
- The model's priority - some models can be conceptually more important and hence require more details.
- Hysteresis can be used to minimize the popping-up effect (new models unexpectedly appear on the screen) and prevent overloading the CPU by delaying the LOD switching.
- There are several perceptual factors that can also be considered. These factors are all based on the knowledge of the user's gaze. Such information allows much more precision and accuracy in changing the models. Hence, they are the most important in context of this chapter, since the model selection is

contingent upon the result from eye and head tracking devices.

Mechanism is the process that creates and modifies the LOD hierarchical structure and determines how each model will be simplified. There are two main approaches. One approach is to work with mesh geometry which is based on the location of the model's vertices or through mesh connectivity which deals with edges and faces of the model. In addition, the simplification algorithms can be fidelity-based, where the user specifies constraints that the simplified model must satisfy or budget-based, where the user specifies the maximum number of features a model can contain. While the first simplification is a NP-complete problem and hence the implementation has to use approximation algorithms, the second model relies on empirical observations to achieve best results.

Error metrics quantify the difference between the original model and the rendered one. If performed manually, they are used to evaluate the overall technique; in automatic settings it is possible to modify the process in real-time to control error.

Problems: Despite many benefits of the LOD techniques, there are several issues associated with hierarchical LOD generation. One of the most common problems is the pop-up effect. Although, it is mainly associated with discrete LOD, any

miscalculations in selecting the proper level or insufficient intermediate levels can lead to a dramatic change between the models in other techniques as well. Another issue that can happen, specifically, in continuous LOD and view-dependent LOD, is the presence of cracks between blocks. This attributed to the case, when a single model is constructed from different blocks with different LOD. As a result, if the silhouette (the bounding edge) is not maintained for all blocks, the blocks will not fit in together [108]. Hence, the reconstruction of the mesh geometry requires maintaining connectivity between blocks and the model has to be perceptually adequate [134]. By the same token, even if the silhouette is maintained, there is still a potential for artefacts, if the algorithm implements a suboptimal approximation [77]. Finally, it is important to reiterate that all hierarchical techniques require significant storage and/or can be computationally expensive.

Drawbacks of the Model-Based Approach One of the main problems that designers of the model-based systems face is the understanding of the user’s perception of simplified objects. In particular, they have to determine when to stop the simplification process, since at some level of simplification the objects lose features that help people to identify those objects. Thus, models semantics have to be taken into account when transforming a model. Possibly this could be achieved with eye tracking software that could learn about observer’s gaze patterns and thus

determined the most important features of interest that should be preserved at any level of LOD. Model-based approaches usually can perform only one type of visual simulation, and more specifically, acuity simulation. Other visual simplifications of features such as shading, shadows, depth cues have not been widely explored. For example, only few studies have tried to implement colour degradations in context of model-based techniques [48, 221].

2.3.3.3 Temporal Model

The previous sections dealt with techniques applicable to the spatial domain. Yet it is possible to apply similar algorithms in the temporal domains as well. For example, Böhme et al. [23] developed a system similar to Perry and Geisler [64] that operated in the temporal rather than the spatial domain. One of the main purposes of having temporal GCD is to process moving objects. For example, Böhme et al.'s [23] system masked moving objects in the observer's periphery while keeping the stationary details intact for the purpose of keeping the user's attention on the area of the task instead of peripheral distractors.

Software Implementation Usually the temporal algorithms are image-based and as a result they work with an image sequence. Nevertheless, a model-based implementation could also be used. Performing filtering on the entire video is very

computationally expensive and requires a very large amount of memory. Fortunately, most movements only last for several frames, which allow the system to utilize a temporal window technique instead. The temporal window defines the number of frames that will be used in constructing the next level in the hierarchy. In many cases, the temporal window requires processing frames that follow the current frame. For example, the system implemented by Böhme et al. [23] needs to know whether the movement happens in the next frame or not. In other words, it needs to look ahead and is therefore not causal. This in return limits their system to pre-recorded content or delayed response and cannot be used in real-time unless latency can be tolerated.

Temporal algorithms also use a resolution map, but the information encoded is the temporal resolution at each point of a participant’s visual field. In a similar manner, as in the scene-based spatial algorithms, the resolution map is aligned with the POR. The main difference is that the level of detail is achieved by interpolating between the levels of a temporal rather than spatial multi-resolution pyramid. The system designer must be aware that the human visual system is very sensitive to motion. For example when the temporal filtering is not accurate or inappropriately masked, sudden flicker or motion artefacts can be very noticeable and distracting. Also, as mentioned earlier there is still need for efficient and robust real-time algorithms. In terms of casualty, another way to approach this will be through

developing prediction algorithms that would be integrated with GCD techniques.

2.3.3.4 Hybrid Models

The previous sections discussed the primary visual rendering approaches. These can be combined in a hybrid system. For example, scene-based and model-based approaches can be combined. Similarly, temporal and spatial approaches can be combined as well.

Model-Based A model-based hybrid technique was proposed by Murphy [134]. The hybrid model combines the benefits of both families of algorithms. The technique still works with the 3D model, but instead of modifying the entire model, it samples the model’s geometry by using a ray casting technique. To generate direction vectors for the rays and for storing the resulting sampled data, the technique uses an intermediate mesh (virtual retina) located between the view points and the scene. By using fragment shaders to speed up the ray casting algorithm, the system utilizes the GPU capabilities. It also uses FBO and thus turns 3D geometry into 2D space, where other image processing operations can be performed more efficiently.

Spatio-Temporal Model Dorr et al. [42] built a system that uses space-variant filtering by using a spatio-temporal Laplacian pyramid. The system combined the techniques described in the previous sections. It also has the same problems as the

temporal only approach and hence was limited to pre-recorded content.

2.3.4 Gaze-Contingent Collision Handling

All the visual techniques discussed in the previous sections deal with changing the rendered content without interfering with the processes that led to a particular visual image. However, correctly simulating the physical processes can often take more resources than the rendering. Fortunately, since multidimensional movements can be very complex [145], people are often insensitive to anomalous physical behaviour. As a result, approximations of these processes can be sufficient for observers and thus greatly reduce the computational load without taking away from the realistic experience. For example, modelling proper behaviour of objects and particles during collision can be very computationally expensive. O’Sullivan and Dingliana [145], investigated perceptual factors such as eccentricity and object separation that determined whether a collision appeared believable to an observer. They showed that knowledge of a user’s gaze position can help in determining the required LOD to use for each collision. In addition, gaze position helps to determine when the system should simulate very realistic collisions that are believable to the user, and when to mask collision anomalies. O’Sullivan and Dingliana [145], break collision handling into three computationally expensive processes that can be optimized by using gaze tracking technology. The collision handling processes are

as follow.

1. Detection - This process is responsible for detecting contact or interpenetration between objects in a scene.
2. Modelling - The modelling process identifies the points or areas of contact between objects.
3. Collision Response - This process simulates the consequences of the collision by either moving the objects back or by simulating real world behaviour such as deformation.

The complexity of the processes depends on the number of objects one has to handle, the required fidelity of the response, the properties of the simulated materials and the shape complexity of each object. Usually a collision handler works by progressive refinement of collision data from the most simplified hierarchical volume representation (the sphere) to a more complex representation. Thus, checking collision between two objects can be very computationally expensive. A contingent system can assist in deciding about what level of detail the refinement process should stop. This can be achieved by prioritization of the models based on their eccentricity from the user's gaze position. The computational cost can be further reduced if all the models are divided into a small number of priority groups. Each priority group is then allocated its share of processing time by the scheduler, with

more processing being spent on higher priority groups. Due to the fact that the system uses hierarchical models, it must deal with many similar issues associated with LOD techniques. Hence, it has to address issues such as popping artifacts. Yet, in complex scenarios this approach can help to reduce the computational effort of a system.

2.3.5 Saccade-Contingent Displays

This set of displays is based on exploiting the insensitivity of the user to stimuli and change during rapid saccadic eye movements, typically for one of two purposes. First, saccade-contingent display can be used to study visual perception, including visual representations, visual attention, and change blindness [126]. In addition, it can be used in the context of interactive graphical displays in order to mask sudden changes that otherwise would be disturbing to the user or to ensure important changes are not missed [5]. In contrast to conventional GCD based on gaze location, this type of display requires identifying the type of movement performed. On-line eye movement classifiers require analyzing eye tracker data over a finite window. As a result, saccade-contingent displays are very sensitive to latency, since the change has to be executed during the saccade and saccade detection and classification takes substantial time. If parameters are not set correctly, some important saccades can be missed or other eye movements can be misidentified as saccades. This in return

can cause many unwanted artifacts and compromise task performance.

2.3.6 Depth and GCD

Another very important component of human vision is depth perception. However, when dealing with depth the full range of depth cues should be taken into account. For example, one has to look at cues such as accommodation and blur. Currently, there are two types of displays that have been explored in the context of gaze-contingent depth. They are the accommodative displays and the DOF blur simulation displays.

2.3.6.1 Accommodative Display

Accommodative displays are based on triggering an accommodation (optical focus of the eye) response from an observer. In the real 3D world objects are at different distances in depth and hence users need to change their focus to different focal distances [100, 232]. One way to approach the problem is to build a multi-focal display. For example, Neil et al. [136] and Paige et al., [183] built a 3D display that presented multiple virtual depth planes, and by doing this, they provided vergence and accommodation cues. Similarly, Akeley et al. [4], developed a display that formed three image planes at different physical distances. In particular, the virtual distance of the object from the viewer specified which plane the object was

rendered on (they also explored depth interpolation by blending across planes). In their prototype, the users observed three superimposed images through plate beam splitters redirecting light from three separate planes at different focal distances. One needs to ensure correct handling of view-dependent lighting effects, occlusions and multiple focal distances along a visual-line. Although accommodative cues can be simulated, this type of display has a limited number of planes and thus limited viewing volume or precision (these factors trade for a fixed number of planes). It also requires an extensive and expensive rendering procedure. In addition, it forces a fixed viewpoint. On the other hand, Shiwa et al. [175] built a stereoscopic 3D display with accommodative compensation (3DDAC). This display is optically translated in focal depth in real-time based on the observer's gaze position. Once again, this display has several limitations. First, the movement limits the working volume. Also such display is bounded by the speed of the mechanical movement of the screen and hence introduces latency and frequency response limitations to the system. Furthermore, both displays are intended to be minimized to the size of a wearable device yet there are many technical issues that must be resolved in order to achieve this.

Another group of displays that can provide accommodation cues are volumetric displays that generate imagery for a projection view of 360° . This is achieved with a high-speed video projector and spinning mirror. Such a display supports multiple

views for the rendered object, and hence most of the time people perceive correct ocular accommodation cues [85]. However, beside the fact that such a system is very complex and not available to an average user, the images cannot provide proper shading or occlusion cues [70]. Holographic displays, on the other hand, are image-bearing diffraction optical devices that can also provide accommodation cues produced by 3D objects and scenes. This is due to their ability to project light as it would have been projected by a real object. Holograms based on photography can provide realistic depth cues and present impressive static 3D displays. Yet, video or real-time image generation based on these principles demand a very high data bandwidth and therefore practical high-resolution digital holograms are not currently practical [169]. Similarly, light field displays preserve the directionality of light distribution across an image as well as the intensity (i.e. a 4D rather than 2D representation of the image) and thus support accommodation and focal blur cues in the display [229, 223]. Super multi-view autostereoscopic displays present multiple distinct views with a spacing smaller than the observers pupil diameter, approximating a light-field and reportedly supporting accommodation [185]. These light field type displays show promise but are currently expensive, exist only as research prototypes, and are very limited in terms of viewing volume, resolution or other important figures of merit.

2.3.6.2 DOF Blur Simulation

DOF blur simulation is another method to aid 3D perception of the rendered world. The DOF blur effect simulates optical blur by blurring the image of objects in front of or behind the focus point. Most DOF applications are not gaze-contingent. Instead developers often assume that the focus point is the centre of the screen. In addition, to simplify the rendering processes, instead of gradually changing the image quality, they pre-define a cubical volume that includes all the objects that will be rendered sharply and anything outside this volume will be blurred. A more ecological solution should allow the user to define the focus of fixation freely within the 3D space and the blur should be a continuous function of depth from the focus point. Brooker and Sharkey [26] proposed a system that generated a synthetic DOF centred on the user's fixation. Unfortunately, the experimental results did not show a significant performance improvement, and hence did not provide evidence for the benefit of DOF simulation. This could be attributed to several factors, such as system latency, low precision of the eye tracking device and inappropriate task (the task was depth estimation of textured surfaces). Hence, it is possible that by improving the system and the experimental task, significant benefits could be obtained. In the last few years gaze-contingent DOF have become a very popular topic [209, 122, 123, 16, 50, 51]. Most of the up to the date studies concentrate on

the user experience and whether or not DOF improves user's satisfaction and performance when gaze-contingent DOF is added to non-stereoscopic and stereoscopic content; results have been mixed.

2.3.6.3 Foveaglyph

Foveaglyph created by Çöltekin and Haggrén [35] is an example of stereoscopic foveation. The idea is similar to the hybrid model described in the previous section. It also uses the depth distance from the user's gaze to select an appropriate LOD level for each pixel. The system takes stereo images as input and returns foveated stereo images. Then the system performs image matching and creates a disparity map. Following this, the system performs a 2D foveation based on an approximation of a human visual system. Since it has to be performed in real time for an individual user only, such a system cannot be generalized to a multiviewer scenario, unless a generic user is established.

2.4 Conclusions

In this chapter, I outlined a variety of domains and applications where GCD have been used and where they could be used in the future. Specifically, I believe the main domains where GCD should be actively used are domains that incorporate VR techniques. In particular, the emphasis should be simulating visual defects in

VR simulators (driving and flight simulators). It also should be for the purpose of improving gaze depth cues in VR settings and finally to the purpose of providing a more accurate multi-sensory experience. I outlined the many challenges and problems that one might encounter when building gaze-contingent applications. In the next three chapters, I will demonstrate different GCD systems that try overcome the challenges outlined in this chapter and to bridge the gaps identified here. My hope is that in a few years, gaze-contingent applications will mature and become a useful tool for a range of problems in the HCI domain.

3 Simulating Visual Defects for Studying Heading and Steering

3.1 Introduction

Gradual and modest decline in visual function is part of healthy ageing. However, neurological diseases such as glaucoma or AMD can produce profound and debilitating effects on visual performance. These diseases can significantly impair people's ability to perform daily activities including their ability to freely and safely move within space. Often, people are not even aware of their condition or do not understand the extent of damage caused by a visual disease that affects them. What is more, the spread of the disease is not uniform and can impact different regions of the retina. A critical feature of diseases such as AMD or glaucoma is the destruction, distortion or degradation of selective regions of the visual field. This chapter examines the role of individual visual field regions and their integration with each other on the ability to judge and control locomotion. Understanding

the mechanisms underlying how the visual system integrates information acquired from different regions of the visual field will provide better diagnostics for identifying people suffering from visual impairments. For instances, Papageorgiou et al. [148] noticed that people with homonymous visual field defects have more fixations in the affected region. The authors suggested that such observation can be used for developing more accurate diagnostic tools. It is possible that similar observation can help in developing tools allowing earlier diagnostic of visual diseases such as AMD. Understanding unusual patterns in the behaviour of such individuals, may also allow us to develop different strategies for them to move within space (as part of rehabilitation). For example, in the reading domain Tarita-Nistor et al. [188] taught patients to read by fixing in the optimal preferred retinal loci and then moving the text relative to their fixation rather than shifting gaze from one word to another as people do in traditional reading. A similar approach can be applied to navigation strategies, possibly by fixating or tracking specific features in the scene whereby people can be more accurate in their navigation strategies. But this must be verified. Therefore, in this chapter, I will present two heading experiments that looked at participants' performance with different gaze strategies (see section 3.5.2.1 and 3.5.3.1).

3.2 Theories in Locomotion

Although locomotion is a basic and integral part of everyday activity for most people, it is a very complex behaviour that is controlled by many neurological and psychophysical mechanisms. One can decompose the task of locomotion by considering what is required in order to reach a desired location (goal). Consider the following scenario: First, a person needs to pick the course or path that she/he will take in order to achieve the goal. Second, the person may have to approach the goal while staying on the path. If she/he do this, she/he will follow the path to reach the initial goal. But how does the person stay on the path? How does she/he know in which direction she/he is actually moving (heading)? In other words, how does a person judge heading direction from observing the sensory consequences of her/his current locomotion? Then how does she/he control her/his motion in the correct direction? For example, how does she/he steer towards their goal, avoid obstacles and correct for deviations from their intended motion? Over the years, people have looked at different theories about perceiving heading and mechanisms underlying the control of steering. Despite considerable research into these questions there is still debate about: (1) the exact mechanism/mechanisms underlying heading perception; and (2) whether these mechanisms are actually used when controlling locomotor behaviours such as steering.

3.2.1 Optic Flow and Heading

Gibson [66] proposed an influential conceptual framework for the visual control of locomotion. He theorized that to control locomotion, humans and other animals rely on information extracted from the *optic flow* (the gradient of optical velocities generated by whole body observer motion). Gibson posited that the active observer uses invariants in the optic flow, such as the focus of expansion (FoE), to directly control locomotion. For instance, when the observer translates the optic flow radiates outward from a FoE lying in the current direction of heading. Thus, the FoE could be picked up from the optic flow and controlled to, in turn, control the direction of locomotion. While optic flow is fixed to the locomotor axis, it is not to be confused with *retinal flow* that depends on both the optic flow and gaze. The FoE is not necessarily visible (even within the field of view) and the FoE of retinal flow does not necessarily correspond to the FoE of the optic flow. Furthermore, judging the optic flow from the retinal flow with a moving eye is not trivial; eye movements add an extra rotational component into the retinal flow field and the zero flow point does not correspond to the heading [157]. Therefore, there have been many proposals and studies of mechanisms to separate optical flow from retinal flow or to extract heading along curved paths [9, 163, 164, 217, 219].

3.2.2 Egocentric Direction cues and Steering points

Many researchers have wondered if it is possible to judge heading or steer in the correct direction without analyzing the optic flow. For example, Rushton et al. [167] asked their participants to walk to a target while wearing base-left or base-right prisms in front of the eyes. By doing so they altered the egocentric direction of objects relative to the locomotor axis, but the FoE relative to the objects in the scene stayed intact. It was predicted that if participants used optic flow to guide their locomotion, then they should compensate for (or rather ignore) the error due to the prisms' deviation and navigate in a straight path to the target. In fact, the participants negotiated a spiral path corresponding to continuous motion in the direction of the target (as perturbed by the prisms), rather than moving directly to the target based on the FoE of the optic flow. Therefore, the authors proposed that locomotion is guided by *perceived visual direction*. That is people can perform locomotion by controlling and monitoring the egocentric visual direction (VD) of targets in the environment, also known as steering points [167, 216]. Simply speaking, this mechanism is based on continuously adjusting motion, so that the target is always in a fixed direction, for example, 0° eccentricity, relative to the driver's body axis. Equivalently this can be described as iteratively adjusting self-motion so that a target is maintained at a fixed angle from the line of motion by

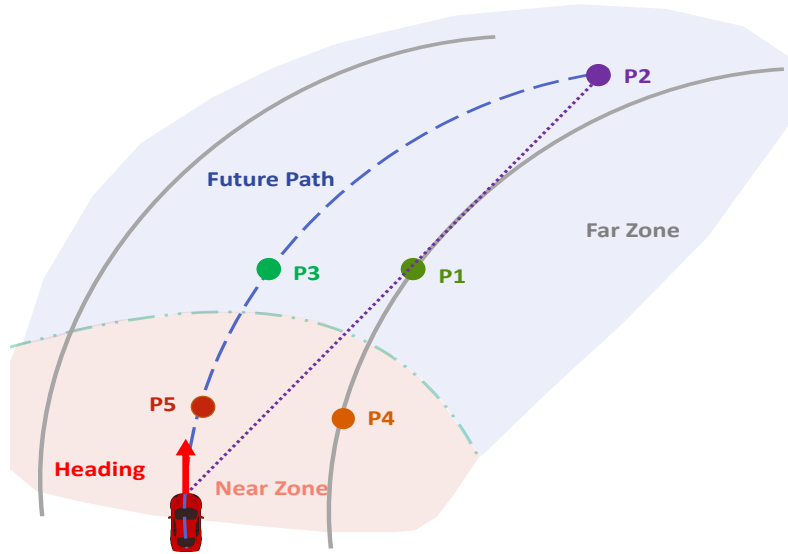


Figure 3.1: Bird's eye view of different steering points and visual regions an observer might potentially use to assist steering on a single lane curved road (indicated by gray lines). P1 — Tangent point on road edge; P2 — A way-point on the future path beyond the tangent point; P3 — A way-point on the future path adjacent to the tangent point; P4 — A near point on the edge of the road edge; P5 — A near point on the future path; Figure adapted from Lappi et al. [103]

moving on an equal-angular spiral [166, 165]. It is possible that more complicated process are involved when the locomotion is extended to manoeuvring a vehicle with restricted turning ability or even worse, manoeuvring such vehicles on a constrained road [214]. In those cases, gaze direction appears to play an important role in starting and executing a curved trajectory.

3.2.3 Gaze and Steering Behaviour

To determine how drivers negotiate a turn one needs the answers to several key questions: What is the optic and retinal flow experienced by people when steering on a curved path? Where do they look throughout their manoeuvre? Do people look at the end goal or do they find sub-goals that, once reached, will lead to their final goal? Do they try to minimize some aspect of their path such as minimizing trajectory length or angular velocity? These questions are complex to answer empirically especially in the context of real world steering behaviour. To answer these questions several theories have been proposed based on observation of gaze behaviour during curve negotiation. These theories have, in turn, been influential in guiding further empirical work. Figure 3.1 shows the locations of possible points of interest when negotiating a turn. Some of these are located on the curve's edges (inside and/or outside), which can help to identify information about the curve (but that the observer should avoid steering toward), other points are on the *future path*, which consists of a set of points that an observer will or would like to pass through. One of the most commonly identified and discussed features on the road is the tangent point (TP), which is located on the inside edge of the curved road, where the observer's line of sight is tangential to the road's edge. In other words it is at the apex of the curve. Several other points presented on the figure are actually

geometrically related to the TP. Specifically, P2 is a waypoint on the future path beyond the tangent point and P3 is a waypoint on the future path adjacent to the tangent point. Lappi et al. [103] provided a thorough review of different strategies that utilize different points on the road. For our present purposes it is sufficient to note that most theories of steering behaviour propose that gaze must be directed to specific locations during the negotiation of a turn. Land and Lee [101] proposed a *tangent point oriented curve negotiation* strategy, which posited that the driver could maintain their steering angle proportional to the gaze angle for successfully steering around a curve. Specifically, the driver should fixate the TP and judge the curvature of the road based on gaze direction relative to the locomotion axis VD_{TP} and lateral distance to the lane edge. Even though there have been numerous studies examining this theory and confirming a general tendency to fixate near the TP, there are still questions about whether drivers might actually use other points in the vicinity of the TP. For example, some studies examined the possibility that drivers might look at a future path point (P2) lying beyond the TP [214]. Although this point is further away than the TP, in terms of visual direction these two points are very close. Another point one might choose to look at is a waypoint on the FP adjacent to the TP (P3) [22]. This type of behaviour is consistent with Boer's *control point theory*, in which drivers need to choose a specific point on their future path trajectory that would allow them to minimize the lateral acceleration required

to steer around the curve and stay within the lane. Similarly, one could use a waypoint on the future path beyond the TP [214] to act as a control point. Finally, some theories are based on utilization of two points on the path for successful steering [101, 168]. All these theories propose that gaze direction is a critical factor. For example, one of the most influential steering studies [101] found that all three of their participants spent 80% of their time looking within 3° of the tangent point.

As all these steering models propose that the driver's gaze is a critical factor, but what would happen if central vision was compromised. Would performance on steering and heading tasks depend on the information acquired at or around the fixation point or could useful information (such as that based on the optic flow) be acquired peripherally? Is it still possible to successfully perform self-motion tasks (such as driving a car) without an intact visual field? What happens when there are anomalies in either the visual field or the field of view of the driver/observer? Would distortions of visual direction have an impact? Would observers modify their gaze behaviour to compensate for central visual field disruption?

3.2.3.1 Visual Defects and their Simulations

Experiments that use controlled visual field modifications allow for quantification of these effects on performance. They could also identify the nature and benefits of behavioral changes that drivers adopt to compensate for the limitations caused by

such impairment or that theoretically could result in improved task performance. Occupational therapists and low vision technology developers could then promote these compensation mechanisms through targeted training and device design. For example, if a disease affects the size of the visual field, more frequent head turns could compensate for the narrow field of view [58]. Therefore, when it comes to the study of visual field functionality, researchers can look at each area of the visual field separately, or find a way to control the stimulation of the entire visual field. One possibility is to study people suffering from visual diseases that alter the visual field such as retinitis pigmentosa [107] or glaucoma [6], which degrade peripheral vision. On the other hand, one can look at AMD patients to assess the effects of central visual field loss. Although conducting experiments with real patients [196] can help identify many interesting gaze patterns, too often, these experiments cannot be replicated because of individual differences in the nature or the extent of the deficits. Hence, another approach is to create virtual simulations of either real or idealized visual defects [118, 59, 210, 3]. The simplest approach is to ask participants to wear goggles that restrict their peripheral field-of-view [192] or to use lenses to alter the entire visual field [167]. Such simulations do not account for eye movements. To resolve this, one might actually ask people to fixate a certain region or perform guided eye-movements. Yet the most natural type of interaction with these simulations is GCD since it allows the user to make natural head and

eye movements while controlling the visual field. With this approach, obtaining realistic simulations entails several requirements. The first requirement is a real-time simulation that can track a driver's eye location and orientation when they move their head or eye. The second is the ability to map a visual deficit to the retinal topology of the driver's eye. Satisfying these requirements allow for an interactive gaze driven system that can update the graphics so that the simulated ocular damage is mapped in real time to the retinal image. In other words, the visual content should change based on how the visually impaired individual might see the real world at every instance.

Many studies simulated visual performance on various tasks when parts of the human visual field were completely missing. However, many visual field deficits are not all or nothing phenomenon and involve degradation or distortion. One particular case is metamorphopsia, which can be modelled as a visual perturbation of the image on the retina that is usually associated with early stages of AMD. In many cases symptoms go unnoticed until AMD progresses into more advanced stages and people can no longer see at all at that region. Despite the importance of such symptoms, it is not clear what happens behaviourally, when some region of visual field, although responsive, acquires visual information that is not congruent with the rest of the visual field.

3.3 Gaze-Contingent System Implementation

3.3.1 Platform

A platform for locomotion experiments has to allow for easy setup of different virtual environments and enable different modes of navigation within those environments. Unity3D was chosen as a development platform, since it provides flexible facilities for virtual scene modelling and rendering, combined with full support for modern shader techniques necessary for high-performance GCD. The platform allowed us to implement different input modalities required for subject input and control including the keyboard and a steering wheel. It also allowed us to incorporate a separate set of scripts that acquired data from the eye tracker, generated rendering content for eye-tracker calibration, and setup of several camera objects that each had their own shaders and scripts responsible for simulating different visual defects for one or both eyes.

3.3.2 Apparatus

Visual displays were generated on a desktop computer with the following specifications: Dell T3500 with NVIDIA Quadro 5000, Windows XP Professional 2002 Service Pack, Intel[®] Xeon[®] CPU, 2.53 GHz, and 3.50 GB RAM. In experiment 2, an EyeLink 1000 was used to track online gaze location and orientation of observers

at 1000 samples per second. The stimulus was presented on a SONY Trinitron Multiscan G420 Cathode Ray Tube (CRT) display monitor (34 cm H by 25.5 cm V, with a pixel resolution of $1024H \times 768V$ and a refresh rate of 120Hz). The high sampling rate of the eye tracker and the high refresh rate of the monitor allowed us to minimize system lag. Our system’s end-to-end latency from eye movement to associated screen change was measured at 39.5 ms.

3.3.3 Defect Simulation

For visual defects, I used methods described in [210, 206]. The visual defects that I studied are presented in figure 3.2. The general idea was to build a system that would use a *resolution image* map to describe an aspect of visual performance for every region of the visual field. The simplest, of course, is the resolution map for occlusions, where one can specify the shape of an occlusion with one color and the unaffected (transparent) regions with another color. For visual perturbation, the image on the resolution map indicates the amount of perturbation. The general idea was built on specifying the direction of distortion with a *resolution image*, which is an image depicting the entire visual field and with each color indicating the amount of perturbation. By using three colors one can describe a 3D distortion with each colour specifying distortion along the corresponding axis. As a result, visual perturbation was achieved by displacing each pixel in the original scene by

the amount indicated in the resolution map, as follows:

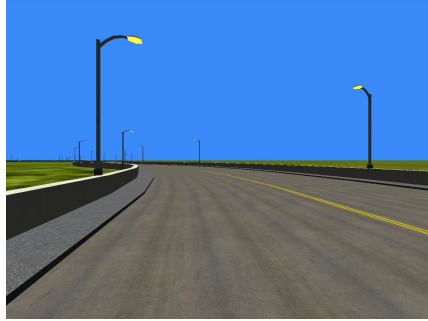
$$X_{new} = X_{Current} + 0.1resolutionMap.R;$$

$$Y_{new} = Y_{Current} + 0.1resolutionMap.G;$$

I tried to module relative large distortions but that would correspond to significant shift between regions in the visual-field. Precisely, for the horizontal distortions, the affected part of the display was shifted to the left by 49 pixels (2.07°), which is close in size to fovea. The Gaussian perturbation used both red and green channels to achieve both horizontal and vertical shifts. Greatest shift was applied to the centre of the distorted region and the degree of shift gradually subsided towards the edges of the distortion region according to a truncated 2D Gaussian profile. The maximum shift was 102 pixels (4.31°) horizontally and 76 pixels (3.34°) vertically, while the minimum shift at the border was 25 pixels (1.057°) horizontally and 18 pixels (0.79°) vertically. Once again, those values fall within the central visual field.

3.4 Use of the System in the Study of Steering Behaviour

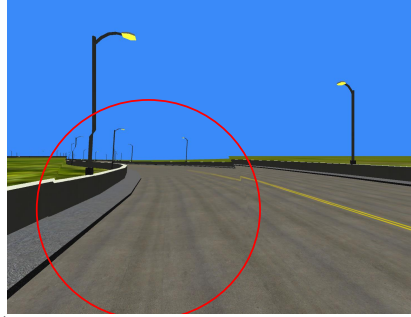
The system was designed to be able to study natural steering behavior in realistic environments with gaze-contingent visual field deficits. For example, in experimen-



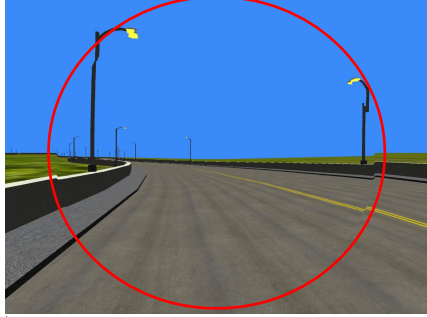
(a) Road with no visual defects



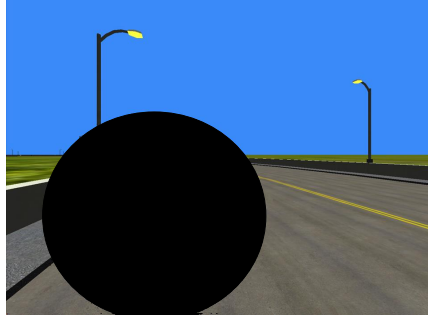
(b) Horizontal perturbation of size 10° .



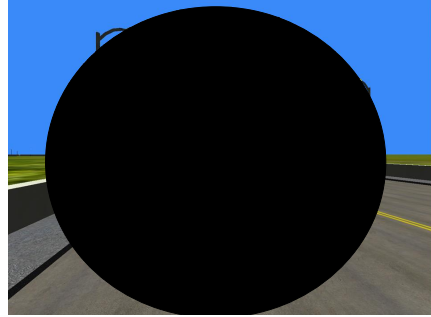
(c) Gaussian perturbation of size 5° .



(d) Gaussian perturbation of size 10° .



(e) Central scotoma of size 5° .



(f) Central scotoma of size 10° .

Figure 3.2: Scenes and visual defects used in the experiment. The blue semi-transparent overlaid circle is for illustration only and is larger than the actual defect to show the distortions that are most visible at the border between the defect and the rest of the visual field.

tal situations people can be instructed to steer on a curved road with different visual defects, specifically with visual perturbations. The system can collect steering behavior concurrently with gaze data allowing for flexible control and study of the

relationship between task (steering) performance and instantaneous gaze. Motor control actions (both steering and gaze) can be expressed in terms of coordinates in the modeled world (global coordinates) or in coordinates centred on the road (global coordinates) as is appropriate (figure 3.3).

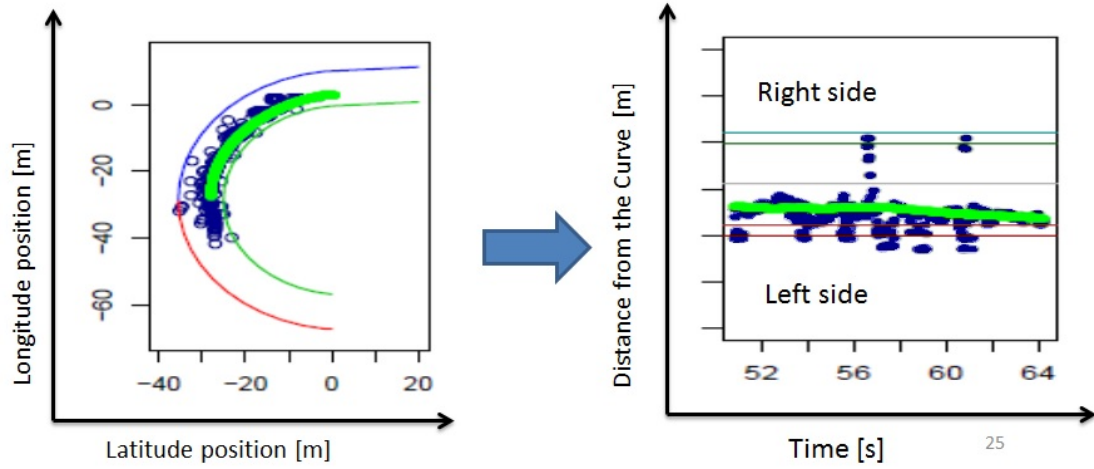


Figure 3.3: An example of driving (green) and gaze (blue line) behaviour during steering on a curve both in global and local coordinates.

3.5 Heading Experiments

Warren and Kurtz [218] looked at the effect of retinal eccentricity (central or peripheral stimulation) and heading eccentricity (Experiment 1: $\pm 1^\circ$, $\pm 2^\circ$, $\pm 3^\circ$, $\pm 4^\circ$; Experiment 2: $\pm 1^\circ$, $\pm 2^\circ$, $\pm 5^\circ$, $\pm 10^\circ$) on heading discrimination. To be precise, they created visual displays that showed visual information only to the central vi-

sual region (circular windows of 10° , 15° , 20° and 25°) or to peripheral regions (the peripheral displays were the converse of the central displays with the inside of the circular region occluded, leaving only dots in the periphery of the $40^\circ \times 32^\circ$ display). They used a forced-choice procedure, where participants had to determine if a vertical target that appeared at the end of each optic flow display was to the right or to the left of the presented heading direction. They observed that heading judgements were more accurate with central visual stimulation (within the central 5°) compared to peripheral stimulation. However, this study did not look at what happens if the occlusions were gaze-contingent and maintained in the same retinal location rather than head based. These issues were addressed by Crowell and Banks [38], who examined the influences of eccentric fixation and translation speed on the ability of observers to discriminate heading direction. Observers viewed movies with different base heading directions (10° , 20° , 40° , 70°) at different speeds (200, 100, 50, 37 cm/sec) and different fixation locations (0° , $\pm 5^\circ$, $\pm 10^\circ$, $\pm 40^\circ$). Pairs of stimuli, one with the base heading eccentricity and the other with a heading offset that was varied according to a staircase psychophysical procedure, were presented and participants indicated whether the heading in the second movie was to the right or left of that in the first. When both the heading direction and direction of fixation were 0° , the threshold was 0.2° ; but at heading eccentricity of $\pm 70^\circ$, the heading discrimination threshold reached 20° . In other words, 20° was, presumably, the

smallest difference between the heading directions in the two sequences that could just be detected. Consequently, the authors concluded that, regardless of the retinal region stimulated (retinal eccentricity), heading judgements were more precise at small heading eccentricities (radial flow field) than at large heading eccentricities (lamellar fields).

Cornelissen and van den Dobbelaars [36] followed up on these earlier studies of the relative effectiveness of central versus peripheral stimulation. In their study the eccentricity of simulated heading through a sparse cloud of dots was continuously varied according to a sum of sinusoids with a peak-peak amplitude corresponding to 70% of the 40° screen width. Unlike the two earlier studies, they asked participants to continuously track the perceived heading direction with their gaze. They claimed this task closely resembles natural behaviour during ongoing locomotion. The coherence between the eye movements and the heading variation was analysed to obtain measures of absolute error and lag between stimulus and response. Retinal stimulation was restricted to the peripheral or central visual field using a gaze-contingent paradigm. In other words, the participant's gaze was used to generate the desired optic flow patterns and also to ensure that portions of the visual field were blocked (simulated scotoma). In addition to controlling the stimulus, the participant's gaze was also the dependent measure in this study. They found that tracking performance improved (accuracy increased and lag decreased)

with an increase in the field size stimulated; that is, increasing the area of either central or peripheral occlusions (tunnel vision) reduced performance. The effects of central occlusion appeared more pronounced but the authors did not attempt to equate the areas of the peripheral and central regions stimulated, or directly compare the effects of peripheral and central stimulation. In a second experiment, they found that tracking response latency was larger for smooth changes compared to step changes in heading direction under the central occlusion conditions, while under the tunnel vision conditions the opposite pattern was found. They concluded that the central retina is more sensitive to smooth motion, while periphery is more sensitive to abrupt change in optic flow.

3.5.1 General Methodology

The two experiments in this chapter extend earlier work on heading perception [36, 218, 38] and I used similar sparse optic flow displays. I augmented these with virtual environments depicting realistic scenes such as an orchard.

To evaluate the impact of visual field regions on heading strategies and performance, I conducted two experiments. The first experiment assessed sensitivity and accuracy of heading judgments as a function of whether the flow field in the vicinity of the future heading point was visible or not. An occlusion was simulated as opposed to a visual field deficit because the occlusion was not gaze contingent.

In the second experiment, I simulated central visual field deficits using GCD. Like Cornelissen and van den Dobbelsteen [36], the visual defects were aligned with the user’s line of sight. However, heading was not continuously varied during the trial nor were participants asked to align their gaze with heading. This avoids confounding task-driven gaze control with the natural gaze behaviour people exhibit in the scenarios.

3.5.2 Experiment 1

Our first experiment replicated and extended the work of Warren & Kurtz [218], who looked at the role of different regions of the visual field in the perception of heading. These experiments, and our attempt at replication, simulated an occlusion of the optic flow in the direction of travel. In the real world, such a situation arises when a near object moving with the head (for example, an object on the dashboard of a car) lies along or near the direction of future travel. If the observer fixates a constant direction then this is equivalent to a scotoma in the visual field but in general this is not the case with a mobile eye. I performed this experiment for two reasons. First, I was interested in dissociating the effects of flow type (radial versus lamellar) and retinal region of stimulation (central versus peripheral). Consequently, I looked at the impact of three independent variables on forced-choice heading discrimination: fixation type, occlusion, and simulated heading direction.

Based on the earlier research, it was hypothesized that the presence of an occlusion would negatively impact the observers' ability to determine his/her heading and this effect would be exacerbated with an increase in occlusion size. I also predicted that observers would be affected less by a central occlusion when the heading direction was more eccentric as they would be able to determine the direction of optic flow by using their peripheral vision.

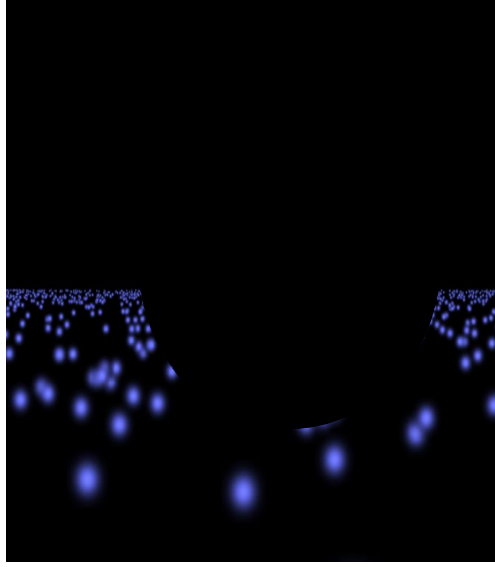
3.5.2.1 Methods

Eight observers participated in this experiment; all were academics or graduate students. Participants ranged in age from 24 to 45 years (31.6 ± 8.6 , $M \pm SD$) and five were female. All participants had corrected distance visual acuity of 20/20 or better and could see clearly at the viewing distance. Written informed consent was obtained from all participants for the study. Observers sat at a distance of 0.24 m away from the display with their head stabilized on a chin-rest, producing a display of $70.62^\circ \times 57.80^\circ$. The display was not gaze contingent and thus participants did not wear an eye tracker.

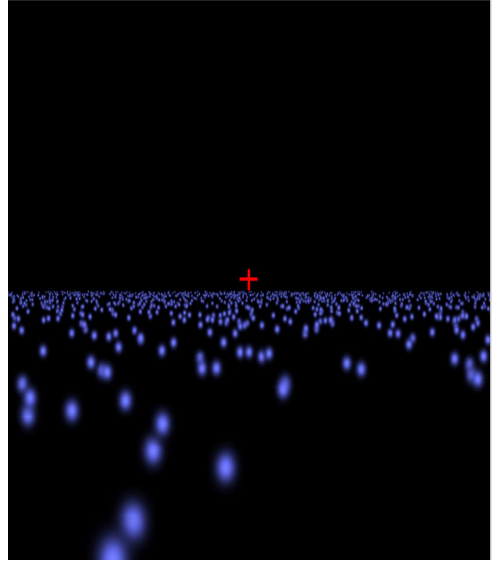
Each trial simulated linear translation over a ground plane that consisted of randomly positioned blue dots extending from the observer to the horizon. The dots were placed against a black background so that only the dots were visible and defined the ground plane (Figure 3.4c). I used random-dot ground plane optic flow

instead of a more naturalistic scene not only for the purpose of being consistent with previous studies but also to eliminate any other cues, such as local expansion that might mislead participants in performing the required task [219]. The speed of the simulated self-motion was always 4 m/s and display motion lasted for a period of 2 s (120 frames). Displays simulated linear self-motion along one of six heading directions, namely $\pm 2.5^\circ$, $\pm 5.0^\circ$, or $\pm 7.5^\circ$ relative to straight-ahead. Within each experimental block the different simulated heading conditions were presented in pseudo-random order. In addition, displays simulated central occlusions of different size (Figure 3.4a) with superimposed black circles that occluded 0° (no occlusion), 2.5° , 5.0° or 10° of the centre of the visual display. Finally, I wanted to determine whether the type of fixation played a role in heading perception. In one condition observers freely viewed the display while in the other condition observers fixated a small red cross ($0.5^\circ \times 0.5^\circ$) placed in the center of the screen (Figure 3.4b). Thus, there were 2 fixation conditions $\times$ 4 occlusion conditions $\times$ 6 simulated heading directions $\times$ 5 repeats for a total of 240 experimental trials. Following each display, a red probe line ($0.1^\circ \times 1.5^\circ$) appeared offset $\pm 0.5^\circ$, $\pm 1.0^\circ$, $\pm 2.0^\circ$, or $\pm 4.0^\circ$ from the simulated heading.

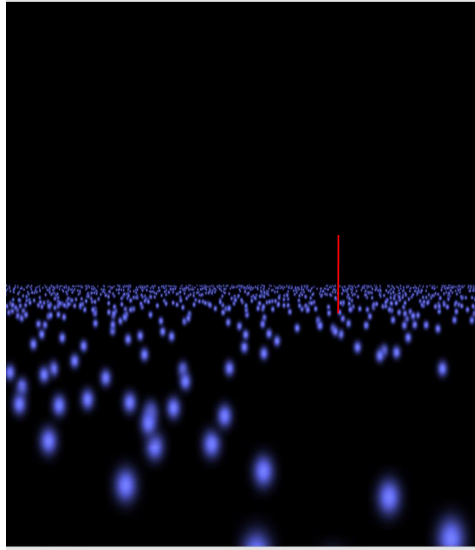
Displays were blocked by fixation mode (free-viewing, central fixation) and occlusion size (0° , 2.5° , 5.0° and 10°). Overall there were eight blocks presented in a different random order for each observer. The whole study took about 1.5 hours to



(a) Random-dot schematic ground plane scene with central occlusion.



(b) Random-dot schematic ground plane scene with central fixation.



(c) Random-dot schematic ground plane scene with a red probe line.

Figure 3.4: Screen captures of the visual displays as seen by an observer in different conditions or at different stages of Experiment 1. The probe line was presented after the optic flow display and observers judged their heading direction relative to the probe.

complete. Participants were provided with short breaks between blocks.

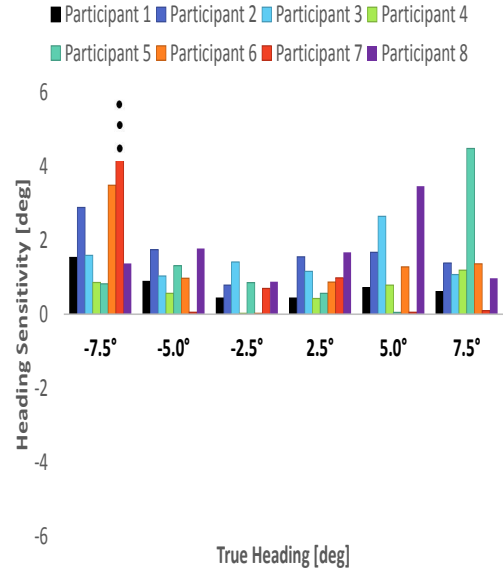
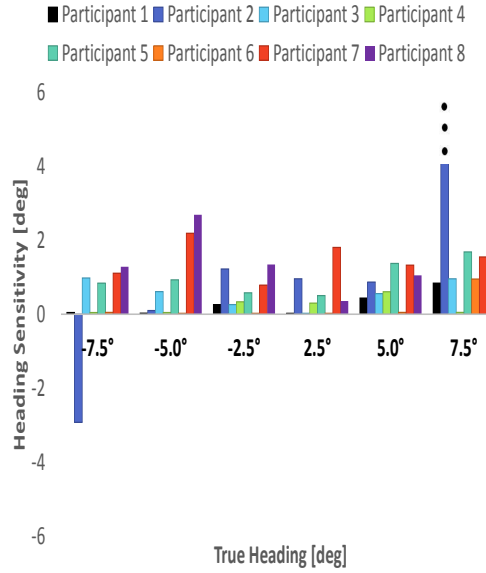
Before the start of the experiment, participants were told that they would view a series of displays simulating heading in different directions. They were notified whether they needed to fixate at the centre of the screen or not on a given block. Within each block, participants viewed displays simulating eight different heading directions paired with eight different probe positions. After display motion ceased, a red vertical probe line appeared at some distance to the left or to the right from the simulated heading. The participants then judged whether this probe appeared to the left or to the right of their perceived heading by pressing either the left or right arrow key on the keyboard. There was no feedback provided about their performance.

3.5.2.2 Results

To calculate a participant’s threshold sensitivity for each condition, psychometric functions as a function of probe offset were obtained. In the absence of bias, if the probe location lay in the simulated heading direction, an equal number of left and right responses was expected. As probe offset from the simulated heading increased, it was expected that participants would become increasingly certain of their responses (e.g., when the offset was far to the left, it was expected that the participant would rarely respond “right”). Thus, as the offset increased from large

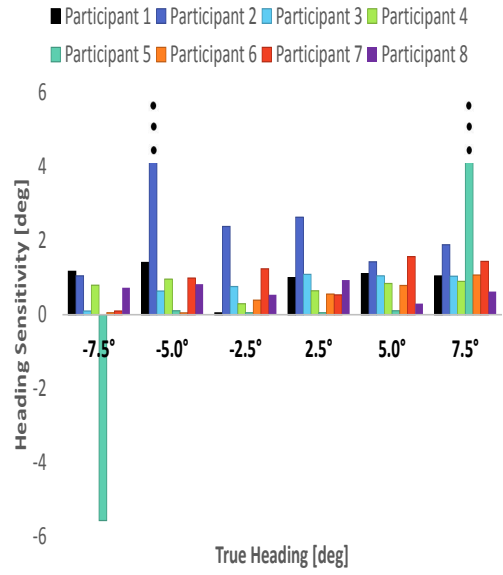
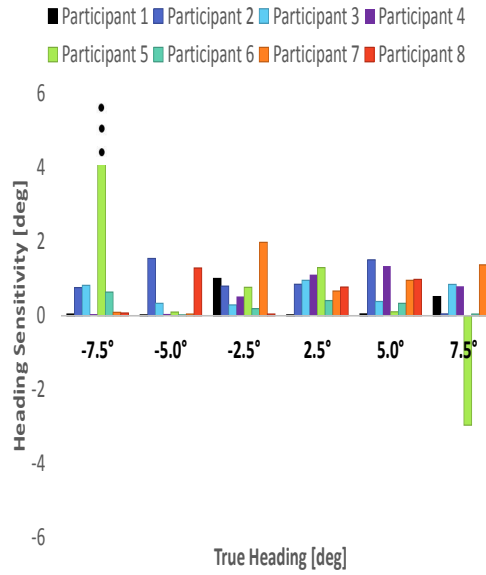
leftward toward large rightward values, it was expected to see an increase in “right” responses from near 0% to near 100%. The resulting S-shaped curve was fitted to a *logit* psychometric function with non-linear regression. The fitted 50% point, or point of subjective equality, was used as a measure of the accuracy of the judgements (bias).

For the threshold, a 75% correct criterion was adopted (in practice estimated here as the difference between the fitted 25% right response and 75% right response offsets divided by 2).



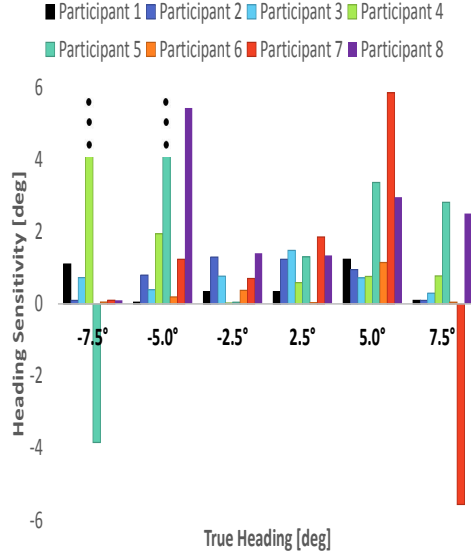
(a) Free viewing with no occlusions.

(b) Fixation with no occlusions.

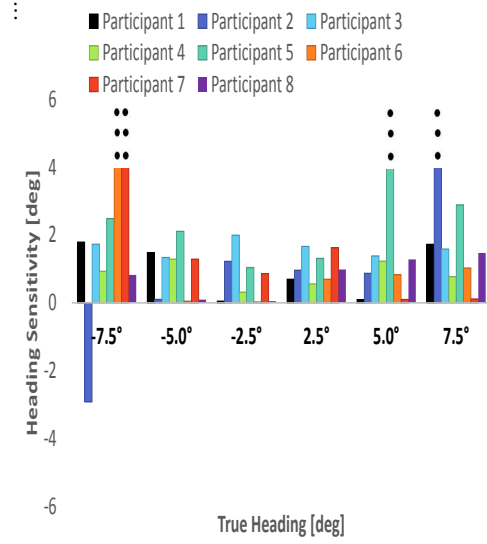


(c) Free viewing with occlusion of size 2.5°.

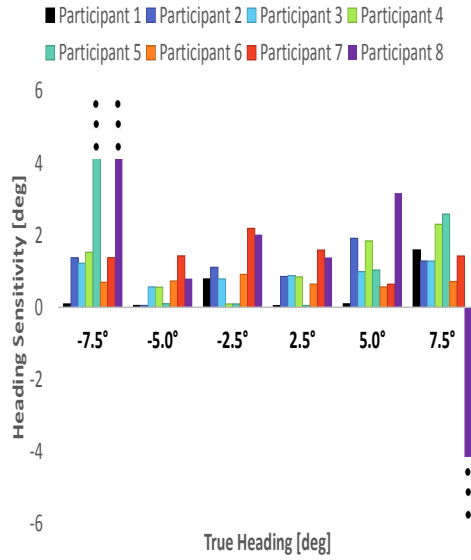
(d) Fixation with occlusion of size 2.5°.



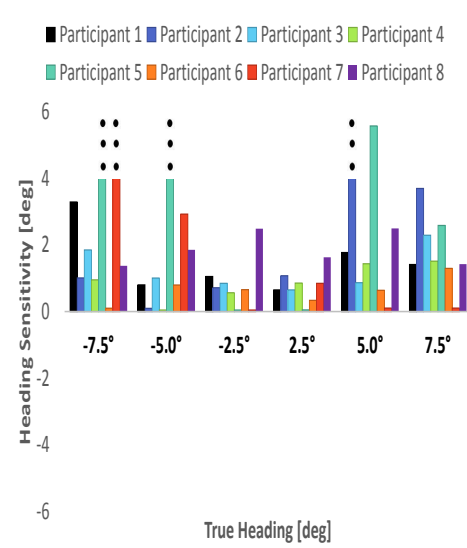
(e) Free viewing with visual occlusion of size 5.0°.



(f) Fixation with visual occlusion of size 5.0°.



(g) Free viewing with occlusion of size 10.0°.

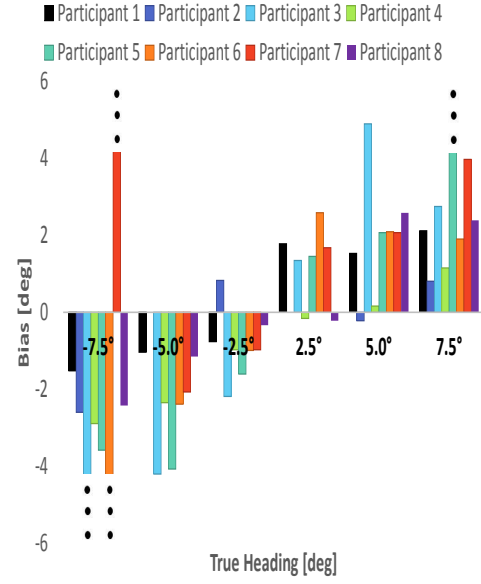
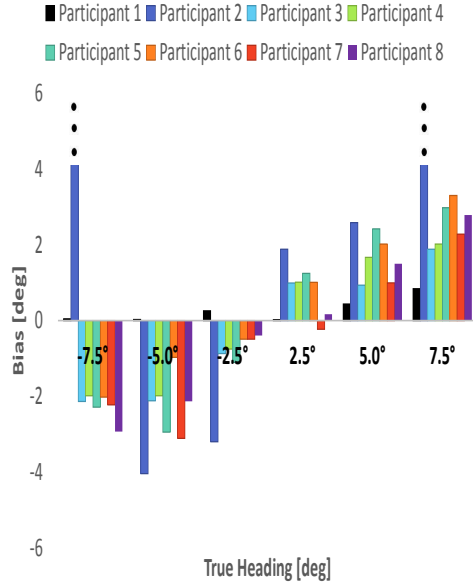


(h) Fixation with occlusion of size 10.0°.

Figure 3.5: Discrimination thresholds as a function of heading directions for each participant for different viewing and visual occlusion conditions.

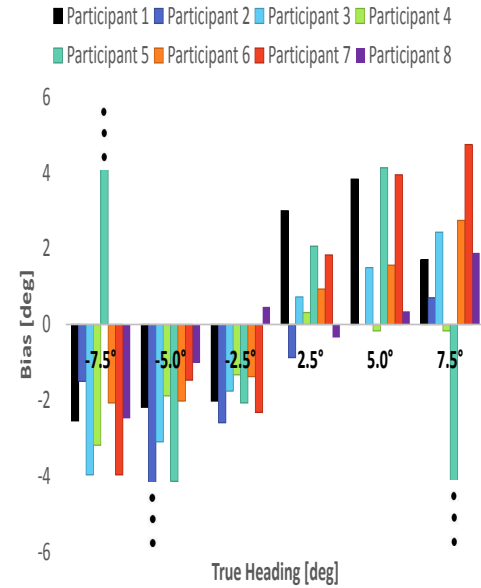
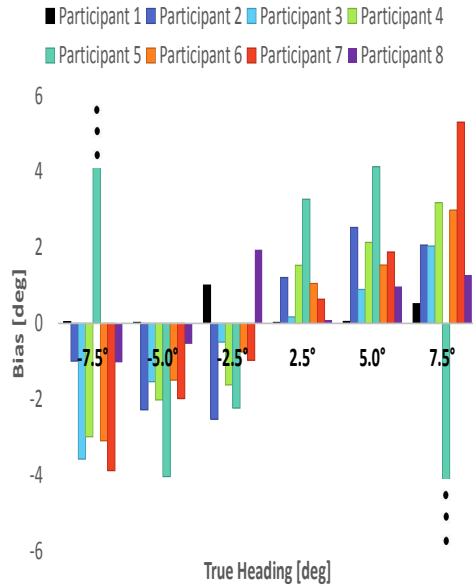
For the baseline condition (central fixation with no occlusion condition) the trimmed mean (17% trimming)¹ for heading discrimination threshold was 1.21° and the trimmed mean for bias was 1.46° . The heading thresholds and biases for all participants as a function of heading eccentricity are presented in figure 3.5 and figure 3.6, respectively. Separate sub-plots show data for each combination of viewing condition (fixation and free viewing) and occlusion size. Also from the figures (3.6 and 3.5), one can see that in some cases performance was extremely poor, so that the participants' heading thresholds and/or biases were very large or impossible to determine (indicated in the figures by three black circles denoting that the value exceeds the chart range). As a result, I could not determine the average across participants. Thus, in order to determine whether these difficulties in discriminating heading depended on any of the independent variables, I analysed the likelihood of the threshold or absolute bias exceeding a criterion of 5° as a function of condition, using logistic regression. The frequencies of thresholds and bias exceeding 5° are illustrated in figure 3.7 as a function of heading direction and occlusion size. It can be observed that the frequency of such large heading thresholds and bias increased with increasing simulated heading direction and increasing occlusion size.

¹The trimmed mean was used to eliminate cases where estimated thresholds were infinitely large from mean calculations



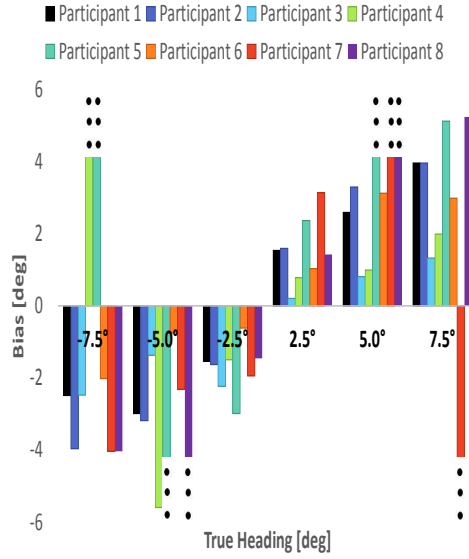
(a) Free viewing with no occlusions.

(b) Fixation with no occlusions.

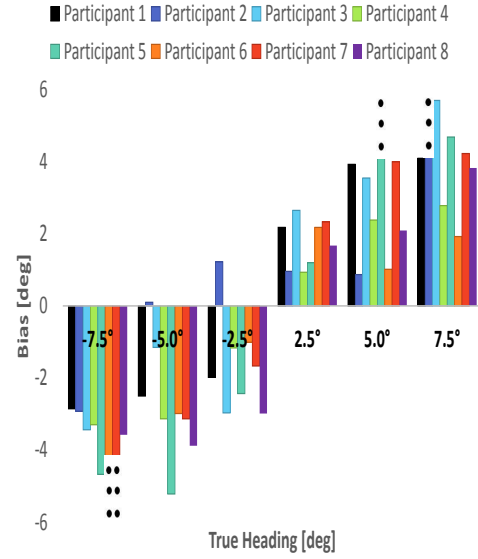


(c) Free viewing with occlusion of size 2.5°.

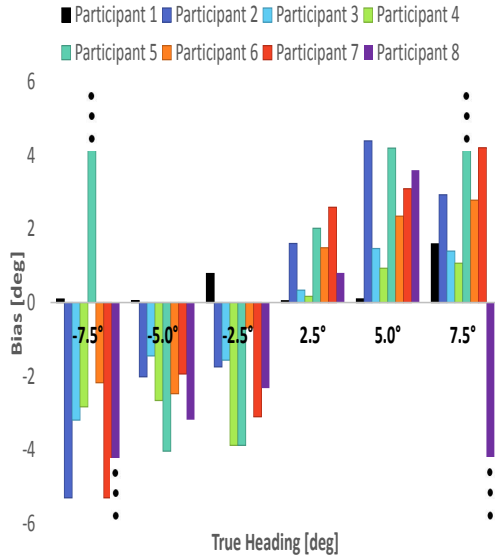
(d) Fixation with occlusion of size 2.5°.



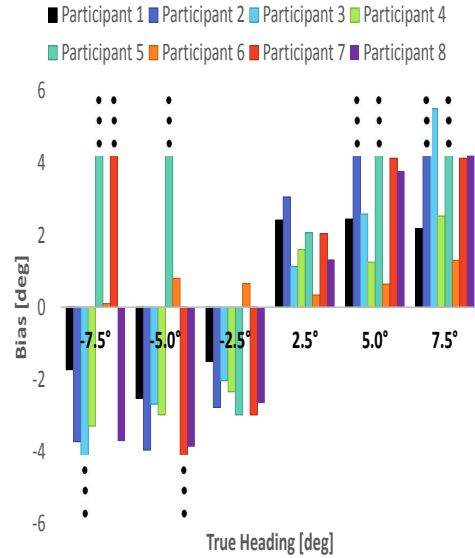
(e) Free viewing with occlusion of size 5.0°.



(f) Fixation with occlusion of size 5.0°.



(g) Free viewing with occlusion of size 10.0°.



(h) Fixation with occlusion of size 10.0°.

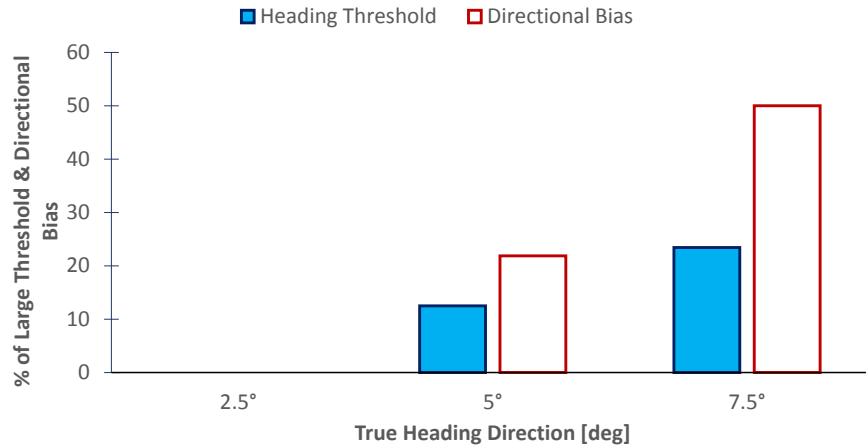
Figure 3.6: Bias as a function of heading directions for each participant for different viewing and visual occlusion conditions.

A stepwise model selection procedure was used to identify model parameters (using the Akaike information criterion, AIC, as a measure of goodness of fit). With this analysis, when I compared free viewing and central fixation conditions, I did not observe significant differences in terms of either heading sensitivity or bias. I collapsed threshold data for simulated heading direction from the right side with simulated heading direction from the left, as the results were symmetric relative to the centre of screen. For the threshold data, the most important factor determining the likelihood of a large threshold was the magnitude of the heading eccentricity ($\chi^2(3) = 18.09, p < 0.001$) followed by the occlusion size ($\chi^2(4) = 4.44, p = 0.04$). Likewise, for the heading bias, the most important factor determining the likelihood of a large bias was the magnitude of the heading eccentricity ($\chi^2(3) = 48.69, p < 0.001$) followed by the occlusion size ($\chi^2(4) = 11.42, p < 0.001$). Inclusion of either the viewing condition or any of the interaction terms did not result in a more parsimonious model and thus the models (for both threshold and bias) selected after the stepwise procedure contained main effects of heading eccentricity and occlusion size. Specifically, Tukey post-hoc analysis with Bonferroni adjustment showed that the frequency of large heading thresholds differed between heading of 7.5° and 5.0° , between 7.5° and 2.5° and between 5.0° and 2.5° (each $p < 0.001$). There was also a significant difference in frequency of large bias between 5° versus 7.5° heading directions with $p = 0.003$; However, there was no significant difference

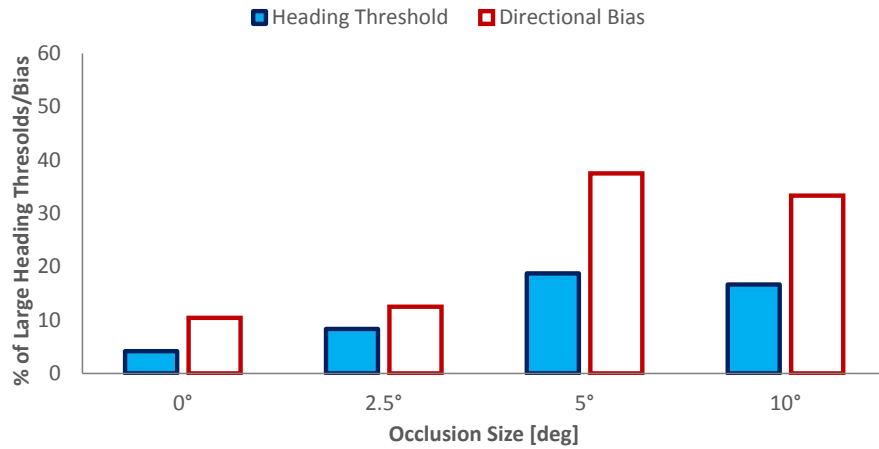
in frequency of large bias between the rest of the heading directions. In term of occlusion size, Tukey post-hoc analysis with Bonferroni adjustment showed that there were significant differences between the no occlusion control condition and all of the occlusion conditions (threshold: 0° versus 2.5° , $p = 0.004$; 0° versus 5° and 0° versus 10° , $p < 0.001$; bias: 0° versus 2.5° , $p = 0.004$; 0° versus 5° and 0° versus 10° , $p < 0.001$). However, there was no significant difference between frequencies at occlusions of sizes for a large heading threshold, but significant differences between different occlusion size for large heading bias (5.0° versus 10° , $p = 0.009$; and 2.5° versus 5° , $p = 0.02$).

3.5.2.3 Discussion

As expected, our results were consistent with Warren & Kurtz [218]. In agreement with the earlier study, heading discrimination threshold and bias increased as the simulated heading direction of the destination point increased relative to straight-ahead. To explain their results, Warren & Kurtz proposed a *functional degradation* hypothesis, that explained heading performance based primarily on the informational quality of the optical flow, modulated by the differential sensitivity of the peripheral and central retina to this information. In other words, observers' heading judgements are poorer, when the simulated heading direction lies in the periphery since this region is normally exposed to locally lamellar retinal flow, and thus is



(a) Frequencies of large heading thresholds and of large directional bias as a function of simulated heading direction collapsed across participant, occlusion size, and fixation condition.



(b) Frequencies of large heading thresholds and of large directional bias as a function of occlusion size collapsed across participant, simulated heading direction, and fixation condition

Figure 3.7: Frequencies of large heading thresholds and of large directional bias as function of simulated heading direction or occlusion size.

tuned to lamellar flow and less sensitive to radial retinal flow. Contrary to this explanation based on retinal flow, I found that the effect of heading eccentricity

was similar under free viewing and fixation conditions suggesting the effect was due mainly to the characteristics of the optic flow and not due to retinal eccentricity. This is consistent with Crowell and Banks work [38], who also noted that people can accurately estimate the direction of lamellar and radial flow in all visual regions. I also looked at occlusions of different sizes to determine the effect of restricting the optic flow to eccentric locations. As predicted, the presence of central occlusion generally made heading judgements more difficult. On the other hand, some participants were not affected by the presence of a central field occlusion (Participant 1 for example). This suggests that at least some people require access to the optic flow in the vicinity of the heading direction to determine an accurate heading direction. It is possible that some people use visual processing from two visual regions (central and peripheral), while people who were not as affected by occlusion size can extract information from only one of these regions. I conducted a control study with two participants, who repeated the experiment while their eye movements were monitored with an eye tracker, and verified that participants did indeed maintain central fixation throughout the fixation trials. Based on this control experiment (and on the stability of gaze observed during the fixation conditions in Experiment 2 below, where eye movements were monitored), we conclude that subjects likely maintained fixation as instructed.

In the next experiment, I examined the effect of retinal eccentricity on heading

perception by utilizing eye-tracking technology, which allowed us to control what people can see in different regions of their visual field while still allowing them to make free eye movements that were recorded.

3.5.3 Experiment 2

Experiment 1 assessed the impact of occluding central portions of the optic flow either with central fixation or with free eye movements. The second study assessed the effects of restricting or distorting information presented to specific regions of the visual field. Specifically, I asked, can heading still be accurately and precisely judged when information provided to the central regions of the visual field is occluded or distorted? To do this I compared heading discrimination performance across six different conditions: no visual field defect and five types of visual defects (5° and 10° horizontal perturbation, 5° and 10° Gaussian perturbation, and 10° scotoma). These visual defects covered the central region of the visual field including the fovea (5°) and the 10° visual defects impacted the parafovea and parts of the perifoveal region as well. It was hypothesized that heading judgements would be impaired by any visual perturbation of the central visual field, including localized uniform horizontal displacement of scene features (i.e., applied only to this region).

In their paper, Warren and Hannon [217], looked at static (when an observer constantly fixates a particular location) and dynamic (when an observer tracks fea-

tures in the scene) heading judgements. They concluded that people could judge heading equally well in the two situations. I was also interested in determining whether this was the case in the presence of simulated visual distortions as well. Therefore, I examined the effects of three types of gaze behaviour — free viewing (when observers could utilize any oculomotor strategy while viewing the self-motion display), directed viewing (when observers had to maintain their fixation on a target that randomly moved between three locations: left, middle and right; and object tracking (pursuit eye-movements were used to track a target object lying on the ground plane during the simulated self-motion) — on the accuracy and precision of heading judgements. I hypothesised that visual distortions would have a detrimental effect on heading judgements, but people would be able to compensate for effects of perturbation with eye-movements that would allow them to sample different regions and hence gather a more complete retinal flow over time. Hence, I expected reduced performance in the presence of gaze-contingent distortion, particularly, if gaze behaviour was restricted.

Finally, I also examined whether performance was affected by the realism of the simulated environment. Specifically, I compared heading judgements made with naturalistic and schematic self-motion displays (the latter simulating self-motion over a ground plane composed of random dots as in Experiment 1). Although in both cases there was looming of scene features, in the first scene there was also

an explicit horizon line and ground plane texturing, while the schematic scene did not have an explicit horizon or clearly defined ground plane. These ground plane differences have been shown to play an important role in determining the apparent glide-slope in landing studies [109, 147]. Therefore, I was interested to determine whether these differences also played an important role in determining lateral heading during simulated travel across the ground. Furthermore, I expected that natural scenes would provide more useful and less ambiguous information and hence support more accurate and precise heading discrimination.

3.5.3.1 Methods

Four observers (academics and graduate students) participated in this experiment. They ranged in age from 28 to 45 years; one was male. All participants had corrected distance visual acuity of 20/20 or better and could see clearly at the viewing distance. Written informed consent was obtained from all participants.

A real-time gaze-contingent system incorporated an EyeLink 1000 (SR Research Ltd) eye-tracker, which monitored the observer's free-head gaze behaviour. Observers sat 0.45 m away from the display, which subtended 41.39° . The experiment was conducted in a darkened room with only the display visible.

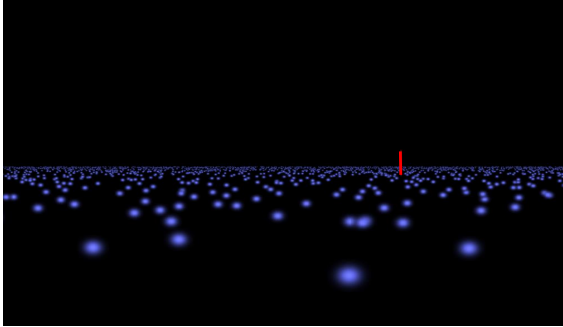
During each trial I either simulated the ground plane used in Experiment 1 (Figure 3.8a) or a natural scene, which included a textured (grass) ground plane,

blue sky in the background and an array of trees equally spaced throughout the scene (Figure 3.8b). As in Experiment 1, each display simulated self-motion over the ground at 4 m/s in one of six possible heading directions ($\pm 7.5^\circ$, $\pm 5.0^\circ$, $\pm 2.5^\circ$). Each trial consisted of 240 frames (display motion lasted for 2 s).

To examine the effects of simulated visual defects on performance, I tested six different conditions: a scene without any visual defect, a central scotoma of size 10° , Gaussian perturbations of size 5° and 10° (Figure 3.8c), and horizontal perturbations of size 5° and 10° .

There were three different viewing conditions: free-viewing, directed looking (defined fixation location) and tracking of a feature in a scene. For directed looking I displayed a red particle subtending $2^\circ\text{H} \times 1.5^\circ\text{V}$, which randomly alternated in position from the center, to two locations to the left or to the right of centre ($\pm 15^\circ$). The particle would disappear and move to a new location every 667 ms, but was stationary between jumps. For the tracking condition, I added a red particle (initially subtending $2.5^\circ\text{H} \times 1^\circ\text{V}$ but looming as it approached the observer) on the ground plane that moved with scene, that is, radiating out from the direction of heading. The particle was similar in size to the blue particles in the schematic ground plane condition. The particle to be tracked always appeared initially at 3.8° below the centre of the screen and continued moving until it reached the end of the visible plane and then restarted at the original position. As before, following each

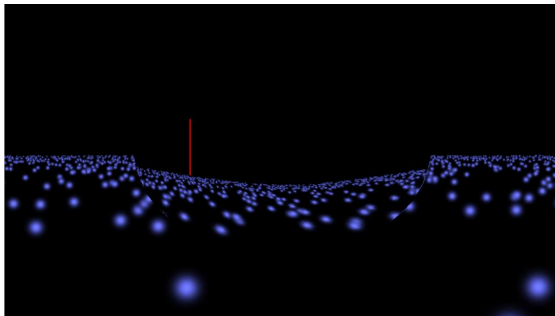
display, a red probe line of length of 5.5° appeared displaced $\pm 0.5^\circ$, $\pm 1.0^\circ$, $\pm 2.0^\circ$, $\pm 4.0^\circ$ from the simulated heading.



(a) Random-dot schematic ground plane scene.



(b) Orchard scene.



(c) Random-dot schematic ground plane scene with Gaussian distortion.

Figure 3.8: Scenes used in Experiment 2.

3.5.3.2 Task

Thresholds were measured in separate blocks of trials for each condition: viewing mode (free-viewing, directed looking and tracking), visual defect type (no visual defect (N), horizontal perturbation of 5° (H5) and 10° (H10), Gaussian perturbation of 5° (G5) and 10° (G10), 10° scotoma (S10)) and environment type (random-dots plane or orchard). Overall there were 36 blocks and they were presented in a different random order for each observer.

Before the beginning of the experiment, participants were told that they would view a series of displays simulating heading in different directions. They were told which viewing mode should be used for each block. Before starting each session participants went through a procedure to calibrate the eye-tracker. The session was also stopped midway for an additional calibration to make sure that there was no significant drift and the calibration was still accurate. As in Experiment 1, within each block participants viewed displays simulating eight different heading directions and eight different probe positions with four repetitions each for a total of 192 trials for each session. After the display motion ceased, a red vertical probe line appeared at some distance to the left or to the right of the simulated heading. The participants then judged whether this probe appeared to the left or to the right of the perceived heading by pressing either the left or right arrow key on the

keyboard. There was no feedback provided.

3.5.3.3 Results

As in the previous experiment, threshold sensitivity for each condition and observer was calculated by constructing psychometric functions as a function of probe offset and fitting a *logit* function using non-linear regression. From these fits I obtained bias and threshold estimates as in Experiment 1. As in the other experiment, some of the thresholds and biases were extremely large. Thus, to determine whether these difficulties in discriminating heading depended on any of the independent variables, I analysed the likelihood of the threshold or absolute bias exceeding a criterion of 5° as a function of condition using logistic regression. A stepwise model selection procedure was used to identify model parameters (using the Akaike information criterion, AIC, as a measure of goodness of fit). With this analysis, when I compared the ground plane with random dots and natural scene conditions, I did not observe significant differences either in terms of heading sensitivity or bias.

As in Experiment 1, I collapsed data for simulated heading direction from the right side with simulated heading direction from the left, because the results were symmetric relative to the centre of screen. For the threshold data, the most important factor determining the likelihood of a large threshold was the magnitude

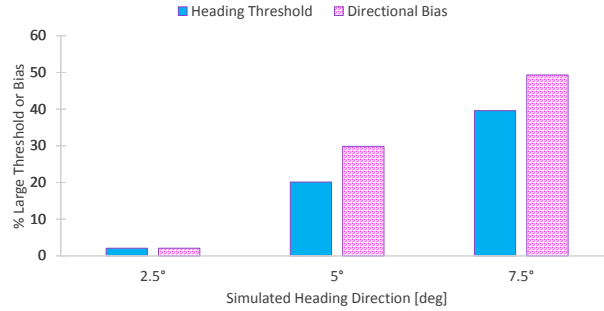
of the simulated heading direction ($\chi^2(3) = 63.50, p < 0.001$) followed by the viewing condition ($\chi^2(5) = 92.08, p < 0.001$) followed by visual defect (type and size) ($\chi^2(9) = 104.12, p < 0.001$). Similarly, for the bias data, the most important factor determining the likelihood of a large bias was the magnitude of the simulated heading direction ($\chi^2(3) = 78.1, p < 0.001$) followed by the viewing condition ($\chi^2(5) = 94.78, p < 0.001$) followed by visual defect (type and size) ($\chi^2(9) = 111.17, p < 0.001$). Inclusion of either the type of environment (schematic ground plane versus natural scene) or any of the interaction terms did not result in a more parsimonious model, and thus, the models selected after the stepwise procedure contained main effects of simulated heading direction, viewing condition and visual defect.

The frequencies of large heading thresholds and directional bias as a function of simulated heading direction collapsed across participant, environment, visual defect, and viewing condition are presented in figure 3.9a. Tukey post-hoc analysis with Bonferroni adjustment showed that the frequency of large heading thresholds differed between simulated heading direction of 2.5° and the two larger eccentricities (5.0° and 7.5° , $p < 0.001$). The difference in frequency of large heading threshold between eccentricities of 5.0° and 7.5° was not significant ($p = 0.08$). Tukey post-hoc analysis with Bonferroni adjustment showed a significant difference in the frequency of bias greater than 5.0° between simulated heading of 2.5° and other

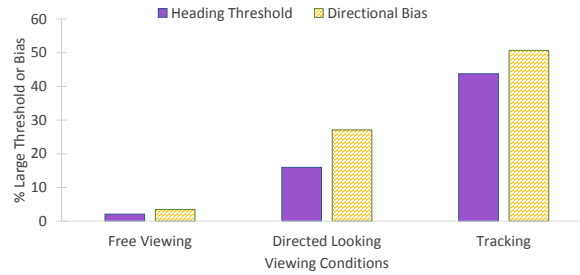
simulated headings (5.0° and 7.5° , $p < 0.001$) but no significant difference between 5.0° and 7.5° .

The frequencies of large heading thresholds and of large directional bias as a function of viewing condition, collapsed across participant, environment, visual defect, and simulated heading are presented in figure 3.9b. From the chart one can see that participants had the fewest number of large heading thresholds and large directional biases for the free viewing condition. Tukey post-hoc analysis with Bonferroni adjustment showed a significant difference in likelihood of large heading threshold between free-viewing and both directed viewing and tracking ($p < 0.001$). However, the difference between directed viewing and tracking was only marginally significant ($p = 0.05$). A similar pattern was observed with the large directional bias, where the main difference was between free viewing condition and the other two viewing conditions ($p < 0.001$). Because the viewing conditions were so different, I examined the eye-data collected during the trials and these observations are reported in the next section.

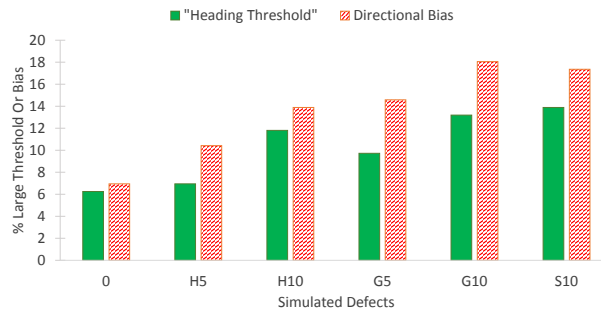
The simulated visual field defects also influenced the participants' threshold sensitivity, however the pattern was complex. When I compared the no defect condition with all other conditions, Tukey post-hoc analysis with Bonferroni adjustment showed that the difference was significant ($p < 0.001$) for all comparisons except for H5 ($p = 0.002$). However, there was no significant difference between the



(a) Frequency of large heading thresholds and of large directional bias as a function of simulated heading direction collapsed across participant, environment, visual defect, and viewing condition.



(b) Frequency of large heading thresholds and of large directional bias as a function of viewing condition collapsed across participant, simulated heading direction, environment and visual defect.



(c) Frequency of large heading thresholds and of large directional bias as a function of scotoma condition collapsed across participant, simulated heading direction, environment and fixation condition.

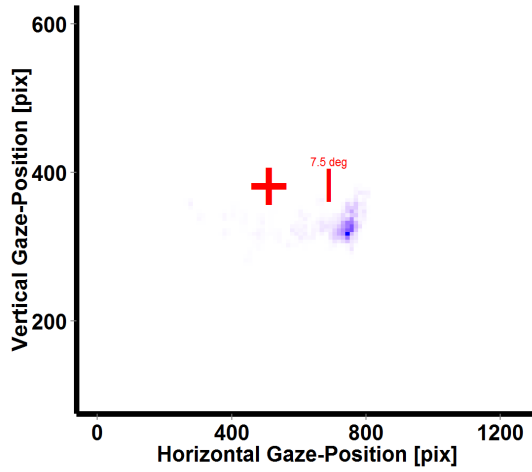
Figure 3.9: Frequency of large heading thresholds and of large directional bias.

individual defects. As with thresholds, Tukey post-hoc analysis with Bonferroni adjustment showed a significant difference in frequency of large directional bias between the no defect control and all the other defects ($p < 0.001$), but there were no significant differences between defect types.

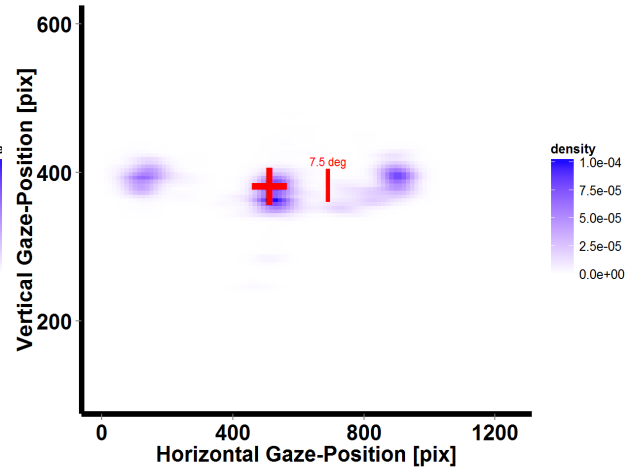
Eye-Data Analysis An example of a heat-map for raw gaze positions for Participant 1, when exposed to 7.5° simulated heading direction for all three viewing conditions is presented in figure 3.10.

For convenience of data analysis, I clustered raw gaze positions based on their proximity along both the x- and y-axes. I then looked at different characteristics of each cluster (number of samples per cluster's area (NC) , major and minor axes of best fit covariance ellipsoid. The gaze patterns presented in the example are representative of other simulated headings (shifted appropriately) for this participant and for other participants. Finally, I ran repeated-measures ANOVA analysis to assess the influence of the independent variables on these gaze measures (Table 3.1).

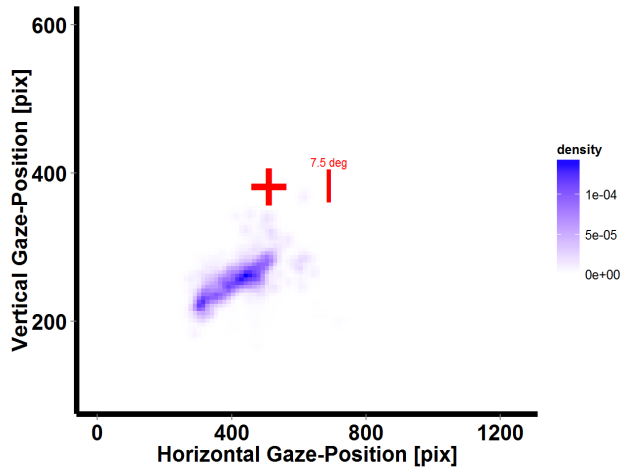
From Table 3.1, one can see that the viewing condition was an important factor for the determining all of the measures listed: number of gaze clusters, the duration of each cluster, cluster density, the length (major axis), and width (minor axis) of the gaze clusters. Consequently, it is possible to conclude that the pattern of gaze



(a) Free viewing condition.



(b) Directed looking condition.



(c) Tracking condition.

Figure 3.10: Heatmaps for raw gaze positions for Participant 1 pooled across all trials for 7.5° heading eccentricity for all three viewing conditions. These examples were for the no defect condition. The intensity of the blue color specify the density of raw gaze samples at each location.

varied with viewing condition as expected. Specifically, it can be observed that during the free viewing condition, people mainly looked at a horizontal eccentricity

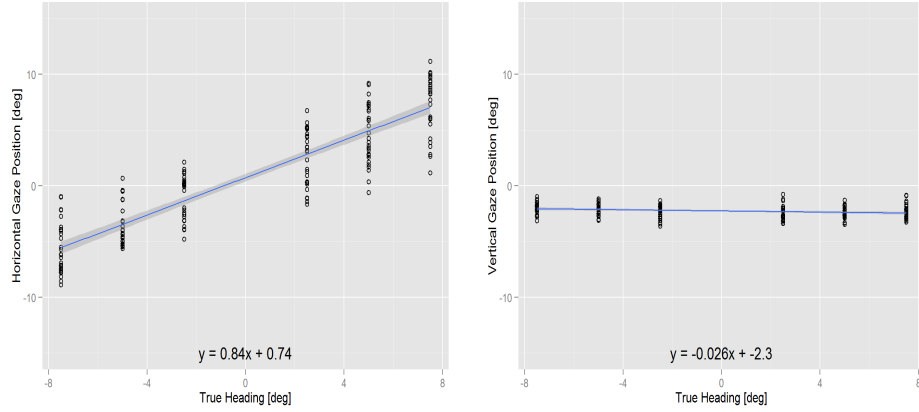


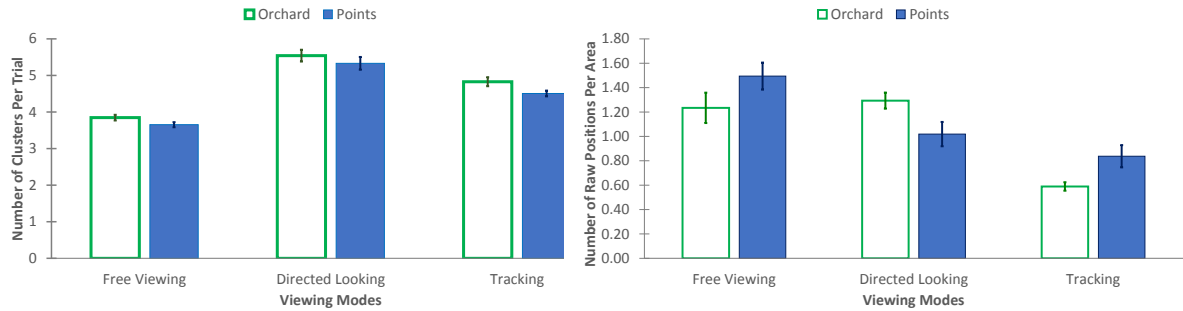
Figure 3.11: Participant's 1 gaze position for simulated heading locations during free viewing conditions for the random-dot scene.

			<i>F</i>			
			Clusters			
	df1	df2	NC	Density	Major Axis	Minor Axis
Viewing	3	2	128.67***	28.66***	56.91***	32.77***
Viewing:Environment	3	2	0.22	6.04**	7.13**	5.61**
Viewing:Simulated Heading	3	10	0.81	0.21	1.06	0.18
Environment:Simulated Heading	3	5	0.29	0.11	0.16	0.01
Viewing:Environment:Simulated Heading	3	10	0.24	0.09	0.07	0.07

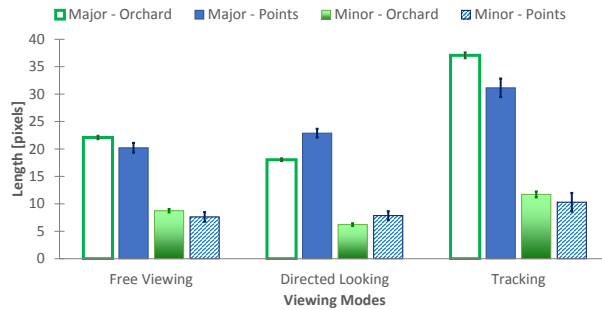
Table 3.1: Repeated measures ANOVA analysis of the independent variables on gaze measures for no defect conditions for individual clustered raw gaze positions. Significance of each contribution is shown by stars ($p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

near the simulated heading direction and their vertical gaze was near the horizon line (figure 3.11). This suggests they were attempting to fixate near the FOE of the optic flow.

The directed looking pattern showed that observers did initiate saccades and



(a) Mean number of cluster per trial across viewing modes. (b) Mean density across viewing conditions.



(c) Mean cluster sizes (major and minor axes) across viewing conditions for different environments.

Figure 3.12: Different gaze cluster characteristics averaged across participants for all viewing conditions for no defect trials. The error bars represent the standard error of the mean for each condition.

fixated at the shifting targets. A typical example of gaze behaviour for the directed looking case is shown in figure 3.10b. One can see that most of the gaze data fell somewhere near the static fixation target but participants often made additional eye movements to the location of simulated heading (note the tendency to occasionally

look in the heading direction). This is also supported by figure 3.12a, which shows that the mean number of clusters was 5.4 (highest of the viewing modes) and not 3 as was expected from participants shifting their gaze strictly from one fixation point to another.

As a next step, I verified that people indeed tracked the moving target in the tracking condition. This is reflected in the fact that the tracking condition had the lowest number of samples per cluster's area (figure 3.12b) and the largest major axis per cluster (figure 3.12c). It is important to note that the results from table 3.1 indicate that the environment (random-dots versus orchard scene) was a significant factor for the number of fixations during a trial, however this effect was very small (random points: $M = 4.50, SD : \pm 0.89$; Orchard: $M = 4.74, SD : 0.93 \pm$). In addition, there was an interaction between viewing and environment for the size of the covariance ellipse (length of the major and minor axis) and cluster's density. The interaction is shown in Figure 3.12c and Figure 3.12b. One can see that the clusters were smaller in the orchard than in the dots condition during directed looking, where in the other two cases (free viewing and tracking) the clusters were bigger in the orchard conditions. A similar pattern was found for density, where the point scene was less dense than the orchard in the direct looking condition, but was denser for the other two viewing conditions.

Finally, I looked at the overall ROI across all conditions per participant. In other

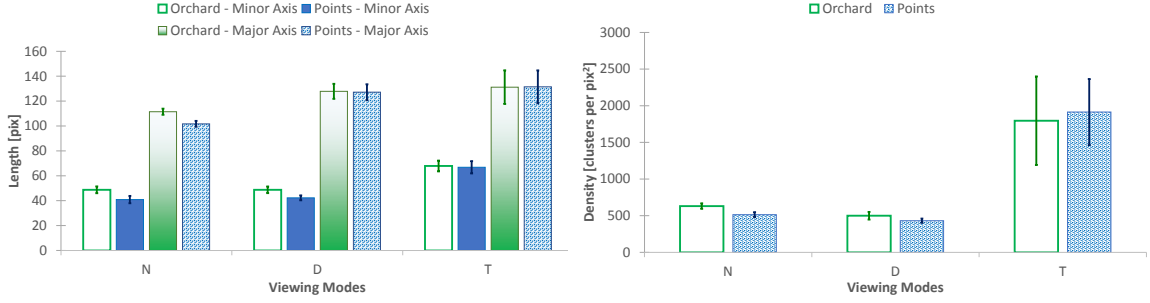
words, I clustered all small clusters from each trial together and looked at the overall shape of these regions. The results of repeated-measures ANOVA are presented in Table 3.3). From the table and figure 3.13, one can observe that viewing modes were significantly different in number of ROI, the length of major and minor axis of each ROI and the number of small gaze clusters per region of interest, which is consistent with the results of the small cluster data. Specifically, the Tukey post-hoc analysis with Bonferroni adjustment showed that the length of the ROI minor axis was significantly larger for the tracking condition than the directed looking and free viewing ($p < 0.001$), while the major axis was significantly smaller for the free viewing conditions than the other two viewing modes (free viewing versus directed looking $p < 0.001$ and free viewing versus tracking $p < 0.05$). This supports the notion that participants focussed their attention more when they were free to choose their viewing strategy. Interestingly, the number of looked only at the no defect condition clusters per ROI was the highest for the tracking condition ($p < 0.01$), indicating that participants indeed tracked the red particle and had only one large region of interest. In addition, the number of small clusters per ROI was significantly smaller for directed looking than for free viewing ($p < 0.05$) reflecting the fact that participants had to divide their gaze attention in the directed looking condition. Finally, the number of regions of interest was significantly affected by all factors included in this experiment (Table 3.3).

	<i>F</i>					
	df1	df2	NC	Minor Axis	Major Axis	Density
Viewing	3	2	15.57***	28.05***	6.01**	13.06***
Environment	3	1	4.08*	3.32.	0.29	0.007
Simulated Heading	3	5	1.98.	0.75	0.84	0.45
Viewing:Environment	3	2	5.06**	0.55	0.26	0.08
Viewing:Simulated Heading	3	10	2.56**	0.48	0.60	0.29
Environment:Simulated Heading	3	5	1.60	0.13	0.19	0.45
Viewing:Environment:Simulated Heading	3	10	2.83**	0.21	0.31	0.46

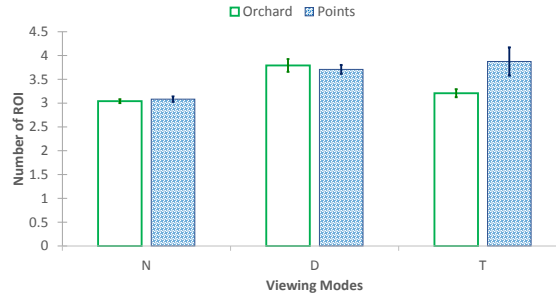
Table 3.2: Two-way repeated measures ANOVA for mean ROI across all defect conditions for no defect condition. Significance of each contribution is shown by stars ($p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

To assess how the simulated defects influenced gaze, I performed a detailed analysis for the free-viewing condition across all simulated visual defects. For this analysis, I looked only at the no defect condition (N) and 10° defects (H10, G10 and S10). Examples of the gaze patterns observed are presented in figure 3.14. Once again, the figure 3.14 show an example of the most attended area of interest with heat-map for Participant 1 for the four visual defects.

Repeated-measures ANOVA analysis showed that simulated visual defects had a significant effect for all measured parameters. Furthermore, there was a significant effect of environment type on minor axis of the bounding ellipse and significant interactions between the simulated defect and the environment for number of clusters per trial and density of raw gaze points in each cluster (Table 3.3).



(a) Mean cluster sizes (major and minor axes) across viewing conditions. (b) Mean number of small gaze clusters per area of ROI across viewing conditions.



(c) Mean number of ROI across viewing conditions.

Figure 3.13: Different ROI characteristics averaged across participants for all viewing conditions during no defect condition. The error bars represent the standard error of the mean for each condition.

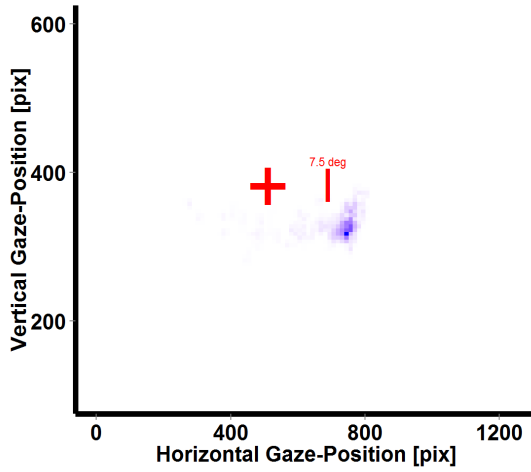
Consequently, I performed a Tukey post-hoc analysis with Bonferroni adjustment to examine specific relations between visual defects. In terms of number of clusters per condition (NC), the no defect (N) condition had significantly fewer regions of interest (clusters) than the Gaussian distortion (G10) and Scotoma (S10)

conditions ($p < 0.001$). Similarly horizontal distortion (H10) had fewer regions of interests than G10 ($p < 0.001$) and S10 ($p < 0.05$) (Figure 3.16b). In terms of density, the no defect had significantly lower density relative to the distortion conditions (Figure 3.16c). Interestingly, the Scotoma condition also had significantly fewer raw gaze positions relative to horizontal ($p = 0.03$) and Gaussian distortions ($p = 0.4$). In terms, of clusters' size (length of the major and minor axes), Gaussian distortion had significantly smaller major axis than the no distortion ($p = 0.04$) and horizontal distortion ($p = 0.03$) conditions and marginally smaller axes than S10 ($p = 0.07$). Finally, minor axis was significantly smaller in the N condition ($p < 0.001$) than in the S10 condition ($p < 0.001$).

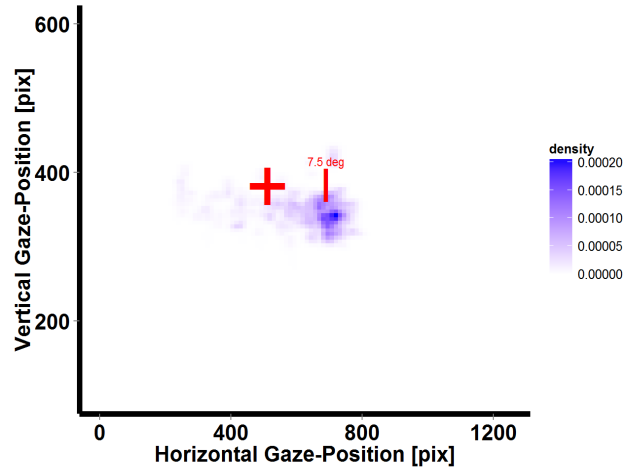
	<i>F</i>					
	df1	df2	NC	Density	Major Axis	Minor Axis
Visual Defect	3	2	22.27***	8.34***	6.62***	7.71***
hline Environment	3	1	0.03	2.85.	0.38	5.68*
Simulated Heading	3	5	0.78	0.86	0.90	0.59
Visual Defect:Environment	3	2	8.11***	2.88*	0.43	1.30
Visual Defect:Simulated Heading	3	10	0.26	0.13	0.18	0.18
Environment:Simulated Heading	3	5	0.40	0.11	0.05	0.04
Visual Defect:Environment:Simulated Heading	3	10	0.48	0.12	0.24	0.25

Table 3.3: Two-way repeated measures ANOVA for mean gaze cluster across all defect conditions for free viewing condition. Significance of each contribution is shown by stars ($.p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

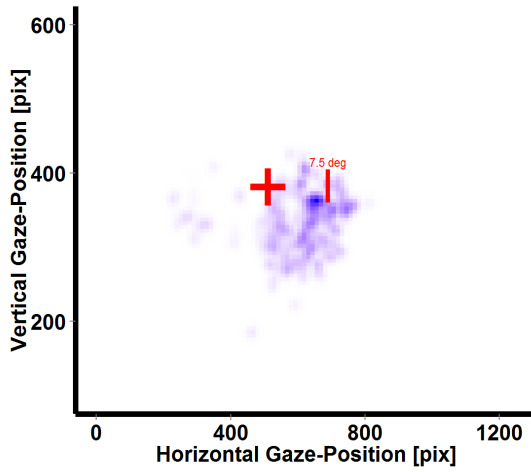
Finally, I looked at the overall ROI across all conditions per participant. In



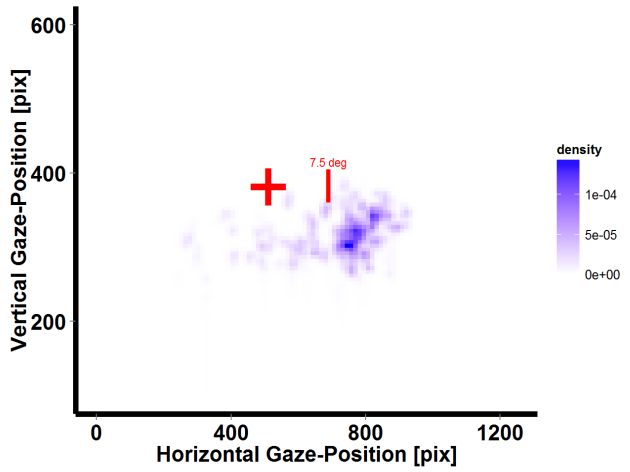
(a) No defect condition.



(b) Horizontal 10° distortion condition.



(c) Gaussian 10° distortion condition.



(d) Scotoma 10° condition.

Figure 3.14: Heat map for raw gaze locations for participant 1 pooled across all trials for 7.5° heading eccentricity for four visual defect conditions. These examples were for the free viewing condition. The color intensity specifies the raw sample density for each location.

other words, I clustered all small clusters from each trial together and looked at the overall shape of these regions. The results of repeated-measures ANOVA are

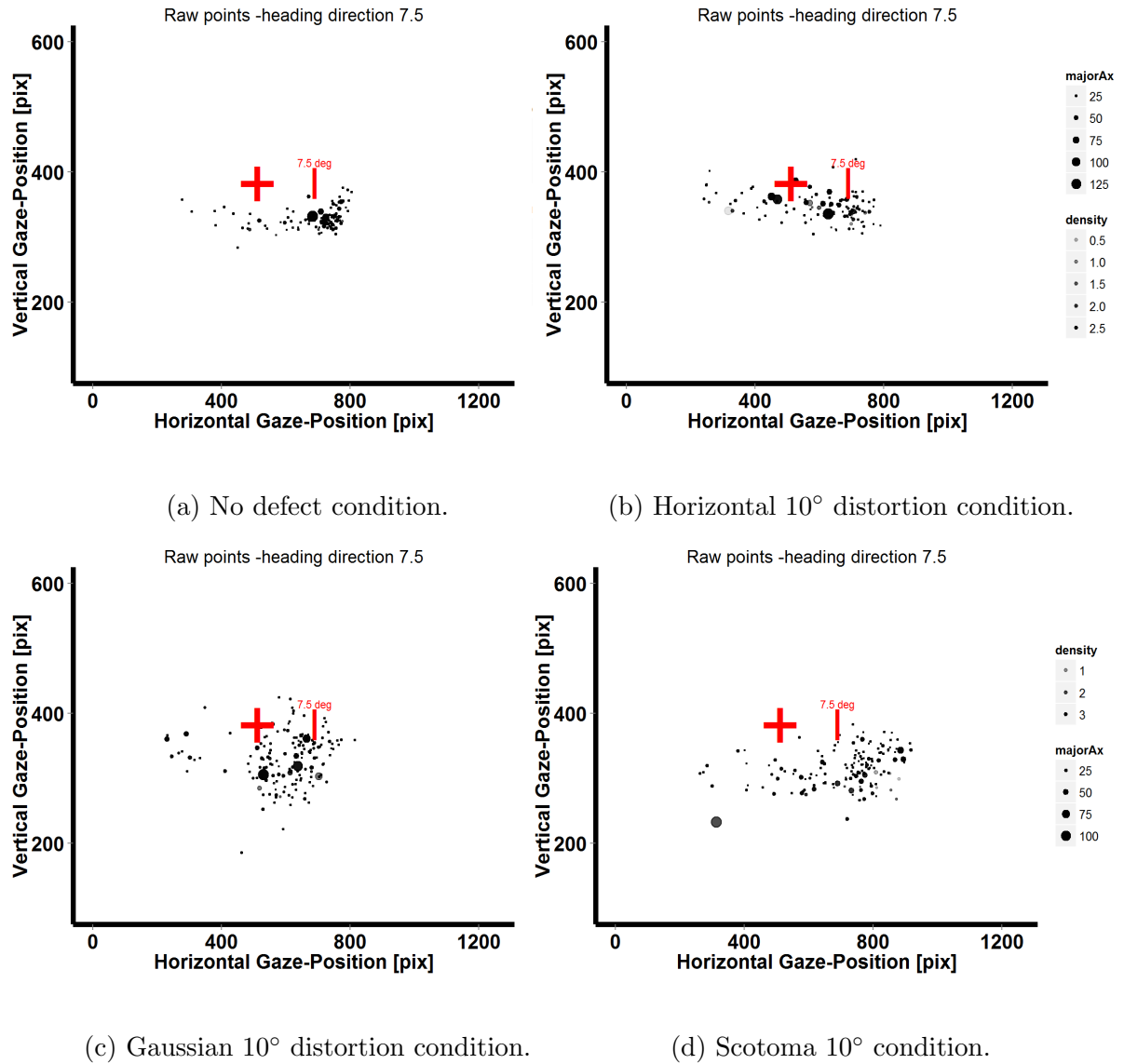
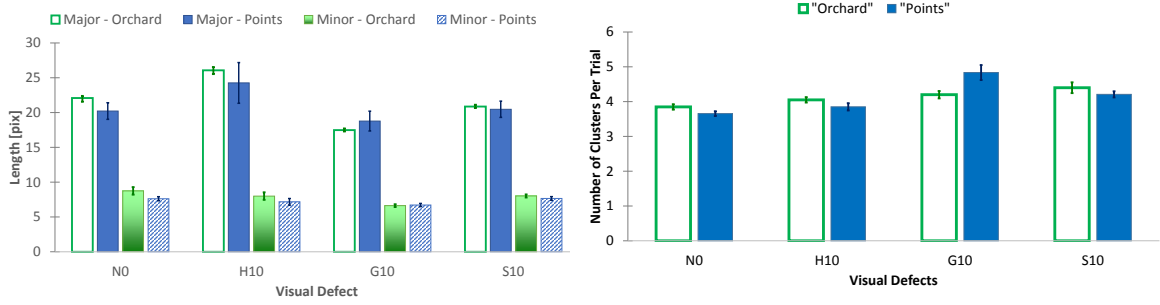
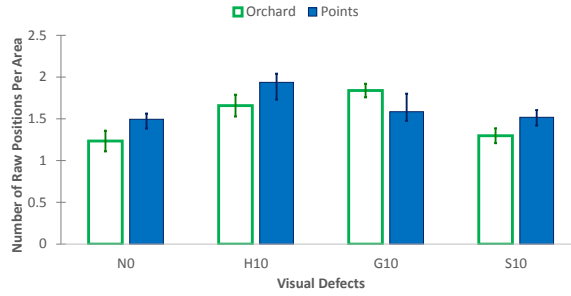


Figure 3.15: Gaze locations for participant 1 pooled across all trials for 7.5° heading eccentricity for four visual defect conditions. These examples were for the free viewing condition.

presented in Table 3.4). From the table one can see that simulated visual defects had a significant effect for all measured parameters (Major, Minor length of ROI



(a) Mean cluster sizes (major and minor axes) across visual defect conditions. (b) Mean number of clusters across visual defect conditions.



(c) Mean density across visual defect conditions.

Figure 3.16: Different cluster characteristics averaged across participants for all visual defect conditions during free viewing. The error bars represent the standard error of the mean for each condition.

and number of small clusters per each ROI(density)). From the figure 3.17a one can see that all horizontal distortion and scotoma conditions had a significantly larger major axis compared to no defect condition (Tukey post-hoc analysis with Bonferroni adjustment had $p < 0.001$ for N0 versus H10 and N0 versus S10). G10

on the other hand had a significantly smaller major axis relative to the no defect and scotoma conditions (Tukey post-hoc analysis with Bonferroni adjustment had $p < 0.001$ for G10 versus S10 and $p < 0.05$ for N0 versus G10). Tukey post-hoc analysis with Bonferroni adjustment showed that no defect condition had a significantly smaller minor axis relative to all defect conditions ($p < 0.001$). In terms of number of clusters per ROI area (Figure 3.17b), the no defect condition had significantly smaller number of gaze clusters per trial in the region of interest than H10 and S10 conditions ($p < 0.01$) and G10 had fewer gaze clusters relative to H10 ($p < 0.05$). This indicates that, overall across trials, people were more consistent with their gaze in the no defect and a Gaussian distortion conditions compared to the rest of the conditions. In addition, there was a significant interaction between simulated defect and the environment for length of a major axis of ROI (Table 3.4). It is possible that the difference between environments in the total number of ROI, as in the Orchard conditions, is because there were more interesting features that the observer would like to look at, such as branches and leaves versus uniform particles (figure 3.17c).

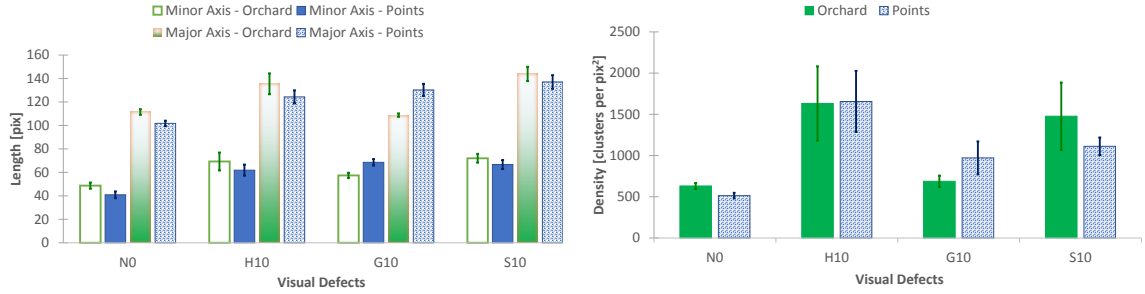
In addition, I examined regions of interest during the trials, and more specifically I looked at the percent of overall gaze clusters near target heading for each condition. I assumed that if people were looking in the direction of simulated heading, their gaze should fall roughly within radius of 2° of the simulated heading direction.

	<i>F</i>					
	df1	df2	NC	Density	Major Axis	Minor Axis
Visual Defect	3	2	0.98	5.37**	14.88***	12.007***
Environment	3	1	15.75***	0.003	0.18	0.55*
Visual Defect:Environment	3	2	0.09	1.42	4.24**	2.09

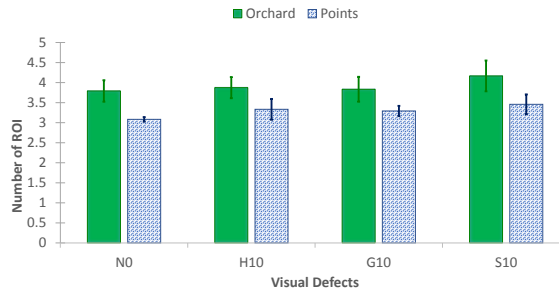
Table 3.4: Two-way repeated measures ANOVA for ROI across all defect conditions for free viewing condition. Significance of each contribution is shown by stars ($p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

Figure 3.18 presents the percentage of gaze clusters within 2° of the simulated heading for all simulated conditions. A repeated-measures ANOVA analysis showed that there was significant effect of visual defects on the number of gaze estimates within the region of interest ($F(3) = 29.28, p < 0.001$). Specifically, Tukey post-hoc analysis with Bonferroni adjustment showed that there was a significant difference between the N and H10 ($p < 0.01$) and between N and S10 ($p < 0.001$). There was also a significant difference between G10 and S10 ($p < 0.01$). Participants were more likely to fixate within the region of interest when the visual field was intact, presumably because it was most beneficial in these conditions.

Environment also had a significant influence on percentage of gaze within the region of interest ($F(1) = 7.84, p = 0.006$). In addition, there were interactions between the defect type and the environment ($F(3) = 3.10, p = 0.03$) as well as between defect type and simulated heading direction ($F(15) = 4.99, p < 0.003$).



(a) Mean cluster sizes (major and minor axes) across visual defect conditions. (b) Mean number of small gaze clusters per area of ROI across visual defect conditions.



(c) Mean number of ROI across visual defect conditions.

Figure 3.17: Different ROI characteristics averaged across participants for all visual defect conditions during free viewing. The error bars represent the standard error of the mean for each condition.

Figure 3.11 illustrates that during free viewing the participant fixated near the heading direction when there was no defect. I compared the slopes and intercepts for the regression between fixation and heading for the different defects. The mean values are presented in figure 3.19. The figure shows that regression slopes for

vertical gaze location were small (close to 0). In contrast, the regression slopes for the horizontal location were large for all the defect conditions, where the highest condition was for no defect condition and the lowest was for scotoma condition.

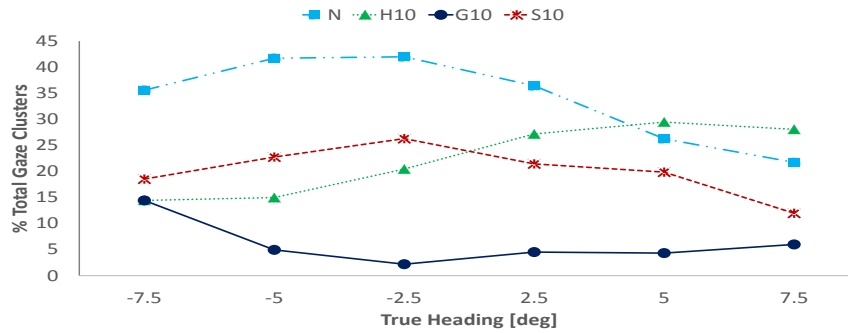
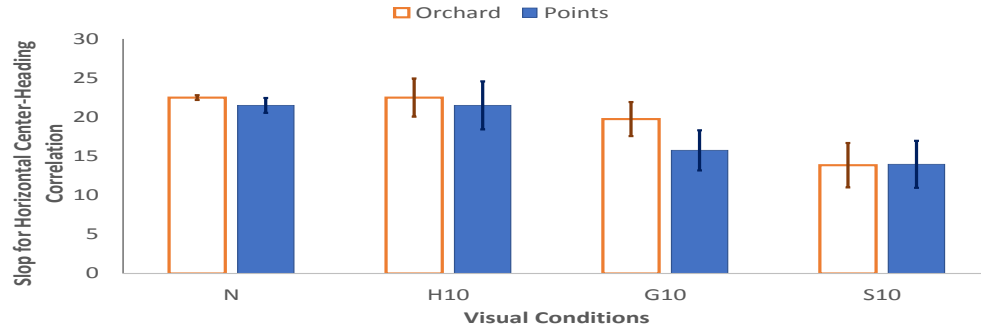


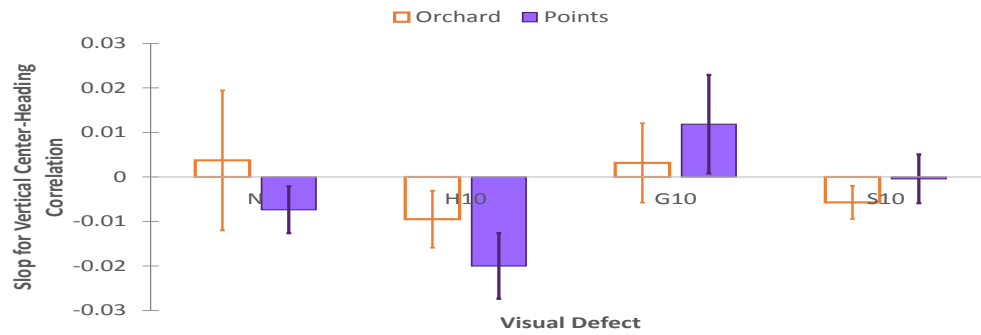
Figure 3.18: Percentage of gaze clusters that fall within a radius of 2° from simulated heading for four defect types across all heading directions.

3.5.3.4 Discussion

There were no significant effects of environment on stimuli in terms of performance data, however, there was a significant effect of environments on stimuli in terms of gaze data. This suggests that it is important to examine heading perception both in simplified schematic displays but also in natural displays as viewers might acquire visual information differently. In terms of sensitivity to heading directions, I observed that the frequency of large heading thresholds and large heading bias grew with eccentricity from the centre of the screen regardless of the visual techniques I asked participants to follow. Once again this supports the theory that the effect was



(a) Mean slope value for the relation between the horizontal centre of gaze clusters' location relative to the simulated heading direction across visual defect conditions.



(b) Mean slope value for the relation between the vertical centre of gaze clusters' location relative to the simulated heading direction across visual defect conditions.

Figure 3.19: Mean slope value for the regression of gaze clusters centre location as a function of heading direction across visual defect conditions.

due mainly to the characteristics of the optic flow and not due to retinal eccentricity.

I also showed that viewing mode had a significant impact on heading thresholds, particularly for large heading eccentricities. I was able to identify distinct pattern in gaze behaviour for each viewing condition. It is interesting to note that free-viewing produced the most precise heading judgements. By contrast, when observers tracked a specific feature on the ground plane their heading judgements

were significantly impaired at heading eccentricities of 5° and 7.5° . This was found to be the case for all of the simulated visual defect conditions, which suggests that when participants actively tracked the particle with their central vision, they failed to utilize their peripheral vision to help them to determine the correct heading direction.

Our psychophysical results also demonstrated that visual defects have an impact on heading judgements. Interestingly, this was true not only for a large scotoma but also for a visual perturbation. It is important to note that during debriefing many participants indicated that they were not aware of the distortions in the perturbation condition. This is reminiscent of many AMD patients, who are not aware of their metamorphopsia.

However, gaze data showed that people used very distinct patterns with each visual defect. For instance, people spent more time fixating during simulated defects, in particular during scotoma, relative to the no defect condition. This could indicate that it took them longer to acquire important information. On the other hand, the number of samples per cluster was higher for the distortion condition than in the scotoma condition. This is consistent with the literature that indicates people with simulated scotoma experience fixation instability [14]. This is also true for those with real central scotoma [187]. Also the nature of simulated defects was reflected in the percentage of gaze clusters that fell within a radius of 2° from sim-

ulated heading. Specifically, the no defect condition had the highest percentage of gaze clusters fall within this area of interest. This was not the case for Gaussian distortion, although the performance improved with simulated heading to the left. This suggests that the Gaussian distortion had a significant influence over where the fixation location. From all the observations above, it can be concluded that the gaze patterns indeed reflected the actual nature of the simulated defect. Consequently, it appears that navigation with visual perturbations can be dangerous as people have to use inaccurate information for navigation tasks such as steering.

3.6 Conclusions

In this chapter, I demonstrated the use of gaze-contingent simulation of visual defects to study locomotion. I conducted a series of experiments that looked at people's ability to correctly establish the heading direction of a simulated locomotion. I established that the ability to judge heading is significantly degraded with heading eccentricity, as well with simulated visual defects. Finally, I also established that very distinct gaze patterns were developed as a result of introducing simulated visual defects. I believe it maybe possible to use these observation in order to develop methods for early diagnosis of visual diseases such as AMD.

4 Gaze-Contingent Depth of Field Simulation

4.1 Introduction

Stereoscopic displays are no longer the exclusive prerogative of the cinema and research laboratories. People nowadays can afford to buy stereoscopic 3D (S3D) TVs and even stereoscopic monitors, tablets and smart phones. Soon the average user will have an option to choose between 2D and 3D displays for any device they might use. Hence, there is increased interest in developing S3D applications for a variety of displays and devices. It is also important to pay attention to the design of the 3D interface to avoid the issues with quality and perceptual human factors that famously contributed to ending the stereoscopic movie fad of the early- to mid-1950's [234].

Stereoscopic displays spark interest not only among users and developers but also the research community. Over the past few years, extensive research has evaluated different aspects associated with these displays. Yet, many questions remain open, including what types of content are best suited to stereoscopic displays, which

tasks benefit the most, how long will the user be able to effectively and comfortably interact with such applications and whether stereoscopic displays can be used effectively by everyone in the general population.

One key problem associated with stereoscopic displays is that stereopsis is only one among many of the cues that help people determine depth. Depth cue omissions can significantly impair depth perception and cause viewers to perceive the observed space flatter than it would appear in real life [222, 190]. So, which cues are missing in typical stereoscopic displays and is it possible to add those cues to improve depth perception?

In real life, the clarity of the retinal image of an object depends on its relation to an eye's fixation in the scene (the point of regard). In other words, the retinal image for a well-focused eye is the sharpest for objects at the focal distance, and is increasingly blurred as the depth of the object from the focal distance increases [69]. The cornea provides most of the optical power contributing to image formation in the eye but the cornea has a fixed focal length. To focus on the POR, the eye must adjust the shape of its intraocular lens to bring objects nearer than infinity into sharp focus; this process is known as accommodation. The ciliary muscles are responsible for adjusting the shape of the lens to accommodate the eye on an object of interest. Conversely, accommodation provides a physiological cue to the distance of an object: by monitoring the focal state associated with the fixated object the

observer could obtain an estimate of the distance of the object. Accommodation is a distance cue that usually is not accounted for in stereoscopic displays. While it is possible to provide this cue to an observer in virtual reality, this requires special volumetric displays. Such displays present targets at different optical distances, for instance by displaying graphics on multiple planes at a variety of optical distances [182, 184, 4].

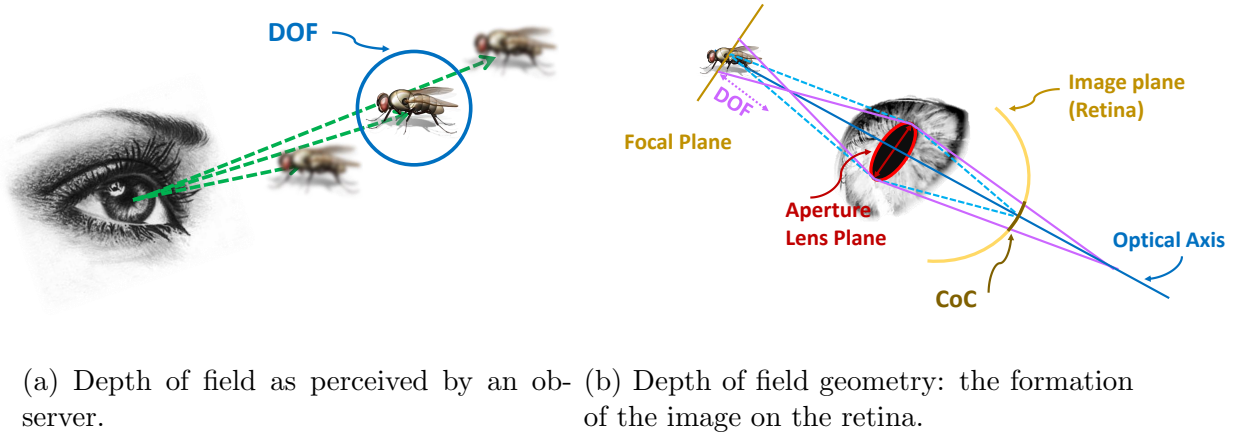


Figure 4.1: Depth of Field.

Although, accommodation provides information concerning the distance of the fixated object, it does not provide static information about depth between objects in the display (although change in accommodation across fixations could be informative). However, when the eye accommodates at a given distance, objects that are nearer or further will be subject to defocus blur and image blur is an informative depth cue [137, 72]. In contrast to the real world, most 2D and 3D graphical

applications do not render selective image blurring. Movies, games, and still photography do, but frequently for artistic reasons rather than realistic simulation and, in these cases, the blur is determined by the camera's focus not the eye's focus.

The range of distances where objects are perceived to be in focus for an imaging system such as the human eye is typically referred to as the *depth-of-field* (DOF) (figure 4.1). In 2D photography, the extent of the DOF is determined by the circle of confusion (CoC), which is a blur circle in the image plane (retina for the human eye). The size of CoC depends on size of the aperture (pupil) and the depth relative to the focal plane. As depth increases, the blur due to the CoC eventually becomes detectable (according to some criterion). Hence, the depth of field depends on both the CoC and the resolution of the sensor (visual acuity of the eye). The border of the CoC is not distinct in a real eye but the CoC is a useful model of DOF. The diameter of the CoC (b) can be approximated by different models. For example, Pentland [154] used a thin-lens model to describe b as follows:

$$b = A \frac{s_0}{d_0} \left| 1 - \frac{d_0}{d_1} \right|$$

Where A is the pupil diameter, s_0 is the distance from the lens to the retina, d_0 is the distance from the lens to the focal plane and d_1 is the distance from the lens to another object, whose image forms behind an image plane. Consequently, image blur due to lens defocus can be simulated by calculating blur circles for different objects. By adding DOF to 2D displays, one can contribute to depth perception

and improve depth qualitatively [120, 209]. In addition, Mauderer et al. [123] found that gaze-contingent DOF increased perceived realism in 3D images.

Accommodation and defocus blur are less important for the cinema as they are effective cues to distance only for relatively near targets, say less than two meters away. Presentation of such images in the cinema is rare since these would correspond to objects presented at extreme depths with respect to the screen. However, smaller stereoscopic displays are typically viewed at closer distances. In a such scenario, people often rely on an additional distance cue, which is *vergence*. Vergence is a physiological distance cue associated with the movements of the eyes. In vergence, the two eyes rotate in opposite directions. A principal function of vergence is to align the high-resolution fovea of both eyes on a target of interest to get a sharp binocular image of the object. As a result, observers need to increasingly cross, or converge, their eyes as the distance to an object of interest decreases. Typically, convergence and accommodation are tightly coupled [170]. However, this is not the case for stereoscopic displays. The problem arises from the fact that, when an observer views a stereoscopic 3D display, she needs to converge her eyes to fuse stimuli located off the screen, while accommodating her eyes at the screen distance. This is known as an *accommodation-vergence conflict* (figure 4.2). This conflict can lead to a range of negative side effects, such as discomfort, eye-strain, headache, and visual fatigue [113, 131, 215, 98, 76]. One solution is to try to

null the conflict by a quickly adjusting binocular disparities to keep objects of interest near the screen plane [16]. However, as with accommodative displays, such a solution is often limited by the number of discrete physical screens (leading to a limited number of real distances). Such displays also cause unnatural shifting of the rendered scene relative to the screen as points of interest change. Finally such accommodation displays have noticeable artefacts associated with fast disparity adjustments. Hence, a possible solution to alleviate the impact of the negative side effects is to provide an artificial simulation of defocus blur [26, 205]. This would not resolve the accommodative-vergence conflict but would provide the natural relationship between retinal image blur and binocular disparity.

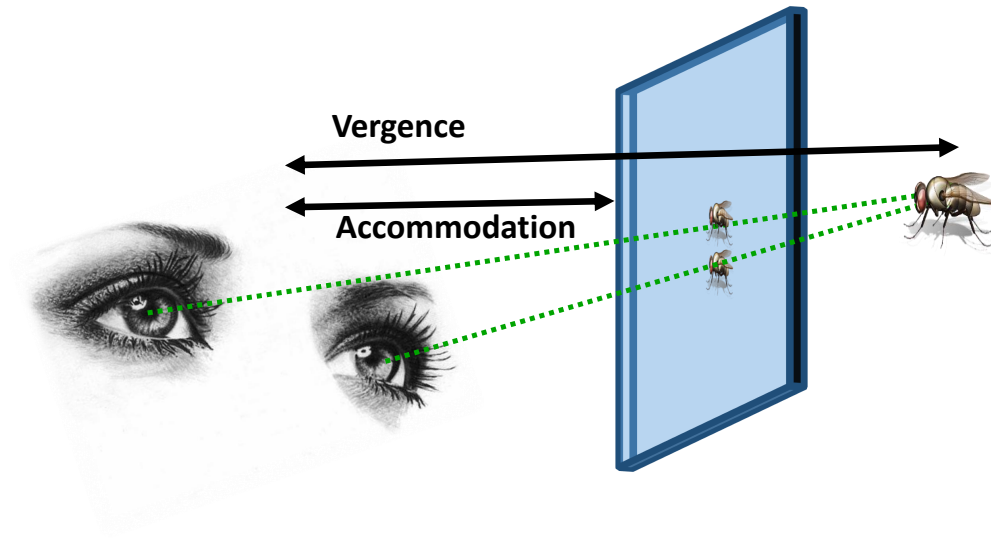


Figure 4.2: Accommodation-Vergence conflict.

Such an approach requires an invisible user interface that responds to the user's

action, in this case their gaze movements, in real-time without any explicit user intent. In order to provide the blur cues present in the real world, the interface needs to measure the POR and update the display in a naturalistic fashion. Ideally, such a simulation will provide a more natural and comfortable interface that users could tolerate for reasonably long periods of intensive use. A DOF simulation has to be congruent to the user's POR at every instant. The best way to achieve this by incorporating GCD implementation. Gaze-contingent DOF simulations estimate the POR in 3D space and then alter the visual scene accordingly. Such techniques could be very useful in games and applications that require interacting with 3D content. For example, Hillaire et al. [73, 74] used gaze-contingent DOF in the context of navigation and user experience. They observed that participants had a positive user experience with DOF added. Although they did not report any performance or psychophysical results in their first paper [73], they conducted a subjective evaluation in their second paper [74]. In the latter paper, they asked participants to rate their experience navigating in the virtual environment without DOF, with static DOF fixed to the centre of the screen and finally with gaze-contingent DOF. They found that with the gaze-continent DOF, participants had the best user experience in terms of: rendering realism, fun, depth perception and immersion. Mantiuk et al. [115] also built a a display system which incorporated gaze-contingent DOF. They then used their system to test different levels of simulated blur as well as

static versus contingent simulations in terms of their impact on subjective ratings of realism. They found that participants gave static DOF simulations very low scores, where gaze-contingent with medium blur scored the highest. Leroy et al. [104], on the other hand, used head-coupled DOF in the context of a pointing task. They examined the impact of blur on both pointing performance and visual fatigue. However, their system was head-contingent and not gaze-contingent. Therefore, it is important to evaluate the benefits of gaze-contingent DOF in the context of interactions with the 3D environment. In addition, the benefits of DOF on a user's performance during extended sessions, where the effects of cue conflict are expected to be manifest most significantly, are not established. Finally, Duchowski et al. [51] tested stereoscopic visual comfort with gaze-contingent DOF. Their results showed that subjective measures, such as visual fatigue, were reduced with DOF, yet participants expressed dislike of DOF on the preference measures. Although Duchowski et al. attributed this to noticeable temporal lag, it is possible that other factors, such as an individual's ability to accommodate and see clearly at certain distances, affected user's ratings.

This chapter is based on my preliminary work [209], where I evaluated the benefits of DOF as a depth cue and determined the crucial perceptual factors that might impact the strength of DOF as a cue. Specifically, I investigated whether simulation of DOF could improve a user's perceptual range with emphasis on depth

perception, general comfort, and aesthetic preferences. In my original study, I tested young adults under the age of 30 years. In a follow-up experiment, I extended this to older adults to see whether there are significant differences between older and younger adults.

Furthermore, there are at least two different aspects or stages of visual discomfort. The first stage, often referred to as visual stress, is usually associated with an unpleasant sensation that results almost immediately upon exposure to a stimulus. Consequently, researchers that are interested in how a display or stimulus produces visual stress usually show images to the participants for several seconds (0.5 to 15 s is typical) and have them judge or compare the images in terms of visual comfort [90, 21, 141, 209]. The second stage is known as visual fatigue and builds up over time, often in response to prolonged visual stress. Visual fatigue often involves visual discomfort, both when the stimulus is present and when it is removed. To study visual fatigue, extended exposures of 30 min to several hours are required [76, 131, 200]. Many papers have addressed visual discomfort in visual displays using a variety of different techniques and measures. Yet, the short and long term effects of visual stress on visual discomfort and visual fatigue have not always been distinguished. I believe that it is critical to make such a distinction to fully understand the relationship between stereopsis and DOF on user comfort. Therefore, in this chapter, I discuss issues associated with short-term visual com-

fort, and the effects of stereopsis and dynamic DOF cues on visual fatigue with extended viewing.

4.2 System Concept

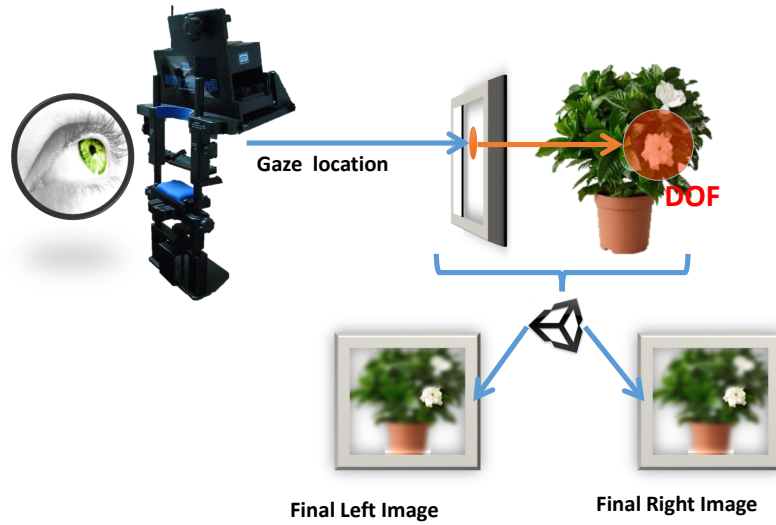


Figure 4.3: System architecture. The eye tracker was used to determine POR in the 3D scene based on the location of gaze intersection with the display. This point allowed the system to render DOF about an appropriate focal distance for the right and left eyes.

Figure 4.3 shows the system concept for the gaze-contingent DOF system that I developed and used in the present studies. To simulate DOF, I incorporated gaze-contingent image processing and used Unity3D as my 3D interactive graphics platform. As a result, I was able to use different virtual environments with custom shading models and scripts to simulate different static and kinetic scenes. I could

also implement different modalities for input acquisition including but not limited to 3D mouse handling. In addition, I developed separate scripts that supported data acquisition and handling from an eye-tracker. These scripts also supported a calibration interface. Furthermore, my virtual scenes always included several camera objects that used shaders and scripts responsible for final image processing operations. This permitted me to implement a range of different visual simulations such as visual distortions and defocus blur.

For the purpose of this chapter, I will describe the algorithm for the gaze-contingent DOF simulation as follows:

1. In real-time, I collected the participant's gaze location in screen coordinates.

This was used for determining the focal distance in 3D space. Estimation of the focal distance was achieved by determining the POR which is the 3D point, where the participant's gaze intersects with a virtual object [210]. I utilized Unity3D ray projection methods to accomplish this task. This helped to establish the focal-depth and the region free of blur, in other words, the DOF around that point.

2. I simulated DOF and defocus blur based on an average pupil size of 3.5 mm associated with a display with luminance of $220 \text{ cd} \cdot \text{m}^{-2}$ [227]. This aperture size determined the extent of blur at different distances from the focal plane.

To impose blur, I used a post-processing effect that permitted me to alter each pixel value based on its depth relative to the POR. I implemented this post processing with shader scripts that work directly with the graphical processing unit (GPU) to maintain low latency and not to impact frame rate.

3. I extended my system to support stereoscopic rendering. I achieved this by determining DOF for each eye independently and by processing the rendered scene separately. It is important to emphasize that in conditions without simulated DOF (monocular cues and stereoscopic cues), I simulated an infinitely large DOF. In other words, the images were sharp at all distances in the 3D space and the left and right images contained binocular disparity. By monocular, I refer to the depth cues simulated; viewing was always binocular. Thus, to implement the monocular condition, I simulated a non-stereoscopic display by presenting identical (zero-disparity) images to both eyes; thus, all processing and rendering steps were equivalent in the monocular-cue and stereoscopic-cue cases. I should also point out that my system works well with any virtual model as far as it can be correctly scaled to correspond to a physical distance from the model to the user.

4.2.1 Apparatus

The visual scenes were generated on a desktop computer with the following specifications: AMD FirePro W9000 FireGL, Windows 7 Enterprise, Intel® Core™ i7-3770k CPU, 3.50 GHz, 3.50 GB RAM. The stimulus was presented on a Samsung 950 Series 3D monitor (27", with a pixel resolution of 1920 H x 1080 V and a refresh rate of 120Hz or 60Hz per eye). On each frame, a stereoscopic pair of images was presented sequentially to each eye via active shutter glasses. I carefully verified that stereoscopic shuttering of the active glasses did not interfere with the eye movement estimates.

During the experiment, the screen was viewed binocularly at a distance of 0.35 m, and the stimulus subtended a horizontal visual angle of 80.6° (Figure 4.4). A real-time gaze-contingent system was built by incorporating an EyeLink 1000 (SR Research Ltd) eye-tracker. I used the tracker's tower mount setup, for the purpose of achieving a reliable tracking at the close viewing distance. A chinrest was used to stabilize the head and to maintain the viewing distance, with the midpoint between the right and left eyes, centred in relation to the screen. The experiment was conducted in a darkened room with only the stereoscopic display illuminated.



Figure 4.4: Experimental setup. The participant sat in front of the 3D display (partially visible beyond the tower mount) with her head fixed by the tower mount provided with the EyeLink 1000.

4.2.2 Calibration

The calibration involved sequentially fixating nine points displayed on the screen in a pseudo-random order. Each calibration procedure was followed by a validation step, where another set of nine points randomly appeared on the screen in sequence. The distances between expected gaze location and the estimated location acquired during this step, indicated whether the calibration was successful or not. In the case that the tracking error was above 0.5° , the system was recalibrated after necessary adjustments to the tracker were made.

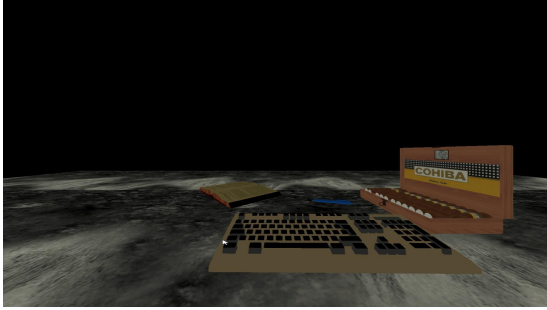
4.3 Depth Impressions, Visual Comfort and Image Quality

In the first series of experiments, I explored the impact of different depth cues, alone or combined, on short-term visual comfort. I was also interested in determining whether short-term visual comfort was affected by age. Numerous studies have shown that the ability to accommodate at different distances changes with age [92, 91, 176]. Accommodation is highly plastic in youth, but it dramatically declines by middle age [92, 91, 176]. Several hypotheses were outlined, which I aimed to support or disprove. First, it was hypothesized that simulated DOF would be significantly more beneficial for people who have reduced accommodative ability than people with a large range of accommodation. I also was interested to determine if a combination of depth cues in VR settings would provide more compelling depth perception than scenes rendered with fewer depth cues. On the other hand, while I expected that DOF would improve visual comfort, I hypothesised that it might degrade perceived image quality. Finally, I thought that scene content would influence participant preference for all three parameters. Specifically, I hypothesized that people would enjoy the presence of DOF in scenes with fine features at different distances.

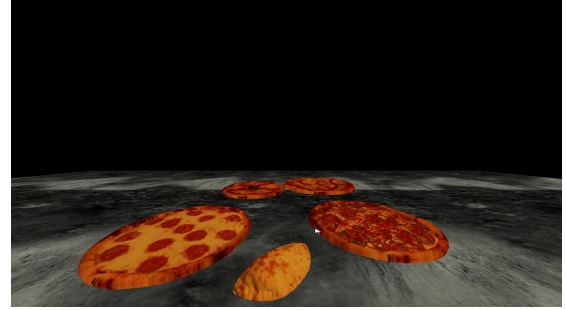
4.3.1 Stimuli

The virtual scenes consisted of several 3D models of everyday objects, such as books. In all cases, the objects were placed on a horizontal plane so that the participant could interpret the objects as lying on a table-top. The scenes were designed to resemble those encountered in everyday interaction with objects and common surfaces within a realistic living space. Figure 4.5 shows the three scenes presented in the experiment. The scene with the flowers was chosen because it had many vertical elements; conversely the scene with stationery items had many details distributed across the horizontal plane, such as the keyboard and cigars. Fine detail was also present on the pages of the book with fine print for the second scene. The scene that displayed the pizzas was chosen due to its high level of detail and texture; in this way, the effect of blur on textures would be noticeable. Each of these scenes was presented with all combinations of provided depth cues (conditions).

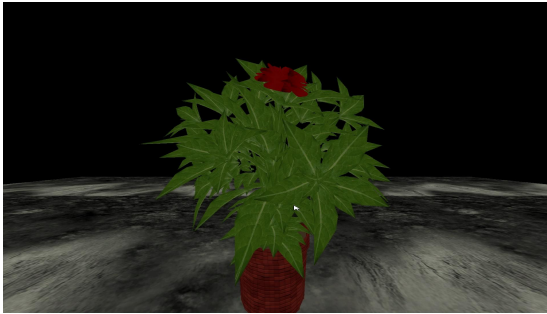
Each trial consisted of two images of a given scene, presented sequentially each with different depth cues: monocular cues (Monoc), monocular cues combined with DOF (DOF), monocular and stereoscopic cues without DOF (Stereo), and monocular together with stereoscopic cues combined with DOF cues (DOF + Stereo) (Figure 4.6).



(a) Stationery scene.



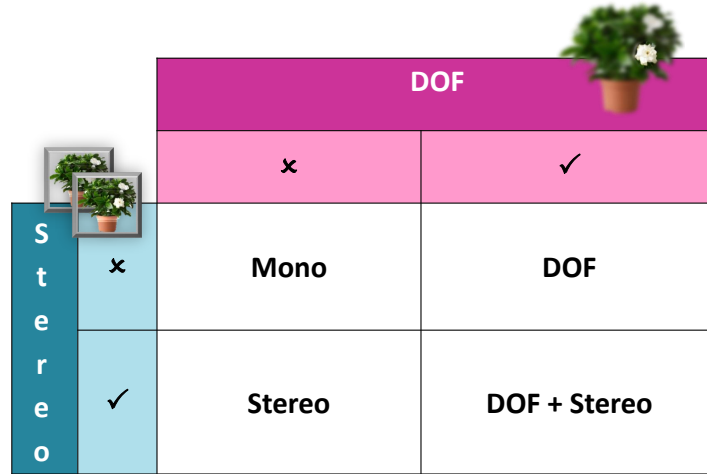
(b) Pizza scene.



(c) Flower scene.

Figure 4.5: The three scene variations.

By permutation there were 12 (4×3) pairings of the different cue conditions, three different scenes and four repetitions for each, providing 144 trials in total presented in a different random order to each participant. Each scene was presented to the participant for eight seconds (4 s per condition). The trials were divided into four blocks and the calibration procedure was performed prior to each block.



		DOF	
		x	✓
Stereo	x	Mono	DOF
	✓	Stereo	DOF + Stereo

23

Figure 4.6: Depth cues for each condition.

4.3.2 Task

Participants answered two questions following each trial. It was impractical to have participants come for two consecutive sessions and hence I conducted a between subjects experiment. As a result, half of the participants compared trials based on the following two questions:

1. Which scene had the most compelling depth? (Depth impression)
2. Which scene had the better image quality? (Image quality)

The second half of participants were asked the following two questions:

1. Which scene was most comfortable to view? (Viewing comfort)
2. Which scene had the better image quality? (Image quality)

Participants were required to choose either the first or the second interval, even if there was little apparent difference (a forced-choice procedure). Participants were also instructed to inspect the scenes thoroughly so that fair comparisons could be made between the image pairs. To ensure that the participants were paying attention to the task and actively visually exploring the images, the experimenter monitored participants' eye movements on-line during the trial, to ensure participants were indeed actively scanning the images and that the eye-tracker was properly recording at all times.

4.3.3 Participants

Participants were split into two age groups - the younger adults (under age 30) and older adults (over age 50). Four groups were randomly selected from the recruited participants: a) younger adults comparing the scenes based on depth impression and image quality (D30-); b) younger adults comparing the scenes based on viewing comfort and image quality (C30-); c) older adults comparing the scenes based on depth impression and image quality (D50+); d) older adults comparing the scenes based on viewing comfort and image quality (C50+);

Adults (30-) The D30- group consisted of nine university students (6 females and 3 males, ranging in age from 22 to 29, average age 24) and the C30- group

consisted of nine university students (6 females and 3 male, ranging in age from 19 to 24, average age 21).

Older Adults (50+) The D50+ group consisted of seven participants (3 females and 4 males, ranging in age from 50 to 70, average age 56.57) and the second group C50+ consisted of seven participants (5 females and 2 male, ranging in age from 56 to 73, average age 60.57).

All participants could clearly see at the viewing distance without the use of eyeglasses. Prior to the experiment, participants were screened for stereopsis, using the RANDOT stereo-test [Chicago, USA]. All participants were able to see depth from disparity 50 seconds of arc or better at distance of 40.6 cm. Written informed consent was obtained from all participants in accordance with a protocol approved by the York University Ethics Board.

4.3.4 Results

4.3.4.1 Depth Perception

For both the 30- and 50+ age groups, a Cochran Q test revealed that there were significant differences in depth rendering preference between conditions ($\chi^2(5) = 35.31, p < .001$ and $\chi^2(5) = 260.4779, p < .001$, respectively). The mean percentage

of trials that one condition in a pair was preferred over the other for both age groups is shown in figure 4.7.

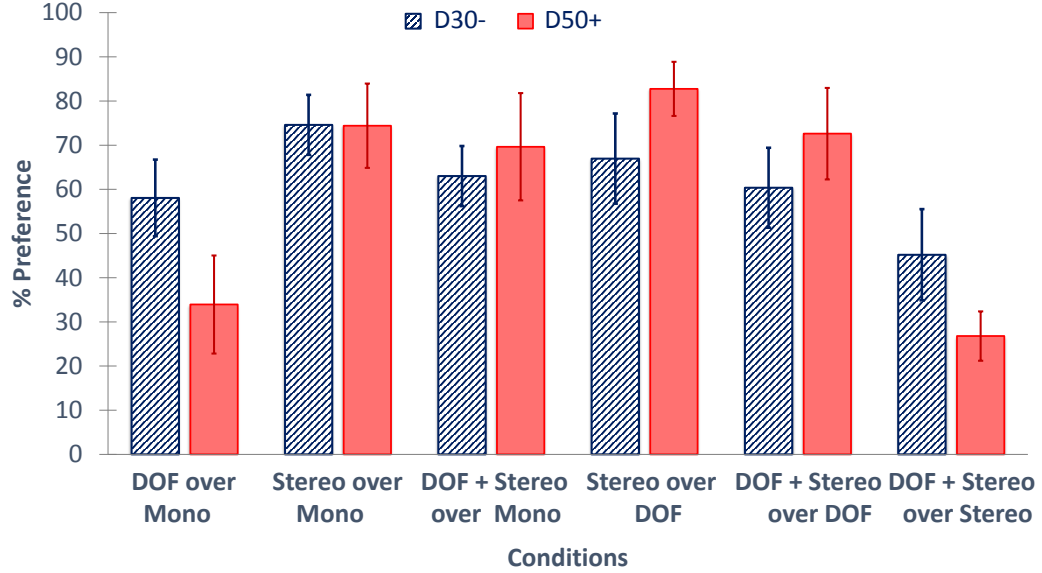


Figure 4.7: The mean depth preference averaged across observers for D30- and D50+ groups. The error bars represent the standard error of the mean for each condition.

In order to better understand the relationship between comparisons, a paired comparison analysis was conducted. A Turner and Firth model [197, 198] was used, as it fits generalised non-linear models (gnm) using an over-parameterised representation. Table 4.1 shows the estimated weights from the fitted model. These estimated weights show the contribution of the cues to preference relative to the base case (in separate rows); a positive estimated weight implies that the case is more likely to be preferred in comparisons. It is interesting to note the preference similarities and differences between the two age groups. Overall, in

both age groups, stereoscopic displays gave stronger depth impressions than non-stereoscopic displays (Mono) and Stereo estimated weights were significantly more positive than the estimated weights for the other cases. The main difference between the two groups was in the effect of the DOF cue, which was positive for the 30- group, indicating a relative preference in comparison to the Mono base case, but was negative for the 50+ group. This can be seen in figure 4.7; for example, in comparisons of the DOF and Mono conditions, the 30- group choose DOF over Mono ($M = 58.05\%$, $z = 2.65$, $p = 0.008$), while the 50+ group preferred Mono ($M = 33.92\%$, $z = 4.24$, $p < 0.001$). While, the significantly positive estimated weights for DOF + Stereo indicated that the combined depth cues were preferred relative to the Mono base case, the estimated weights were significantly smaller than for the Stereo case. This suggests that adding the DOF cue to a stereo display reduced rather than enhanced depth. Consistent with this finding from the model, both groups preferred depth from DOF + Stereo less than from Stereo cues (30-: $M = 45.19\%$, $z = 3.50$, $p = 0.0005$; 50+: $M = 36.78\%$, $z = 5.92$, $p < 0.001$).

Figure 4.8 shows the depth perception preferences for different scenes and both age groups. The younger participants showed that they preferred the presence of DOF as a depth cue over Mono particularly in the flower scene ($M = 69.25\%$). Conversely, the older group preferred the Mono over DOF condition for all scenes, although this preference was slightly smaller in the flower scene. For both age

	D30-				D50+			
	Mono	DOF	Stereo	DOF + Stereo	Mono	DOF	Stereo	DOF + Stereo
Mono	0	0.12**	0.49***	0.33***	0	-0.25***	0.62***	0.26***
DOF		0	0.37***	0.21***		0	0.88***	0.52***
Stereo			0	-0.16***			0	-0.36***

Table 4.1: The chart shows estimated weights of depth cues in determining depth preferences for both 30- and 50+ groups. Each row shows the contribution of each cue (columns) to preference relative to the base case (1st column); a positive estimated weight implies that the case is more likely to be preferred. Significance of each contribution is shown by stars ($p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

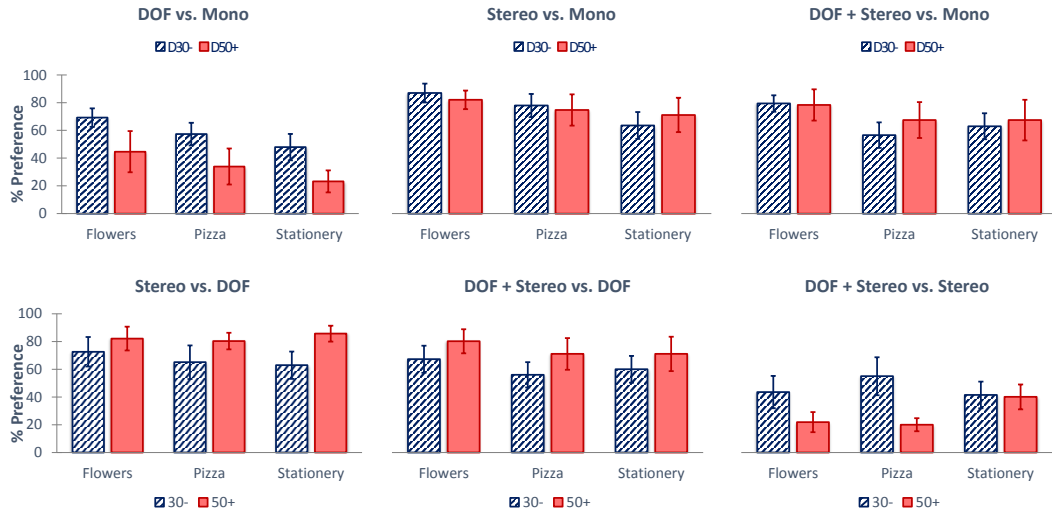


Figure 4.8: Mean depth preference for the three scenes for both D30- and D50+ groups. The error bars represent the standard error of the mean for each condition.

groups the preference for the Stereo condition over Mono was most pronounced for the stationery scene. The preference of the 50+ group for Stereo over DOF was near 80% and similar for all three scenes, but for the 30- group, this preference was stronger for the flower scene than for the pizza and stationery scenes. Finally, when participants had to compare DOF + Stereo over Stereo, while 30- had the same

preference pattern for both flower and stationery scene, 50+ group choose DOF + Stereo more often in the stationery scene than in the rest of the scenes.

4.3.4.2 Viewing Comfort

For both the 30- and 50+ age groups, a Cochran Q test revealed that there were significant differences in comfort preferences ($\chi^2(5) = 555.162, p < .001$ and $\chi^2(5) = 99.321, p < .001$, respectively). The means and standard error for the comfort comparisons for both age groups are shown in Figure 4.9, and the estimated weights for both age groups are presented in table 4.2.

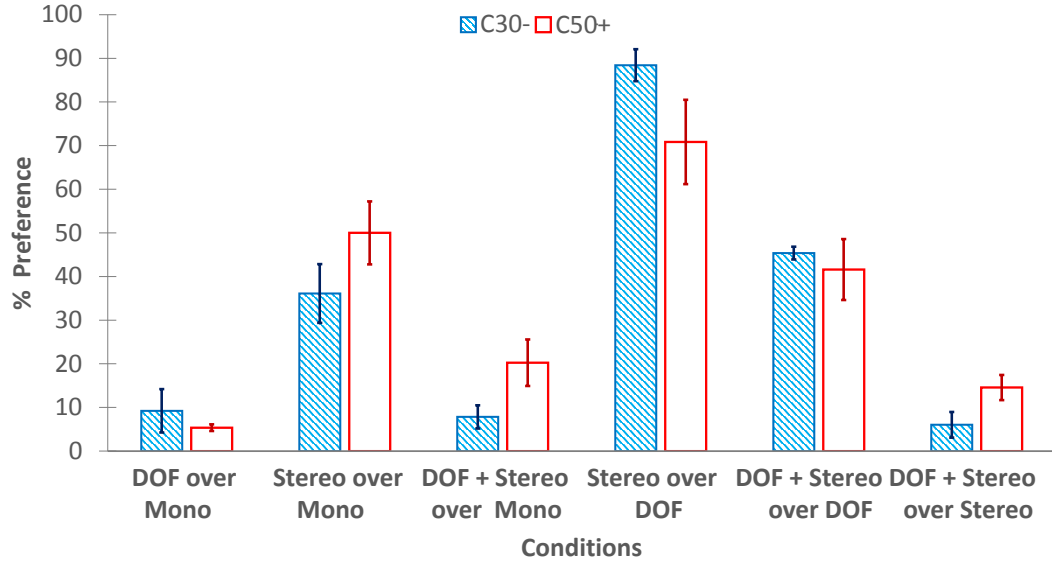


Figure 4.9: The mean viewing comfort preference averaged across observers for C30- and C50+ groups. The error bars represent the standard error of the mean for each condition.

From the table one can observe that for both groups, the pattern of viewing

comfort preference was opposite to depth preferences and much stronger. In other words, conditions with DOF were typically considered less comfortable, while the Mono condition was considered to be the most comfortable condition. This was true even when combined with stereoscopic rendering and stereoscopic rendering was judged more comfortable than DOF + Stereo. It is likely that the DOF simulation determined these ratings since weightings of DOF and DOF + Stereo were similar and much larger than Stereo. Adding the Stereo cue to DOF had only a minor negative impact on comfort (DOF + Stereo was chosen slightly less often than DOF in direct comparisons $M = 45.37\%$, $z = 2.07$, $p < 0.04$ for 30- group and $M = 41.60\%$, $z = 1.85$, $p = 0.06$ for the 50+ group) but adding DOF to a Stereo display had a large impact (DOF + Stereo was chosen over Stereo only $M = 6.02\%$, $z = 11.69$, $p < 0.001$ for 30- group and $M = 14.56\%$, $z = 15.06$, $p < 0.001$ for the 50+ group). Furthermore, it is interesting to note that the bias against Stereo was smaller for the 50+ group (-0.12) than for the 30- group (-0.22). In particular, when Stereo was compared with Mono, the younger group felt more comfortable with the Mono condition ($M = 36.11\%$, $z = 1.88$, $p < 0.06$) while the 50+ group had no preference ($M = 50\%$, $z = 3.61$, $p < 0.0003$). However, the 30- group did have a stronger bias for stereo in Stereo versus DOF and DOF + Stereo versus Stereo comparisons.

Viewing comfort, for different scenes and age groups, is presented in Figure

	D30-				D50+			
	Mono	DOF	Stereo	DOF + Stereo	Mono	DOF	Stereo	DOF + Stereo
Mono	0	-1.24***	-0.22***	-1.36***	0	-0.79***	-0.12.	-0.90***
DOF		0	1.02***	-0.12*		0	0.67***	-0.11.
Stereo			0	-1.14***			0	-0.78***

Table 4.2: The chart shows estimated weights of various depth cues in determining comfort preferences for both 30- and 50+ groups. Each row shows the contribution of each cue (columns) to preference relative to the base case (1st column); a positive estimated weight implies that the case is more likely to be preferred. Significance of each contribution is shown by stars ($p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

4.10. Generally, the comfort preferences were similar across all scenes with a few exceptions. Data for the 50+ group, suggests that the older group found it more comfortable to view the flower scene with the Stereo condition over the Mono condition ($M=66.07\%$), but had no preference between these two conditions for other scenes ($M = 39.25\%$ for pizza and $M = 45.15\%$ for stationery). On the other hand, the young adults preferred Stereo the least with the stationery scene ($M = 26.38\%$), while for the other scenes they proffered stereo about 40% of the time ($M = 43.98\%$ for flowers and $M = 38.49\%$ for pizza). Another interesting pattern was observed for the DOF + Stereo versus DOF conditions, where the 50+ preferred DOF + Stereo for all scenes about 45% of the time ($M = 41.67\%$ for flowers and, $M = 50\%$ for Pizza and $M = 44.44\%$ for stationery), while the 30- also preferred DOF + Stereo about the same percentage as the other group for the flower ($M=55.36\%$) and pizza ($M=44.25\%$) scene, but preferred DOF + Stereo only 25.03% of the time for the stationery scene.

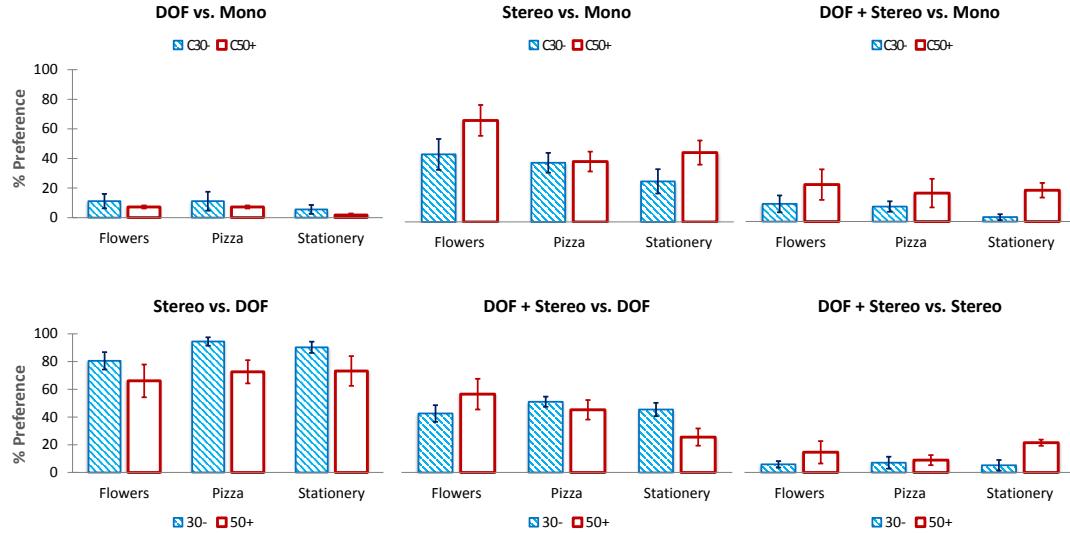


Figure 4.10: Mean viewing comfort preference for the three scenes for both C30- and C50+ groups. The error bars represent the standard error of the mean for each condition.

4.3.4.3 Image Quality

Image quality preference data are presented in Figure 4.11. A Cochran Q test revealed that there was a significant difference in image quality between conditions for both the 30- group ($\chi^2(5) = 282.22, p < .001$ for C30- group; $\chi^2(5) = 182.69, p < .001$ for the D30- group) and for the 50+ group ($\chi^2(5) = 235.78, p < .001$ for the D50+ group; $\chi^2(5) = 86.08, p < .001$ for the C50+ group). Overall, I observed that the pattern of image quality ratings closely resembled the pattern of comfort ratings. The Turner and Firth Model gnm estimates for both age groups are presented in Table 4.3. From this data, one can see that there were differences in preferences

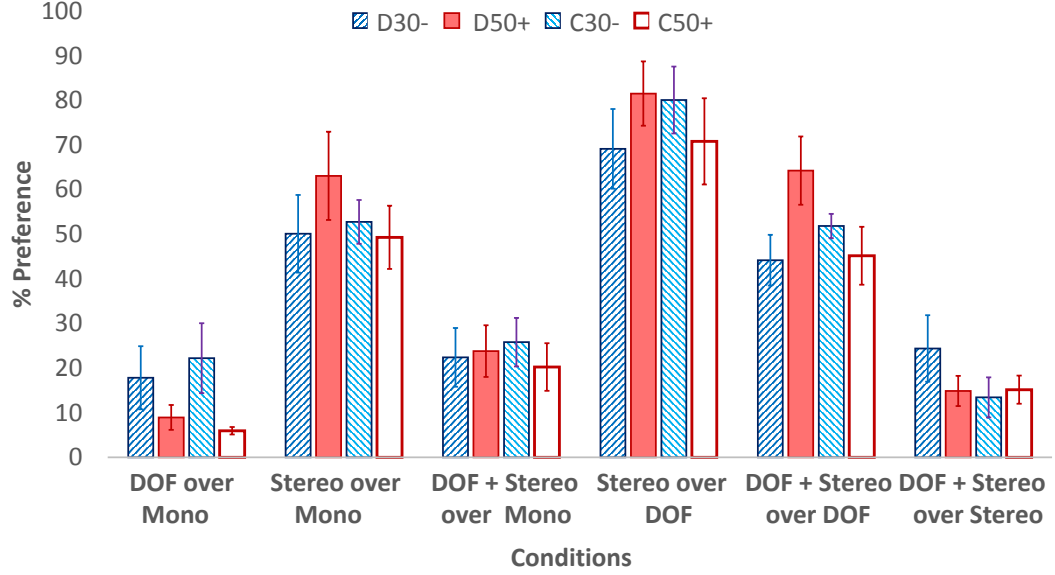


Figure 4.11: The mean image quality preference averaged across observers for 30- (D30-, C30-) and 50+ (D50+, C50+) groups. The error bars represent the standard error of the mean for each condition.

for both age groups, when image quality question was coupled with depth question as opposed to when it was coupled with the viewing comfort question. The data showed interesting trends across different comparisons. For example, when participants were asked to compare Stereo against Mono, only the D50+ and C30- groups preferred Stereo (0.17 and 0.16, accordingly), while C50+ had an opposite preference (-0.13) and D30- had no significant preference (-0.05) at all. A similar pattern was observed with the DOF + Stereo versus DOF condition, where only C50+ preferred DOF + Stereo over DOF (0.26) and the rest of the groups did not have significant preference for one condition over the other.

The image quality preferences for the three scenes is presented in figure 4.12.

Group	Mono	DOF	Stereo	DOF + Stereo
	<i>Mono</i>			
D30-	0	-0.61***	-0.05	-0.64***
D50+	0	-0.87***	0.17**	-0.62***
C30-	0	-0.67***	0.16**	-0.62***
C50+	0	-0.81***	-0.13*	-0.87***
	<i>DOF</i>			
D30-	0.61***	0	0.56***	-0.03
D50+	0.87***	0	1.04***	0.26***
C30-	0.67***	0	0.83***	0.05
C50+	0.81***	0	0.68***	-0.06
	<i>Stereo</i>			
D30-	0.05	-0.56***	0	-0.59***
D50+	-0.17**	-1.041***	0	-0.78***
C30-	-0.16**	-0.83***	0	-0.78***
C50+	0.13*	-0.68***	0	-0.74***
	<i>DOF + Stereo</i>			
D30-	0.64***	0.03	0.59***	0
D50+	0.62***	-0.26***	0.78***	0
C30-	0.62***	-0.05	0.78***	0
C50+	0.87***	0.06	0.74***	0

Table 4.3: The chart shows estimated weights of image quality in determining preferences for both 30- and 50+ groups. Each row shows the contribution of each cue (columns) to preference relative to the base case (as indicated in sub-titles); a positive estimated weight implies that the case is more likely to be preferred. Significance of each contribution is shown by stars ($p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

From the figure one can see that participants preferred Stereo over Mono condition during the viewing of a flower scene, but they disliked Stereo condition for the other two scenes. A similar pattern was observed for the DOF + Stereo and Stereo comparison, where, once again, during the flower scene participants preferred DOF + Stereo, but disliked it when had to observe the stationery scene. It is interesting to note that although participants did not like DOF + Stereo over Mono for all scenes,

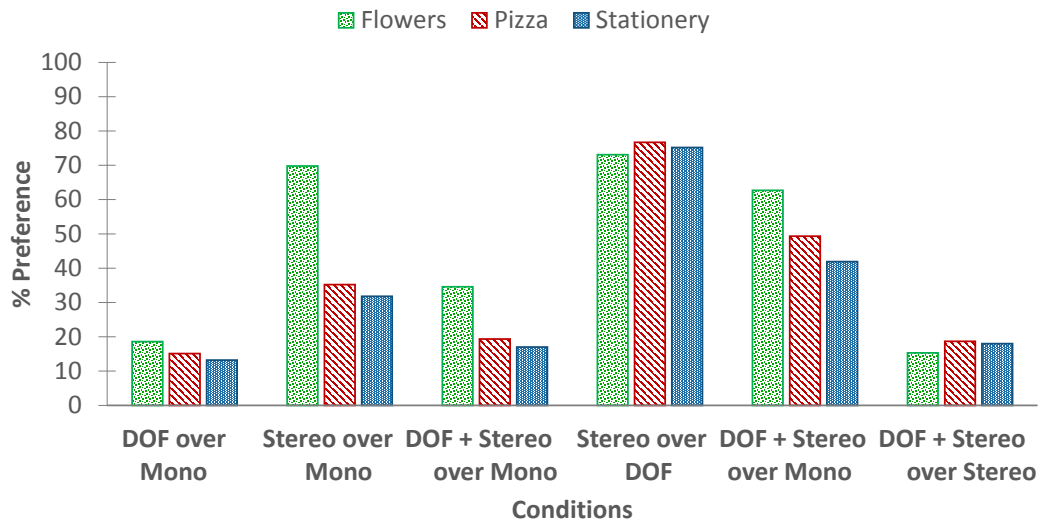


Figure 4.12: Mean image quality preference for the three scenes collapsed across groups.

they were less critical in the flower scene. Thus, in all cases, the positive influence of stereo cues was larger in the flower scene. For the rest of the comparisons the results were very similar across the scenes.

4.3.5 Discussion

In terms of depth preferences, I observed that in some cases DOF contributes to more compelling depth perception, specifically, when added to non-stereoscopic displays. The relatively weak influence of DOF on depth can be possibly explained by the fact that focal blur, in general, is not a very strong depth cue. One reason is that static focal blur itself does not provide the sign or direction of depth relative to

the focal point. Hence, participants need additional depth cues to fill in the missing information. Many of these cues, such as chromatic aberration and accommodative microfluctuations [137], are not provided by gaze-contingent DOF. In many cases, this sign ambiguity can be resolved by other depth cues such as stereopsis [67, 154, 117, 137]. These findings are also consistent with Duchowski et al. [51], who also observed that participants do not like gaze-contingent blurring effects. Their findings highlight the need to improve spatio-temporal fidelity, which includes, but is not limited to improving the accuracy of eye-tracking and reducing temporal lag.

In terms of short-term viewing comfort, I found that both viewing comfort and image quality were reduced, when DOF was added as a depth cue for both age groups. This can be explained by the fact that participants might notice the blurriness that results from DOF simulation. Furthermore, I observed very similar preference patterns between image quality and viewing comfort, but patterns were different for image quality and depth. This suggests that image quality, in particular, image blur is an important driver of comfort judgements in my displays. This is supported by medical studies of subjective *depth-of-focus*² [133, 7]. These studies suggest that participants’s tolerance of blur due to defocus can change with age independently from their actual ability to accommodate at different depths

²Atchison and Woods [7] defined subjective depth-of-focus as “the range of focusing errors for which the image of the target appears to have the same clarity, contrast and form as the optimal in-focus image”. That is not to be confused with an objective DOF that measures the actual change in blur with depth.

and from their objectively measured DOF. The subjective impression of increased DOF with presbyopia is a potentially important factor in the applicability of these techniques to older participants.

The age of the participants also had an impact on participants' preference for depth cues. As people age, they slowly lose their ability to accommodate to near objects [155, 92, 91, 176, 71] and possibly slowly stop using DOF as a depth cue at all. Additionally, the older-adults group was more tolerant to the presence of DOF in terms of viewing comfort. This is possibly due to the fact that, over the years, they have become less sensitive to the presence of blur than the younger population [133, 213]. For example, in their study Mordi and Ciuffreda have shown that subjective tolerance to DOF substantially increases with age, implying that image quality degradation that would be tolerated by a typical older observer would bother the typical younger observer [133]. A relative insensitivity to blur (increased subjective DOF) may explain why DOF was a less effective depth cue in the older observers.

When designing a gaze-contingent DOF, one has to also understand the scene and what features will attract a participant's attention. For example, people often direct their gaze to the edge of an object, where a tracking error or depth estimation error can lead to incorrect location of simulated DOF [202]. Binocular eye-tracking can provide a convergence angle signal to fixation depth that is independent of

the actual geometry and screen parameters of the virtual scene. For example, Duchowski et al. [50, 51] built such a system. However, vergence estimation for binocular eye-tracking is very sensitive to noise. Furthermore, observers often have fixation disparity, and this might be more pronounced in virtual environments. For instance, Duchowski et al. [50] compared depth estimation in a virtual environment with a physical environment. They found that vergence error varied with target distances. Moreover, this error was larger when participants looked behind or in front of the screen. Thus, depth estimation from vergence can lead to inaccuracies in depth judgements. In this study, the most accurate setup possible with the existing eye-tracker was utilized, which was a monocular set-up. Moreover, out of the three scenes, participants favoured the presence of DOF in the flower scene, as this was the scene with the highest number of edges and most variable depth.

The difference between age groups and their scene-dependent preferences was also evident in the stationery scene. In this scene, several big objects contained many small objects and were rich with edges and small cavities. For instance, the keyboard had many small keys and several cigars were lined up in the cigar box. As a result, this stimulus provided participants' with a strong monocular perspective depth cue. Indeed, this was evident, when participants compared (a) the Stereo condition with the DOF condition and (b) the DOF + Stereo condition with the DOF condition. In both cases, although, participants preferred the stereo

conditions, their preferences were weaker for the stationery scene than for the other scenes, on all three questions (depth impression, visual comfort and image quality). Similarly for the image quality question, participants preferred the Mono condition over the Stereo condition in the stationery scene. This suggests that when dealing with a detailed scene with fewer depth cues (monocular versus stereo) participants preferred to see the details as clearly as possible.

The preference for Stereo in the stationery scene could possibly be explained by the peripheral layout of the scene. When I designed the stationery scene, I tried to minimize occlusion of small objects and hence I spaced out the large objects throughout the virtual scene. Therefore, most of the big objects ended up either in the bottom of the scene or at the far right edge of the scene. Yet, DOF can only be an effective cue to depth or influence image quality, if the blur introduced can be detected by the visual system. In the stationery scene, when participants fixated a particular object, the rest of the objects in the scene fell in the peripheral region of the observer's visual field. However, the blur of the object on the retina is due to two factors: defocus and peripheral blur. Perhaps, people could not detect the defocus blur in the periphery or attributed blur in their periphery not as defocus blur but rather as peripheral blur. Unfortunately, there is very little research that addresses the combination of defocus blur and peripheral blur. It is possible that simulated blur due to defocus has to be stronger in the periphery in order to provide

less ambiguous depth cues to the user. van der Linde [202] was one of the few people who incorporated both types of blur into a gaze-contingent system, but he did not include any empirical evaluations.

4.4 Long Term Comfort

In the previous section, short-term exposure to different depth cues and observers' preferences was examined. In the study presented below, I was interested in examining observers' comfort ratings, when they were presented with different depth cues for a prolonged periods of time.

The main objective in conducting this experiment was to evaluate and compare a user's visual fatigue with stereoscopic and gaze-contingent DOF displays. The main hypothesis was that by combining DOF with stereo, it is possible to alleviate some of the negative effects associated with stereoscopic displays in general and specifically, with accommodation-vergence conflict (or more precisely with disparity-blur conflict).

The intent was to assign a representative and realistic task to a user for a prolonged period of time (such as object selection in 3D space). I also wanted to select a task that would require participants to interact with the VE - reaching and shifting their gaze to different regions of 3D space. The game of chess was very appropriate and met our requirements, as the game requires a detailed scan of all

the pieces on the board. The rules of the game also require players to perform piece selection and transfer to a new location. A between-subjects experimental design was used with half the participants playing the game with stereoscopic images (Stereo condition) and infinite DOF and half playing with stereoscopic images and realistic DOF (DOF + Stereo).

4.4.0.1 Stimuli

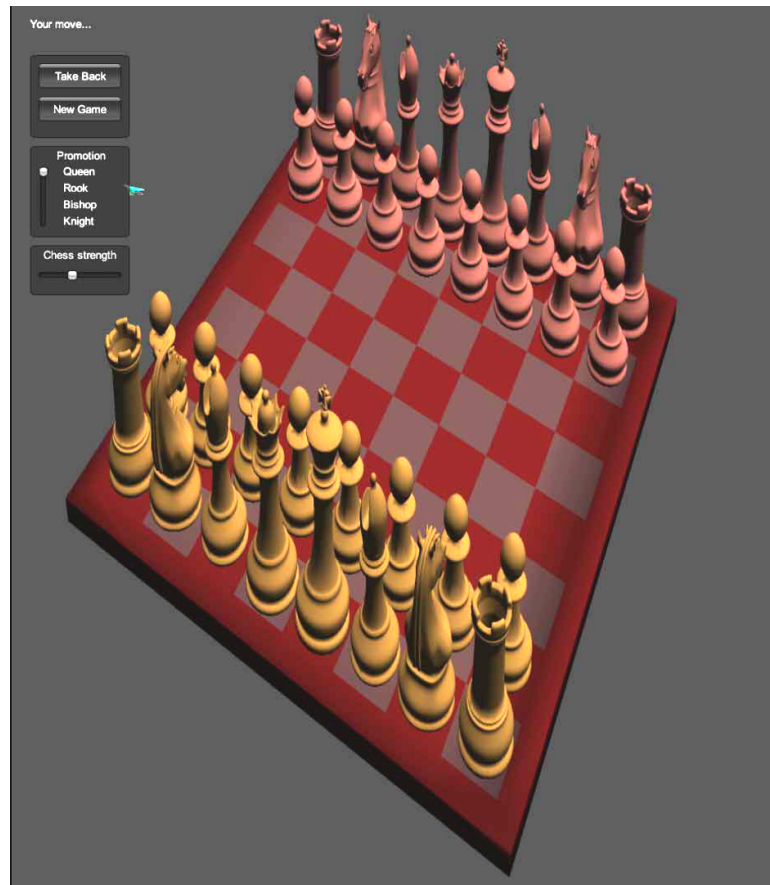


Figure 4.13: Stereoscopic chess board.

I adapted and incorporated a freely-available Unity3D chess project into the gaze-contingent system [32]. Consequently, the virtual scene consisted of a chess board and chess pieces (Figure 4.13). I controlled the camera convergence (horizontal image translation) so that the chessboard partially protruded from the screen. Thus, part of the board on the player's side was presented stereoscopically out of the screen plane, while the rest of the board that belonged to the computerized player was set further into the screen. By performing these manipulations, I created a focal range of 0.45 m. Specifically, the participant sat at a distance of 0.35 m away from the screen, the near part of the board was at about 0.27 m from the participant and the furthest point was about 0.72 m from the participant. I rotated and fixed the chessboard in such a way that the participant could easily select any square in 3D space. I also adjusted the color scheme of the board and chess pieces to minimize the ghosting commonly associated with stereo displays [194] so that it would not be perceptible by the user.

4.4.0.2 Procedure

The experimental session proceeded in three stages. The first stage was a pre-test stage. During that period of time, participants underwent a series of visual tests and viewed a short video about how to play chess. They also had a short non-stereoscopic demonstration on how to play the chess game. At the end of this

stage, participants had to fill out a fatigue questionnaire that served as a baseline for comparison. The second stage was the game stage, where participants played the chess game for 30 minutes. If a game was completed (by checkmate or draw), then another game immediately followed until the 30-minute interval was completed. It was very important during this stage that the eye-tracker was recording participants' gaze location accurately. Therefore, participants went through a calibration procedure before beginning the game. Special care was taken to limit the overall calibration time to no more than five minutes (up to three setup adjustments and calibration/validation attempts). To minimize visual fatigue due to calibration, the experimental protocol required dismissing participants who required longer calibration time (however, there were no cases that required prolonged calibration). Every five minutes throughout the game, calibration was evaluated and a drift correction was applied, if required. If the drift correction result exceeded a threshold of 0.5° , the eye-tracker was calibrated again. Although the drift correction check was performed approximately five times during the game, care was taken to limit calibration time, including any re-calibrations, to five minutes total as discussed above. This was done, again, in order to minimize visual fatigue. The third and final stage was a post-test stage. During that time participants repeated the fatigue questionnaire and visual tests performed in the pre-test stage. Each session concluded with a debriefing.

4.4.0.3 Participants

Each of the two groups of participants consisted of 11 university students (5 females and 6 males in each group, ranging in age from 18 to 29, average age 21.23). All participants had uncorrected distance visual acuity of 20/30 or better and could see clearly at the viewing distance without glasses. Written informed consent was obtained from all participants in accordance with a protocol approved by the York University Ethics Board.

As part of the pre-test, participants were screened for stereopsis (see Section 4.4.1) and were excluded if stereoacuity was worse than 50 seconds of arc at a distance of 40.6 cm.

4.4.1 Evaluation

4.4.1.1 Subjective Measures

One of the most common measures of visual fatigue involves questionnaires to probe the subjective experience. Unfortunately, a fatigue questionnaire that is sensitive, valid, reliable and robust has yet to be established [98]. Some researchers have asked only one general question. For example, Yano et al. [231] asked participants to rate how tired they were after an exposure. They used a five-point Likert category scale (“not tired”, “sense a little tired”, “little tired”, “tired”, “very tired”).

Similarly, Kooi and Toet [90] also asked their participants to rate viewing comfort with a five-point scale. On the other hand, other researchers used longer questionnaires that were modified from other standards and requirements like ISO 9241-3 [2]. Peli [153] asked participants to rate aspects of visual discomfort such as eye dryness, irritation in eyes, difficulty in focusing, postural discomfort and headache on a scale from none to severe. Kuze and Ukai [96] conducted a subjective evaluation of visual fatigue caused by moving images using a more extensive list of symptoms querying different eye sensations and postural pains. Similarly Duchowski et al. used questions such as “How tired are your eyes?”, “How clear is your vision?”, “How tired and sore are your neck and back?”, “How do your eyes feel?”, “How does your head feel?”. Interesting, for their main study they choose a 0-100 rating over the typical Likert scale claiming it provided a more detailed report of participants’ experiences. For my study, I adopted the questionnaire of Sheedy et al. [173, 174] that was developed based on clinical experiments and observations. This questionnaire identifies different symptoms based on sensation type and location, as well as internal and/or external symptoms. Specifically, the questionnaire can be divided into three categories. The first category concerns eye symptoms, and it included questions on tired eyes, irritated eyes, dry eyes, watery eyes, pulling feeling on eyes, burning sensation in eyes, ache behind or in the eyes, eye strain, uncomfortable vision, blurry vision, difficulty focusing and headache.

The second category is associated with motion, and is related to simulator sickness symptoms in virtual environments. It included questions on nausea and difficulty in concentrating. The third category was a musculo-skeletal category that included questions on neck pain, postural discomfort and muscular stress. In their review paper, Lambooi [97] recommended to not only use Sheedy’s questionnaire, but also to extend the questions to query additional background information such as previous experience. As a result, I added two more questions intended to reflect the general physical state of a participant. Specifically, I asked them to rate tiredness and sleepiness. Finally, I also included a demographic questionnaire that inquired about participants’ previous experiences with DOF + Stereo displays.

4.4.1.2 Measures of Visual Function

Visual fatigue is often associated with impacts on visual function, which can impact task performance. As a result, and due to the lack of standard, validated subjective measures, there has been interest in measuring oculomotor or visual correlates of visual fatigue. Two of the most commonly used measures are dynamic accommodative response and refractive error [231, 62, 81, 144, 200]. One of the main problems with these measures is that they require special equipment with very specific installation requirements. In addition, some devices provide very coarse results [97]. Due to the nature of the physical setup, this type of measure could not be incorporated.

On the other hand, changes in pupil diameter are associated with accommodation-vergence eye movements and hence have been proposed as a correlate of visual discomfort [201, 200]. It has been observed that the pupil constricts with near viewing to provide a large DOF and dilates at far viewing distances [78]. Pupil diameter was chosen over dynamic accommodation since that data was already available from an eye-tracker, while the latter required specialized equipment to measure. I expected to see an ongoing change in pupil size with change in participants' vergence. I also expected to see less variation in pupil diameter over time in cases of people experiencing visual fatigue.

Among many other possible physiological parameters, blink rate has been associated with fatigue and consequently, I wanted to examine if visual fatigue in stereoscopic displays is associated with a higher rate of blinking as found in other studies of visual fatigue [179, 87]. As such, I collected the number of blinks along with other eye-data from the eye-tracker.

Another approach that is commonly used for fatigue examination is clinical optometric measurement. In many cases these measurements are preferred, as they are easily accessible, noninvasive and can be used consistently with a large group of participants. Usually, these measures are taken before and after visual exposure to a stimulus. They are also helpful in the pre-screening process. I chose measures that were similar to Lambooi et al. [97], who recommended performing

the following tests: the visual acuity test, stereo acuity test and convergence test. Therefore, I also included phoria³ (horizontal and vertical) tests to be consistent with other studies that looked at visual fatigue with stereo displays [132]. All of the above measures were important as they allowed me to ensure that participants were successful at seeing the stereo content that was presented to them during the experiment. I also wanted to ensure that the tests I used were validated as standard clinical tests and provided important information about binocular visual performance.

Consequently, I had three types of evaluation. First, I evaluated visual and oculomotor functions with the following instrumentation:

- Stereo Acuity: RANDOT [®] stereo-test [Chicago, USA].
- Visual Acuity: Freiburg Vision Test [8].
- Near Point Visual Test: The head was stabilized on a chin-rest and participants had to fixate at the smallest line they could see on standard near-vision test card at a distance of 0.65 m away. Participants then had to slide test card toward their head along an optical rail until they could just see the selected line in focus. To achieve maximum accuracy, participants were

³Phoria is a visual test that evaluates vergence position of the eyes during monocular viewing, when participant's eyes accommodate at a certain distance. It considered to be a good measure of determining the most comfortable fusion distance for the user.

prompted to move the card to a position, where they detected the text to appear blurred and then move the card back until the text appeared sharp again.

- Vertical and Lateral Phoria: I used Titmus Biopter Vision Test [Petersburg, USA].

Second, I composed my own visual fatigue questionnaire as described above. Finally, I also collected and analysed eye-data, specifically blink rate and pupil size.

4.4.2 Results

4.4.2.1 Measures of Visual and Oculomotor Function

The results for the near point of clear binocular vision measure (near point vision test) are shown in figure 4.14. With a mixed-effect ANOVA, I found a significant interaction between the testing interval (pre-test and post-test results) and the different type of displays (Stereo or DOF + Stereo) ($F(1,20) = 8.248$, $p = 0.009$). This interaction is the factor of principal interest, as it reflects differences in exposure effects between the two display types. Specifically, participants in DOF + Stereo condition experienced an increase in near point distance, while participants in the stereo condition had their near point decreased (i.e., improved) on average

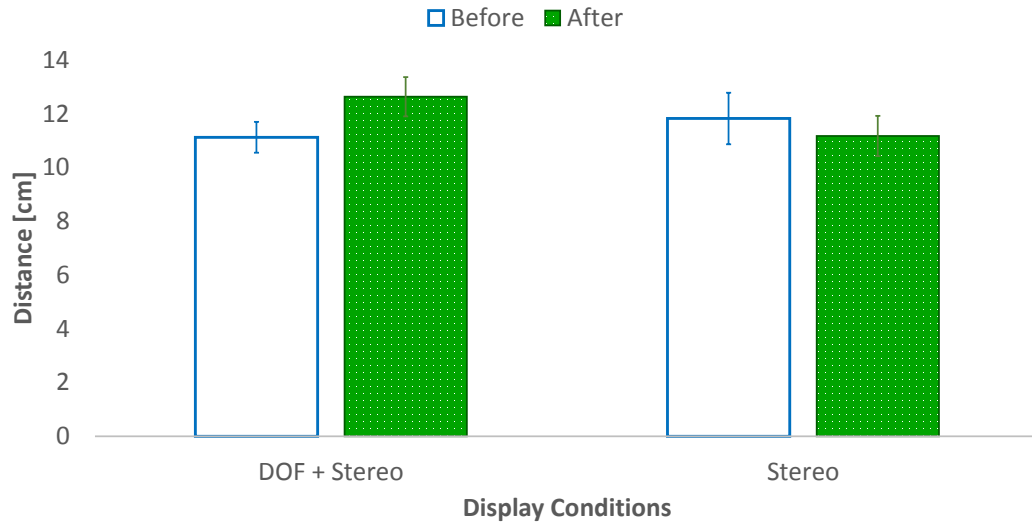


Figure 4.14: Mean before and after results for near point vision test for the two display conditions. The error bars indicate the standard errors of the means for each condition.

after stereoscopic presentation with infinite depth of field.

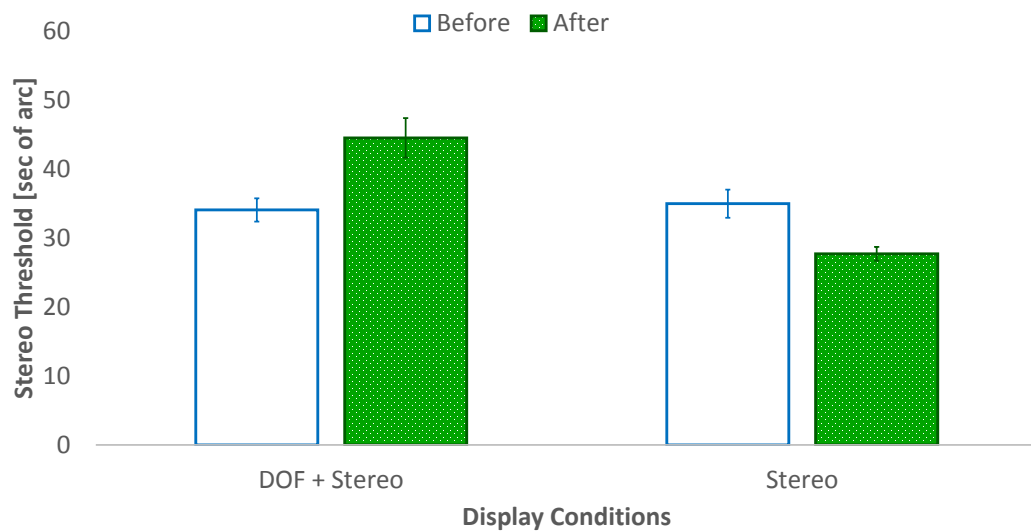


Figure 4.15: Mean before and after results for Randot stereo test for the two display conditions. The error bars indicate the standard errors of the means for each condition.

Similarly, a significant interaction was found between testing interval (before and after scores) and the two viewing conditions on the stereo acuity test ($F(1,20) = 4.67$, $p = 0.043$). In the DOF + Stereo condition, the stereoscopic threshold was increased on average relative to the baseline, while following stereoscopic displays without DOF simulation stereo threshold improved (declined) on average relative to baseline (results are shown in Figure 4.15, note that the decrease in stereo acuity actually corresponds to an increase in threshold seconds of arc).

The rest of the effects and interactions for the other measures were not significantly different between the two viewing conditions (Table 4.4), except for an interaction between gender, display type and the pre- and post- scores on the lateral phoria test. This seems to be mainly due to one female participant in the stereo condition, who had a very large change in phoria measurement. It is possible this is due to gender differences in perception of S3D [124].

Visual and Oculomotor Function	$F(1,25)$	p
Randot Score	6.183	0.0219 *
Binocular Acuity	0.707	0.41
Right Eye Acuity	0.005	0.945
Left Eye Acuity	0.000	0.995
Vertical Phoria	1.250	0.277
Lateral Phoria	1.763	0.199
Near Point	8.248	0.009 **

Table 4.4: Measures of visual and oculomotor functions: the table show the interaction between the objective measures before and after the chess game and the display type (Stereo or DOF + Stereo).

4.4.2.2 Subjective Fatigue Questionnaire

A summary of Kruskal-Wallis Chi-squared tests for the difference in exposure effects between before and after responses on the questionnaires collapsed over display condition is presented in Table 4.5. The results show that there was a significant difference between subjective responses before and after playing the chess game. Specifically, in terms of *general experience*, there was a significant difference between before and after for the *tiredness* question. In the *eye-symptoms* section, there were significant differences for the *tired eyes* and *uncomfortable vision* questions. For the rest of the questions in this category, the difference between the pre and post exposure was marginal. There was some gender interactions in terms of sensitivity to *eye ache* and type of displays presented. However, due to a small sample size, I cannot conclude if the effect was due to an outlier. In the *Motion* section, there was a significant difference for the *nausea* and marginal difference for the *difficulty concentrating* questions. Finally, in the *Musculo-Skeletal* section only *postural discomfort* was marginally significant.

The above findings indicate that participants showed an increase in fatigue after the game. The most relevant question is whether these effects differed between the two conditions. A summary of Kruskal-Wallis Chi-squared tests for the difference in pre- versus post-test change between the DOF + Stereo and the Stereo conditions

questionnaires are presented in Table 4.6. Significant differences between conditions were observed only for three questions. Specifically, in the *motion* (simulator sickness) category, there was a significant difference between stereo displays compared to stereo and DOF displays in terms of increase in nausea. Marginally significant differences were found in the case of difficulty in concentrating, where participants experienced a harder time concentrating following the Stereo condition than the DOF + Stereo. Lastly, in the musculo-skeletal condition, I observed marginally significant differences in ratings of postural discomfort, where participants experienced a larger increase in discomfort following stereo compared to DOF + Stereo. To clarify the magnitude of my results, the results for these three questions are presented in figure 4.16.

4.4.2.3 Eye-Data

The change in pupil size or change in blinking rate did not manifest any significant trends.

4.4.3 Discussion

The purpose of the study was to evaluate visual fatigue after prolonged viewing of stereoscopic displays by comparing two conditions: Stereo and DOF + Stereo. It was expected that there would be a significant alleviation of the fatigue symptoms

Questions	$\chi^2(2)$	p
<i>General Questions</i>		
Q1 Tiredness	8.313	0.004 **
Q2 Sleepiness	2.074	0.150
<i>Eye-Symptoms</i>		
Q3 Tired eyes	9.398	0.002 **
Q4 Irritated eyes	3.428	0.064 ·
Q5 Dry eyes	2.051	0.152
Q6 Watery eyes	2.932	0.087 ·
Q7 “Pulling” feeling on eyes	3.499	0.061 ·
Q8 Burning sensation on eyes	3.113	0.078 ·
Q9 Ache behind or in eyes	3.385	0.066 ·
Q10 Uncomfortable vision	5.484	0.019 *
Q11 Blurry vision	0.012	0.912
Q12 Difficulty focusing	2.834	0.092 ·
Q16 Headache	3.394	0.065 ·
<i>Simulator Sickness Symptoms</i>		
Q15 Nausea	7.401	0.007 **
Q18 Difficulty concentrating	1.177	0.278
<i>Musculo-Skeletal Symptoms</i>		
Q13 Stress	0.258	0.611
Q14 Neck pain	3.713	0.054 ·
Q17 Postural discomfort	3.679	0.055 ·

Table 4.5: Kruskal-Wallis Chi-square for difference in pre-test versus post-test responses to the fatigue questions (collapsed over conditions).

following DOF + Stereo relative to Stereo alone. I indeed observed some mixed indications that DOF might be a beneficial addition to stereoscopic displays.

In terms of visual and oculomotor functions, stereoscopic acuity increased and the near point of convergence decreased (i.e., both improved) post-game with stereoscopic viewing, but worsened with DOF. The improvement with stereoscopic viewing is consistent with stereoscopic adaptation that has been shown in the literature

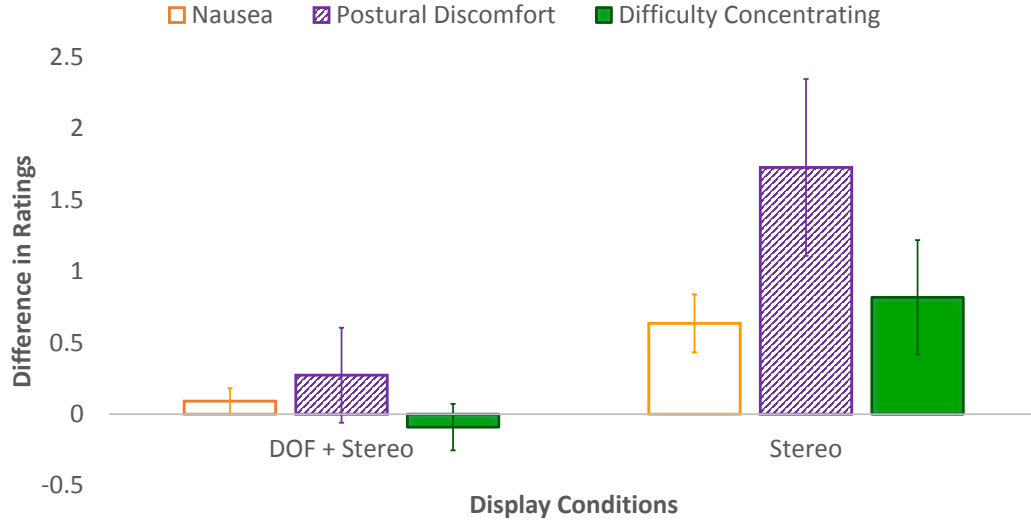


Figure 4.16: Change in subjective ratings after display exposure averaged across observers. The difference was calculated by subtracting the before ratings on the Likert scale from the after ratings on the scales. Error bars show the standard error of the mean for each condition.

following practice with stereoscopic stimuli [111, 20]. It is possible that this adaptation effect was absent in the DOF + Stereo condition because of the additional depth cue provided by DOF that reduced reliance on stereopsis and hence the stimulus to adaptation. As a result, in the DOF + Stereo condition, there was no significant improvement in stereo acuity and near point. As argued by Mather and Smith [121] the effective range of stereopsis and blur are complementary and in some cases visual blur is expected to predominate for relative depths that exceed the range of disparity mechanisms. Specifically Mitchell [129] showed that blur can limit the range of disparity processing and thus DOF simulation may have reduced the range of stereoscopic depth processing which, in turn, limited adaptive changes

Difference between before and after	$\chi^2(2)$	p
<i>General Questions</i>		
Q1 Tiredness	0.011	0.917
Q2 Sleepiness	1.279	0.258
<i>Eye-Symptoms</i>		
Q3 Tired eyes	2.016	0.156
Q4 Irritated eyes	0.287	0.592
Q5 Dry eyes	0.341	0.559
Q6 Watery eyes	1.014	0.314
Q7 Pulling' feeling on eyes	0.479	0.489
Q8 Burning sensation on eyes	2.335	0.127
Q9 Ache behind or in eyes	1.417	0.2339
Q10 Uncomfortable vision	1.123	0.289
Q11 Blurry vision	0.787	0.375
Q12 Difficulty focusing	0.121	0.729
Q16 Headache	1.432	0.232
<i>Simulator Sickness Symptoms</i>		
Q15 Nausea	5.091	0.024 *
Q18 Difficulty concentrating	3.2788	0.070 ·
<i>Musculo-Skeleta Symptomsl</i>		
Q13 Stress	0.158	0.691
Q14 Neck pain	1.43	0.231
Q17 Postural discomfort	3.286	0.070 ·

Table 4.6: Kruskal-Wallis Chi-square for differences between the Stereo and DOF+Stereo conditions in exposure effect (Post-test minus pre-test fatigue responses).

in oculomotor functions.

In terms of subjective measures, Stereo display induced more nausea relative to the DOF + Stereo display, which is consistent with a beneficial effect of DOF [132, 230]. Nausea is usually associated with simulator sickness. Therefore, this is a very important subjective measure and these results suggest gaze-contingent DOF might be beneficially incorporated in virtual environment applications such

as simulations and games. I also found that postural discomfort and difficulty concentrating was more of a problem after viewing displays with stereo alone than when DOF was added to stereo. This could be due to a direct effect of reduced cue conflict in the DOF + Stereo reducing visual fatigue. An alternative possibility is that, due to a more realistic display in the DOF + Stereo (compared to the Stereo condition), participants felt more immersed and had experienced fewer physical and cognitive distractions. Mauderer et al. [123] in their recent study have argued that gaze-contingent DOF contributes to the perception of depth (see also Experiment 1) and hence makes the virtual scene appear more realistic. As reported in the results section, no significant differences were observed between the two conditions on any other fatigue questions. This suggests that dynamic DOF may not be enough to alleviate all fatigue symptoms.

4.5 General Discussion

In summary, I built a gaze-contingent DOF system simulating the defocus blur that is associated with a participant's eye accommodating at different distances within virtual space. I also presented two studies to evaluate different aspects of DOF impact. Specifically, I looked at the effect of age associated with short-term exposure to DOF and I also looked at the effects of short and prolonged exposure to DOF. I hypothesized that DOF could improve the perception of depth in an older

population. However, the results showed that DOF degraded depth quality for participants who were not able to accommodate in real-life. On the other hand, for the young participants, I hypothesized that DOF could potentially enhance depth perception and alleviate the accommodation-convergence conflict. I demonstrated that simulated gaze-contingent DOF could indeed enhance depth perception when combined with other depth cues. Furthermore, for all age groups, DOF when combined with stereo did not seem to enhance short-term viewing comfort and has strong negative impact on perception of image quality. This implies that DOF is not a straight-forward solution for accommodation-convergence conflict. Nonetheless, I hypothesized that DOF might alleviate the fatigue associated with long-duration exposure to stereo display. The second experiment demonstrated that DOF indeed alleviates some symptoms of visual fatigue, but not all.

I believe that my studies also shed some light at why there are some inconsistencies between different studies. Based on the first study, I believe that special care should be taken when choosing participants for DOF studies/applications. One needs to be aware that a decrease in the ability of participant to accommodate, will also result in a decrease in the ability of the participant to enjoy the benefits associated with simulated blur. For example, a game for a young audience could choose to add DOF to a non-stereoscopic display and the participants should experience stronger depth, while for an application developed for a mature crowd

it is important to incorporate stereo and reduce blur as much as possible. When catering for the general public, one needs to find a balance to accommodate both age groups.

In their paper Zhang et al. [233] claim that DOF cannot be a primary cause of discomfort. They supported their claim by the O'Hare et al. study [142] that looked at prolonged viewing of stereoscopic images with different levels of DOF. Zhang et al. [233] study is similar to my long-term comfort experiment. It is important to reiterate that I have also found some evidence for a positive influence of DOF in extended viewing scenarios. Therefore, one of the contributing factors to the difference in visual comfort between studies is exposure time, which possibly activates different mechanisms during various periods of time. As I discussed in the introduction, visual discomfort can be broken into two stages: the visual stress associated with short exposure and visual fatigue that results from prolonged exposure. Another important factor that possibly attributes to the difference between studies is that the results from O'Hare et al. were not based on gaze-contingent DOF simulation. Although the task in O'Hare et al. required people to maintain fixation at a predefined distance, it has been shown that there is a significant difference between static and gaze-contingent DOF [74]. This could be due to the fact that eye-movement possibly contributed to depth perception and/or participants do not follow the assigned task.

It is also interesting to note that there was a significant difference between experimental results and the type of displays used during experimentation. For example, in their work, Hoffman et al.[76], looked at visual performance and visual fatigue with or without vergence-accomodation conflict. In their study, they used a volumetric display and they showed that such conflict indeed creates visual fatigue and discomfort. Duchowski et al. [51], on the other hand, tried to replicate Hoffman et al.[76, 75] but instead of using a fixed-viewpoint, volumetric display they used a gaze-contingent DOF stereo display, they did not find any difference in participant's comfort rating.

They attributed this to two flaws associated with the experimental design: the limited scale that did not provide enough room to exhibit the difference between conditions and prolonged breaks between each session and filling in of the fatigue questionnaire. Interestingly, Langer and Siciliano [102] also tried to replicate Held et al.'s [72] study by using DOF stereo displays with controlled fixation. Although the task was not related to visual comfort, but rather to a depth discrimination, both tasks require the viewer to resolve blur associated with defocus. Once again their work indicated that observers did not utilize DOF in contrast to Held et al. [72]. It is possible that there are methodological differences between studies. However, there is also a possibility that simulated DOF omits or competes with other depth cues that are naturally presented in a volumetric stereo display.

Generic simulations of natural DOF based on a thin lens plus aperture optical model have natural limitations. Although I did my best to match the blur to a realistic DOF, I could not account for individual optical differences in the participants' eyes. Such differences are idiosyncratic and dependent on pupil size, chromatic content and other factors [29]. Eye-tracking can provide online estimates of pupil size to allow for modelling of simulated aperture to actual pupil size. However, the DOF simulation modelled focal blur that results from viewing through a finite sized pupil but not the effects of aberrations and micro-fluctuations of accommodation that are believed to be important to disambiguate depth from accommodative blur [137]. Theoretically, if simulated depth of field matched a natural depth of field, the simulation should be indistinguishable from a real scene (in terms of image sharpness). Despite my best effort to minimize latency of the system, in some cases, the change in focus was still noticed by participants and it unfortunately likely reduced the comfort [90, 141]. Kooi and Toet [90, 141] reported that noticeable imperfections in a stereoscopic image can be described as very uncomfortable for an observer. It is possible that these technical limitations on spatial and temporal precision reduced image quality from expectations of natural viewing. However, it is also possible that participants actually preferred the image quality of the non-realistic scene with infinite DOF over their natural viewing. This is a reasonable supposition, since the image quality of the infinite DOF is better than expected

from natural viewing in terms of image sharpness and image information. Consequently, high-frequency content in the scene is preserved everywhere in the image. Therefore, it is possible that participants judged image quality as better in these ‘hyper realistic’ images than in more realistic, but blurry, DOF simulations.

5 Audio Gaze-Contingent Display

5.1 Introduction

When people interact with the real world they are typically inundated by a cacophony of sounds, sights, smells and other sensory information (unless they are in a controlled or sparse environment). How do people cope with the richness and amount of information provided to them through the different senses to make sense of the world around them? Fortunately, the brain has mechanisms that allow people to selectively attend to sources that are the most important in a given situation. This allows people to choose to redirect their attention from one source to another. In contrast, in virtual reality (VR) (also known as VE) applications, the information provided to the users is limited and carefully designed to accommodate the technological limitations (such as computational power) while meeting need for sufficient realism by simulating a subset of the abundance of cues normally presented in the real world. Usually system designers choose to concentrate on providing (simulating) cues stimulating one or two of the most dominant senses. The dominant senses

for human perception are usually thought to be vision and audition. When in conflict, it has been shown that audition is often overpowered by vision, such as in the well-known ventriloquism effect [79]. As a result, many VR applications use this effect and similar approximations to simplify the audio cues in the VR scene [37]. However, scarcity and impoverished spatial modelling of auditory cues compared to visual cues or the mismatch between the two modalities, can significantly reduce scene realism. This is because, in the natural environment, audio-visual cues are often tightly integrated, as in the McGurk effect [127, 41]. Consequently, it is very important to understand how merging the visual and auditory modalities within the VR world affects the perceived multi-modal environment so that the designer can present the user with a rich and realistic simulated world while complying with the system's limitations.

The survival of many species directly depends on successful sound discrimination and localization. Conversely, sound segregation and discrimination can be enhanced by localization [31]. For humans, each ear receives a sound pressure signal that is the sum of all the signals emanating from different sources and reverberating off surfaces in the environment. Auditory scene analysis relies on the brain exploiting consistencies in the world and between the ears, expectations of source and scene regularities, grouping of related signals, and other processes and assumptions to segregate sound sources and perceive objects in the world [172]. Successful auditory

scene analysis depends on selective attention, although there is debate whether this selective attention acts to select auditory streams or to influence the formation of these streams [172]. In turn, selective attention is affected by: (1) different characteristics of the sound environment (open, closed, furnished), (2) temporal, spatial, spectral and other similarities and distinctions between competing sources and (3) unique properties of the sound source (material or animal species) that can help to filter and discriminate spatially distinct audio streams. When dealing with speech, the spectral profile of the speaker's voice (average pitch and speed, sex and accent of the speaker), prominence of transitions such as subject matter, voice dynamics and syntax can be very important when attempting to attend to different speakers. However, it is not always enough to discriminate between sound signals; it is also important to recognize the location (origin and direction) of sound. Binaural sound localization can be explained by Duplex theory [181], that states that sound source location can be determined from interaural time difference (ITD), which is the time difference between the sounds reaching each ear and interaural level differences (ILD), which are differences in sound level entering the ears. Such cues are augmented by spectral cues that depend on the shape of the head and outer ear to spectrally shape the sounds arriving at each ear. The combination of these cues can be described by the so-called head-related transfer function (HRTF) that describes the transfer function of a sound signal from a spatially localized

source as it propagates to the inner ear [13]. Specifically, the HRTF is position, frequency, and distance dependent and describes the interactions of the sound with the listener's pinna (the grooves and notches of the pinna in particular), and to a lesser extent, the head, shoulders/torso. In general, none of these important spatial audio cues are presented in a typical audio-visual interactive application. Even in VR these are typically highly impoverished compared to real sounds and reduced to stereophonic or surround sound presentations.

As with vision, there are many cases when decoding the acquired sounds can be ambiguous. For example, without any additional cues, one cannot distinguish the location of sound sources originating exactly in front or behind a listener. In addition, sound might propagate through objects or reflect from surfaces. Hence, people utilize additional information about the source and environment to resolve ambiguity. For example, when dealing with live conversations, people pay special attention to body language and lip movements; when dealing with objects people pay attention to object's dynamics and special features that can suggest audio characteristics of that particular object; knowing the layout and objects within an environment can also be very helpful. Consequently, one of the fastest ways to collect the missing information is by visual observation. Hence, visual-audio augmentation plays a very important role in our daily interaction with the world. Numerous research studies showed how visual-audio augmentation can enrich experiences and how it can turn

into a distraction [94, 95, 228]. However, there has been very little research done in the area of incorporating visual attention in the augmentation of audio cues for more effective audio-visual simulations of real life scenarios. For example, Maddox et al. [114] compared the effectiveness of guiding attention with visual and auditory cues and found that visual guidance helps in sound discrimination, while auditory guidance does not. Therefore, I proposed to build an audio gaze-contingent system to explore the benefits of audio augmentation.

5.2 Previous Studies

When designing a multi-modal system that incorporates visual-audio attention, one has to understand where such a system could be beneficial and what the important requirements for such a system are. I envision that such a system could be greatly beneficial in any VR setup where multiple avatars are involved. This could be a social game or a virtual conference. To determine how these benefits could be manifested, one has to understand the underlying psychophysical theory, such as the *cocktail party (CP) effect*, and the technological developments, such as *GCDs* that could support such a system.

5.2.1 Cocktail Party Effect

The term cocktail party (CP) effect was introduced by Cherry [31]. It refers to the ability of a person to attend to a single speaker, when exposed to multiple concurrent speakers and also to attend to their name for example which may come from the “background”. The example that inspired the term, is a cocktail party, where guests try to participate in conversations in a room full of speaking people. Interestingly, as complex as such a task may be, people still manage to successfully communicate. During a cocktail party, a guest can practice one of two types of attention (1) selective attention, where a person attends to one source of information among many or (2) divided attention where the guest tries to attend to many speakers at the same time. Thus, CP research looks at selective auditory attention by assessing under which conditions a listener can attend to one of several competing messages.

5.2.2 Gaze-Contingent or Selective displays

One of the strongest indicators of a person’s attentional interest is her/his gaze direction. Although people can attend to peripheral stimuli (covert attention) typically they look at what interests them (overt attention). Thus, a user’s current gaze can be used as a strong indicator of visual attention. Systems that use gaze

to indicate attention and modify the display are referred to as gaze-contingent displays (GCDs). These applications use information acquired from an eye-tracking device and translate this information into adequate output on a display. The term GCD is usually used when such transformation happens without conscious selection of target by the user, while *selective displays* are applications that reflect intentional selection of particular item in the virtual or real world (e.g., an ‘eye mouse’). Often the GCD technology extrapolates gaze location in the scene from raw eye position and modifies the corresponding point or object in virtual space accordingly. Such modifications are usually designed to improve visibility and accuracy of the object of interest. Although most of the time such displays are based on visual (graphic) processing they could be extended to modify auditory content as well. An exception are gaze-contingent audio-visual substitution applications that perform sonification to represent data through sound [199, 112]. Yet, there are no GCD systems that utilize multi-modal (visual-audio) augmentation rather than the replacement of one sense with the other.

5.2.2.1 Midas Touch

As described in Chapter 1, a fundamental problem with interfaces based on gaze is the ambiguity of gaze-contingent selection [80], also known as the *Midas touch problem* [82]. In ancient legend, the gift of converting things into gold by touch

turns into a curse; similarly, the ability to select things by gaze can be a curse. This can be attributed to the fact that gaze is not primarily used for explicit pointing. Instead, most of the time, the human eye performs sensory input and eye movements support this goal with operations such as scanning and tracking. Furthermore, not all eye movements are voluntary. As a result, gaze contingent selection can affect areas in a scene that are not regions of attention. Common strategies to avoid the Midas touch problem are the use of intentional blinks or prolonged dwell time on the area of interest. Unfortunately, these solutions are more applicable when gaze is used as an explicit selection tool rather than in the context of gaze-contingent displays. Selection in GCD requires a natural (implicit) approach, so that the user does not need to consciously control their gaze [186]. To solve this problem in auditory GCD, Bolt [24] proposed using attention history statistics to control the volume level in one of several windows displayed if the participant paid extended attention to it for a significant period of time, and to decrease the volume if a given area was only briefly looked at. In other words, gaze history can be summarized with a heat map overlaying the scene as in Kunka and Kostek's [94] work. This information could then be overlaid with the 3D scene to determine which sound sources should be the most prominent and which should be de-emphasized to the user.

5.3 General Methodology and Experimental Approach

As I mentioned earlier, in a normal 3D world, the participant has rich spatial cues to help segregate and attend to objects and auditory streams. These are usually absent in a VE aside from basic (and usually uncalibrated to participant's HRTF) binaural cues from stereophonic audio. Consequently, in a series of experiments described in this chapter, I was interested in exploring the relationship between visual attention (as measured through gaze patterns) and the aural information acquisition of speech. Specifically, I wanted to identify the primary design choices that can lead to audio-visual human-computer interfaces. More specifically, can GCD be used to improve the way people acquire auditory information in simulations of real-life scenarios? Is it sufficient to manipulate the volume (increasing loudness) and quality (intelligibility) of audio input contingent on the participant's gaze-location to improve audio voice discrimination in 3D scenes? Similar to the graphical studies of Kistler et al. [88] and Pelechano et al. [152], my system simulated a VE with multiple avatars located at different positions in the virtual space. These avatars would speak at the same time to closely resemble a cocktail party scenario. To resemble real scenarios, the intensity of the sound projected from each avatar in the scene depended on two factors: the avatar's spatial location and the area of interest, which was driven by the 's gaze. Participant's gaze was used to identify the specific AOI

and this information was used to enhance the audio streams in a task-dependent manner to ease speech recognition and source identification in a cluttered audio scene. I hypothesized that gaze-contingent attentional selection could augment or substitute for normal spatial attention underlying the real-life cocktail party phenomenon and related effects. I used gaze as an indicator of current conversational interest and highlighted the attended source while de-emphasizing distracting speakers through volume manipulation. This gaze-contingent enhancement allowed simulation of the attentional effects of spatial auditory attention without full high-fidelity spatial sound simulation. I assessed these techniques both in terms of the impact on speech recognition and the naturalness of the VE interface. Two types of experimental design were used. First, I was interested in acquiring an overall user impression of my techniques along different aspects associated with VE. Second, I was interested in directly comparing the techniques to assess the relative effectiveness and naturalness of the interfaces.

5.3.1 System Setup

I used Unity3D to create and present the VE for the experiment. Besides the rapid 3D scene development, scripting options and experimental flexibility provided by this platform, Unity3D has support for manipulating sounds, modelling sound propagation in the 3D environment and customized sound manipulation. Specifically,

it incorporates a virtual microphone (*AudioListener*) that represents the listener. This object acquires all the sounds propagated from different sound sources within the virtual scene with regards to the spatial location of each sound source relative to its (*AudioListener*) location. The *AudioListener* was located at the main camera (i.e, the simulated user location). Unity also correctly distributed sounds to generate proper audio for the stereo headphones.

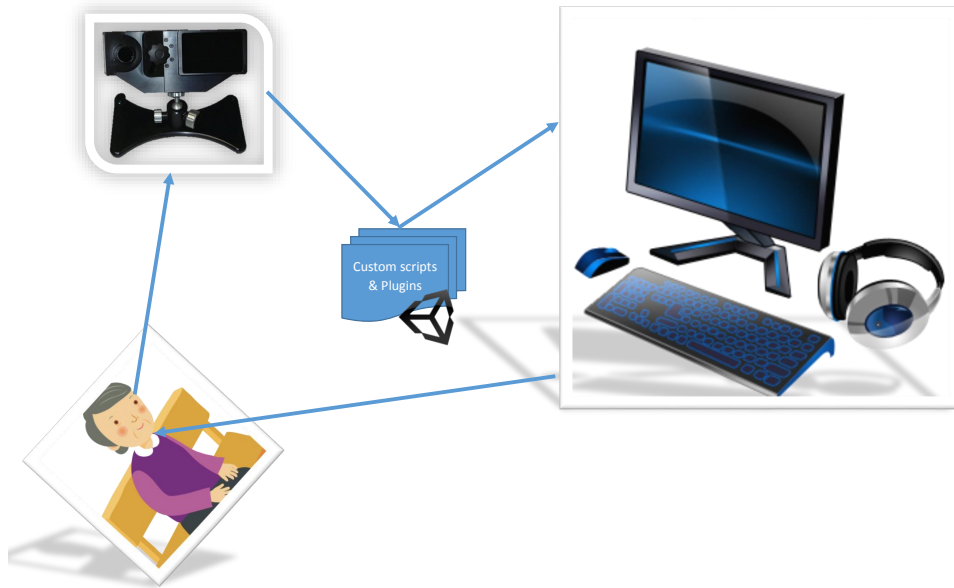


Figure 5.1: System's architecture.

As was noted by Witkin and Leventhal [228], sound proximity is biased by the presence of faces. Choosing the right facial representation (avatar) is very

important. To create animated avatars, I used Crazytalk7 Professional version. This application generated animations for characters, including their lip and facial movements, based on the audio and text files corresponding to their assigned speech segments. The fact that the user's view was fixed in the virtual scene during the experiment allowed me to use avatars with 2.5D representations (with no back meshes), which significantly reduced rendering complexity.

5.3.2 Apparatus

Visual displays were generated on a desktop computer with AMD FirePro W9000 FireGL, Windows 7 Enterprise, Intel[®] Core[™] CPU, 3.50 GHz, 3.50 GB Ram. The stimulus was presented on a Panasonic Viera TCP54VT25 54" diagonal HDTV plasma display with a pixel resolution of 1920 H x 1080 V and a refresh rate of 60 Hz. During the experiment, the screen was viewed binocularly at a distance of 1.6-1.7 m away from the participant. This viewing arrangement provided a window into the VE with a horizontal visual angle of $38^{\circ} - 41^{\circ}$ (as shown in figure 5.2). A real-time gaze-contingent system was assembled by incorporating an EyeLink 1000 (SR Research) eye-tracker, which supports on-line tracking of the participant's eye and head position. To provide maximum comfort to the participant I used a desktop mount that allowed free head movements. As a result, the participant's head location was tracked by localizing an identifying marker placed on their forehead in

the eye tracker video stream. The experiment was conducted in a darkened room to maximize gaze-tracking performance. To achieve realistic 3D sound, I used Sony MDR-XD200 Stereo headphones that transmitted stereo sound to the participant.



Figure 5.2: Experimental setup. The participant sat in-front of a TV screen, while her eyes were tracked by EyeLink 1000.

5.4 Experiment 1

In the first experiment, I evaluated different variations of gaze-contingent auditory highlighting in the context of the CP effect and subjective user ratings. Overall, I hypothesized that such gaze-contingent attentional selection could augment or substitute for normal spatial attention underlying the CP phenomenon and related effects. I assessed how the parameters of the GCD audio enhancement affected the

ability to recognize speech in a cluttered audio environment. Specifically, I compared six techniques that differed in the amount and time course of the attenuation of the background or the amplification of the attended speaker. In addition to examining different techniques, I varied the number of speakers. I asked participants to subjectively rate each technique on the task and on measures related to social presence and immersion. In addition, I had performance-based dependent variables, specifically, the accuracy of identifying the correct speaker and the response time between start of a trial and identification of the speaker. Finally, I recorded both the raw eye movements and the objects of interest for each given instance.

5.4.1 Methods

5.4.1.1 Task

The participants were asked to identify themselves as part of a discussion group, where all of their group members were virtual characters. Each trial consisted of a short discussion session in a VE. During the discussion session, each avatar spoke about a randomly selected topic. Each speech segment was a repeated monologue that was approximately one minute long. At the beginning of each trial (discussion session), the participant was given a keyword, which was the topic of one of the speeches presented by the avatars. The participants' task was to identify which

avatar was speaking about the indicated topic. The participants were instructed to identify the correct avatar in the most accurate and timely manner. As such, each of these discussion sessions is referred to as a *speaker identification trial*.

5.4.1.2 Stimuli

Six different gaze-contingent auditory techniques were implemented. The first technique was an *absolute sound cut-off* (A), where the observer was able to hear only one speaker at a time corresponding to their current conversational interest, while all other sound sources were muted. The second condition was a *partial sound cut-off* (P). This technique presented the fixated sound source at its peak volume, while all other sound sources were reduced to 20% of nominal volume (programmable volume supported by Unity⁴ for all participants.). The remaining techniques had filtered sound modulation, based on the time a particular target was fixated. In other words, the longer a participant looked at a target the louder it got, while other targets became quieter. Consequently, I had fast sound modulation up and down (FF) and slow sound modulation up and down (SS). Finally, I had volume that increased quickly with attention but faded slowly when unattended (FS) as well as the converse of slow increase with rapid fade (SF). Fast modulation from maximum volume to minimum volume was 1.13 seconds while slow modulation

⁴The volume of the headphones on the volume control options was setup to 50%

was 9 seconds. For the four cases (FF, SS, SF and FS) the background drop-off was capped to 20% of maximum volume (programmable volume), just as in condition P, and the foreground volume had the potential to reach 100% of maximum volume (programmable volume). All trials started with 50% of maximum volume(programmable volume). The modulation was exponential ($1 - e^{\frac{-t}{\tau}}$) to model the perceptual sound attenuation that would occur in the real world [180].

Seven distinct avatars were used in the experiment (see Figure 5.3). Despite their distinct facial features, care was taken to ensure that each avatar’s face subtended the same visual angle (5.55°). The avatars spoke behind a wooden stand that was marked with the avatar’s ID number. I expanded the bounding ellipsoid to include both the avatar’s upper body and the wooden stand, because there are instances when people pay attention to what the speaker is saying but no longer look at his/her face [212]. The number of avatars varied: each audio technique was tested with three, five and seven avatars. To control for the size of grouping (the user’s required field of view), care was taken to ensure that the first and the last speakers always appeared at the same locations on the screen; the rest of the avatars were evenly distributed out in the remaining space. Thus, I varied both the number and density of speakers in the scene.

The audio content was also carefully selected. While only male voices were used (it has been shown that people can easily discriminate between male and female

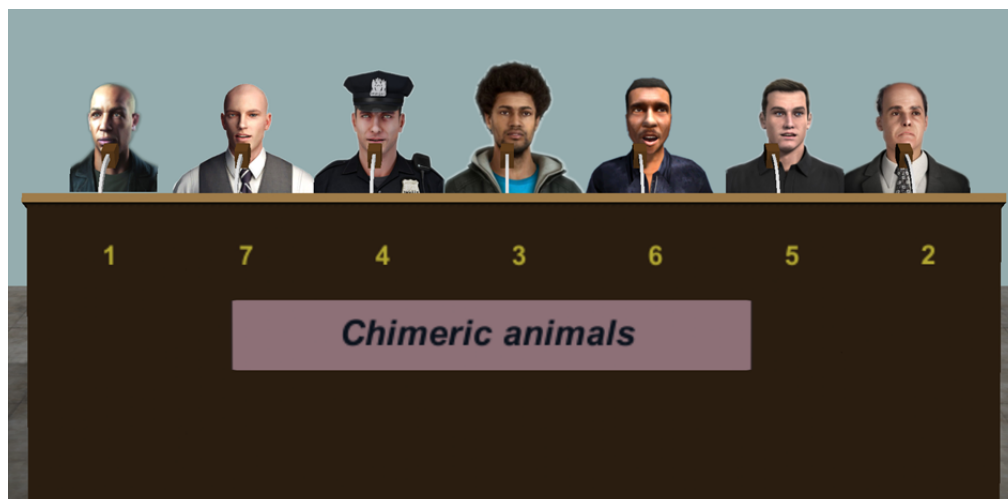


Figure 5.3: Example of a speaker identification trial with seven speakers.

voices), there was a variety of speakers to simulate a realistic scenario. I also used a wide range of topics, where speakers described a given term or situation. These clips were obtained from various pod-casting websites⁵). The keyword and keyword synonyms were repeated multiple times throughout each monologue. An example of text for the term “allowance” is provided below:

*Many children first learn the value of money by receiving an **allowance**. The purpose is to let children learn from experience at an age when financial mistakes are not very costly. The amount of money that parents give to their children to spend as they wish differs from family to family. Timing is another consideration. Some children get a weekly **allowance**. Others get a monthly **allowance**. In any case, parents should make clear what, if anything, the child is expected to pay for with*

⁵(a) ESL Podcast - http://www.ivoox.com/podcast-esl-podcast_sq_f1350_1.html; (b) Learn Out Loud - <http://www.learnoutloud.com/Podcast-Directory/Travel>; (c) Encyclopedia of Life - <http://podcast.eol.org/podcast>; (d) English Listening Lesson Library Online - <http://www.elllo.org/english/0701.htm>; (e) VOA Learning English - <http://learningenglish.voanews.com/>; (f) Culips English Learning Podcast - <http://culips.com>; (g) The English Desk - <http://englishdesk.blogspot.ca/>; (h) ESL Business News - <http://www.eslbusinessnews.com/>

*the money. At first, young children may spend all of their **allowance** soon after they receive it. If they do this, they will learn the hard way that spending must be done within a budget. Parents are usually advised not to offer more money until the next **allowance**.*

Finally, I made sure that there were no similar topics (or similar vocabulary) used simultaneously on any given trial.

The system supported two types of models that affected the sound volume for each speaker. The first was the speech sound model, which was based on the physics of each virtual object and its location relative to the user. This information determined the sound that the participant would hear. For example, further objects presented objects farther away resulted in quieter sounds to the listener's ear as opposed to objects that were closer, assuming both objects had equivalent intensity at their source. This is conventional sound source modelling for 3D VEs and was of limited use to the stationary participant in this experiment with equidistant speakers. The second model enhanced or substituted the user's auditory spatial attention and thus required determining the focus of the user's attention. For instance, avatars that the user looked at projected louder messages than those that were not of interest to the user. This was achieved by real-time gaze tracking [116].

5.4.1.3 Procedure

Each experiment started with participants completing a short demographics questionnaire, followed by an online audio test that checked their ability to hear over a range of frequencies (500 - 4000 Hz) [AuD OnLine Hearing Test]. The participants then performed a short English comprehension test to identify baseline for overall language proficiency and to train participants for the main experiment. Participants that had significant difficulty understanding spoken English were excluded. After completing the test, participants were shown a demo trial and were given instruction about how to respond during the trial.

The main experiment followed and was divided into six experimental blocks. Each block consisted of a calibration procedure, six speaker identification trials and a post-block questionnaire. During each trial, the participant listened to a group of avatars all simultaneously speaking on independent topics, as described in section 5.4.1.1. Once a participant was able to identify the correct avatar, he or she had to press the keyboard space bar. This action prompted to a new screen, where the participant was required to type in the ID number of the avatar that was identified as the target speaker. Participants could change their answers in case of any typing error until they registered it by pressing the space bar once more. After each block of six speaker identification trials, participants completed a

questionnaire that assessed their user experience during the block. This sequence continued for the duration of the experiment until all six blocks were completed. At the end of the experiment participants were debriefed and all of their questions were answered and any additional comments about the experiment were recorded.

Each of the six blocks was associated with a different one of the audio augmentation techniques discussed in section 5.4.1.2. The administration order of the blocks was randomized based on a Latin Square design. Each block consisted of six trials. The trials differed in the number of avatars speaking with two repeat trials for each avatar set size in each block. Trial order was randomized for each participant within each block. Similarly keywords, sounds and speakers' locations were individually randomized. Overall, each subject participated in 36 trials over the duration of the experiment (six audio conditions by three avatar set size conditions, each repeated twice).

5.4.1.4 Participants

I recruited 38 participants for this study; however, 10 were dismissed due to poor performance on the English language proficiency screening and poor calibration. Therefore, 28 participants took part in the full experiment (14 females and 14 males, ranging in age from 18 to 38, average age 25.25). All participants had uncorrected distance visual acuity of 20/30 or better, with good hearing in both

ears. They were all highly proficient in the English language. Written informed consent was obtained from all participants in accordance with a protocol approved by the York University Ethics Board.

5.4.1.5 Questionnaire Design

I developed the post-block questionnaire based on the Biocca et al. [18] social presence questionnaire. All questions were administrated with an on-line questionnaire and used a 7-point Likert scale to rate each question, with 1 as *never* and 7 as *always*. The questions are presented in table 5.1 and are grouped into several categories. The first category was the *Spatial Perception* category and included Questions 1 and 2. I was interested in these questions because, in the real world, to change the volume of the sound, one has to change the distance from the sound source. In other words, one has to approach the speaker or move away from the speaker to hear him better or to stop listening. Consequently, I expected that in their replies participants would indicate whether or not they needed the sound volume to change; I expected that this might be the case for the techniques with slow volume modulation (SF, SS and FS). In addition, I wanted to see if the people immersed themselves in the VE and had the urge to physically influence the sound quality. I asked more direct questions on this topic in the second category related to *VR Immersion* composed of Questions 3, 4, 5 and 6. These questions addressed

the participant's awareness of the discrepancy of the fusion between the real world and the virtual scene. I expected that the unnatural/uncomfortable trials would make participants more aware of the real world. The third category was *Visual Perception* and included Questions 7, 8 and 9. These questions looked at how, if at all, the auditory techniques influenced the participants' perception of visual experiences such as image quality. I hypothesized that there should be an effect of the techniques as the auditory quality was guided by the participant's visual attention. The fourth category was *Auditory Attention* and it included Questions 10, 11, 12 and 13, which inquired about whether the techniques made it easier or more difficult for the participant to direct her/his auditory attention to specific speakers. The fifth category, *Action-Reaction*, included Questions 14, 15 and 16 that dealt with the contingency aspect of my system, and checked whether people could predict, manipulate and get immersed in the virtual experience. I expected that a well-calibrated and responsive system would create a positive and interactive experience. Finally, the last section *Overall Experience* included Questions 17, 18, 19 and 20. These questions asked about the general user experience and I expected to see major differences between the implemented techniques on these questions.

Spatial Perception	Q1.	I wanted to listen to speakers from a closer distance.
	Q2.	I wanted to make specific sounds louder or softer.
VR Immersion	Q3.	I found it easy to forget that I am listening to virtual speakers rather than real people.
	Q4.	I found it easy to forget that I was watching a display.
	Q5.	I felt as if the speakers and I were in different places rather than the same room.
	Q6.	To what extent did you feel consciously aware of being in the real world?
Visual Perception	Q7.	I was distracted by the visual quality.
	Q8.	I liked the visual quality of the speakers.
	Q9.	I wanted to see the speakers more closely.
Auditory Attention	Q10.	I was easily distracted by other speakers, when listening to a given speaker.
	Q11.	Speakers were able to communicate their stories clearly to me.
	Q12.	I could pay close attention to a speaker I was interested in.
	Q13.	I was able to pay attention to multiple speakers simultaneously.
Action-Reaction	Q14.	I felt I knew what was going to happen next (in terms of sound).
	Q15.	Speakers paid close attention to me.
	Q16.	The speakers were affected by who I paid attention to.
Overall Experience	Q17.	It was easy to adapt to the current audio technique.
	Q18.	The task I had to perform was difficult.
	Q19.	I enjoyed listening to the current block.
	Q20.	The block was natural.

Table 5.1: List of questions in the post-block questionnaire (The list was shuffled before it given to participants).

5.4.2 Results

5.4.2.1 Psychophysical results

Errors Participants were required to identify the speaker corresponding to a given conversational topic but in some cases could not do so or they identified the incor-

rect avatar. The average number of these errors was 5.88 (out of 36 trials) over the entire experiment averaged across participants. Error rate was similar across techniques and only the SF condition had a slightly higher error rate (12.35% versus 8.13% across the other five conditions). As it was not clear what would be the participant's performance without augmented audio, it was planned to include a non-gaze-contingent condition, however, after a pilot study, this condition was not possible for the participants to do in a reasonable amount of time, if at all. Without the ability to utilize attention to listen to a particular speaker more closely, participants performed very poorly with only 42.0% correct identifications on average versus the 91.05% correct (across all six conditions) that was achieved in the present study.

Detection Time The keyword detection (target speaker identification) time across techniques and number of speakers is shown in figure 5.4. Analysis with a two-way ANOVA revealed significant main effects for the number of avatars ($F(1.84, 101.13) = 55.21, p < 0.001, \eta^2 = 0.501$) and audio technique ($F(2.971, 173.78) = 10.80, p < 0.001, \eta^2 = 0.164$) on the task performance time. Results also revealed a significant interaction between the number of speakers and audio technique ($F(2.175, 387.58) = 2.175, p < 0.05, \eta^2 = 0.038$). Detection time was larger for larger group of avatars than for smaller groups but the effects of audio technique was larger with an in-

creased number of avatars speaking. A pairwise comparison (with Bonferroni correction for multiple comparisons) revealed significant differences between conditions with three speakers and conditions with 5 speakers ($p < 0.001$), between conditions with three speakers and seven speakers ($p < 0.001$), and between conditions with five speakers and conditions with seven speakers ($p < 0.001$). Similarly, a pairwise comparison revealed significant differences between the A and SS techniques ($p < 0.05$), between the A and SF ($p < 0.001$), between the P and SS ($p < 0.001$), and between the P and SF ($p < 0.001$). Detection times were longest in the SS and SF conditions.

5.4.2.2 Questionnaire Data

I will present the data from the post-block questionnaires in terms of groupings of related questions as discussed in section 5.4.1.5. The data for this section will be presented with the box-and-whisker plots, where each box represents the lower and upper quartiles (top side - 25% and bottom side - 75%), the dark horizontal line inside the box indicates the median, the white dashed line indicates the mean, the extent of the whiskers show the minimum and maximum values that are not outliers and the stand alone points are outliers (more or less than 1.5 times the upper or lower quartile respectively).

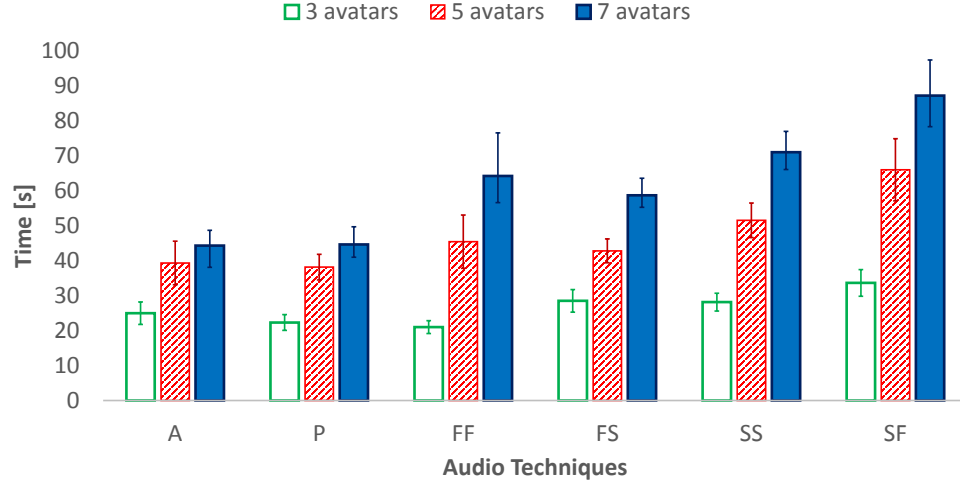


Figure 5.4: Mean keyword detection time as a function of technique and number of speakers, averaged across observers. The error bars represent the standard error of the mean for each condition. Absolute unattended sound cut-off (A); Partial unattended sound cut-off (P); Fast volume amplification for speaker of interest and fast unattended volume drop (FF); Slow volume amplification for speaker of interest and slow unattended volume drop (SS); Fast volume amplification for speaker of interest and slow unattended volume drop (FS); Slow volume amplification for speaker of interest and fast unattended volume drop (SF).

Spatial Perception The distribution of questionnaire responses on Question 1 and Question 2 are presented in figure 5.5. For Question 1 (‘I wanted to listen to speakers from a closer distance’), a Friedman test revealed a significant effect of technique on response type ($\chi^2(5) = 18.22, p < 0.003$). A post-hoc test using Wilcoxon Signed-rank tests with Bonferroni correction (used for all pairwise comparisons of the questionnaire data below) showed significant differences between techniques A and SF ($W = 1, Z = -3.26, p < 0.05, r = 0.44$).

Similarly, for Question 2 ('I wanted to make specific sounds louder or softer'), a Friedman test revealed a significant effect of technique on response type ($\chi^2(5) = 36.68, p < 0.001$). Post-hoc tests showed that participants wanted to change the volume of sounds significantly less often in the A condition than in the other conditions except for FF (n.s.). Responses on this question for the FF condition were also significantly different than SF ($p = 0.001$). Inspection of the figure shows that conditions with a slow increase of the attended to speaker (SF and SS) seemed more troublesome in terms of volume than the equivalent fast increasing conditions (FF and FS). A Tukey post-hoc analysis with Bonferroni adjustment confirmed that the conditions with fast increase in foreground volume were significantly different than those with slow increase ($p = 0.002$).

VR Immersion There were no significant differences between techniques for Questions 3 ('I found it easy to forget that I am listening to virtual speakers rather than real people'), 4 ('I found it easy to forget that I was watching a display'), 5 ('I felt as if the speakers and I were in different places rather than the same room') and 6 ('To what extent did you feel consciously aware of being in the real world?'). The average score for Questions 3, 4 and 5 were 3.4, 3.2 and 3.2, respectively, which indicates a *neutral* attitude. For Question 6 the mean response was 5.0, which suggests that participants were aware of the real world around them. Immersion

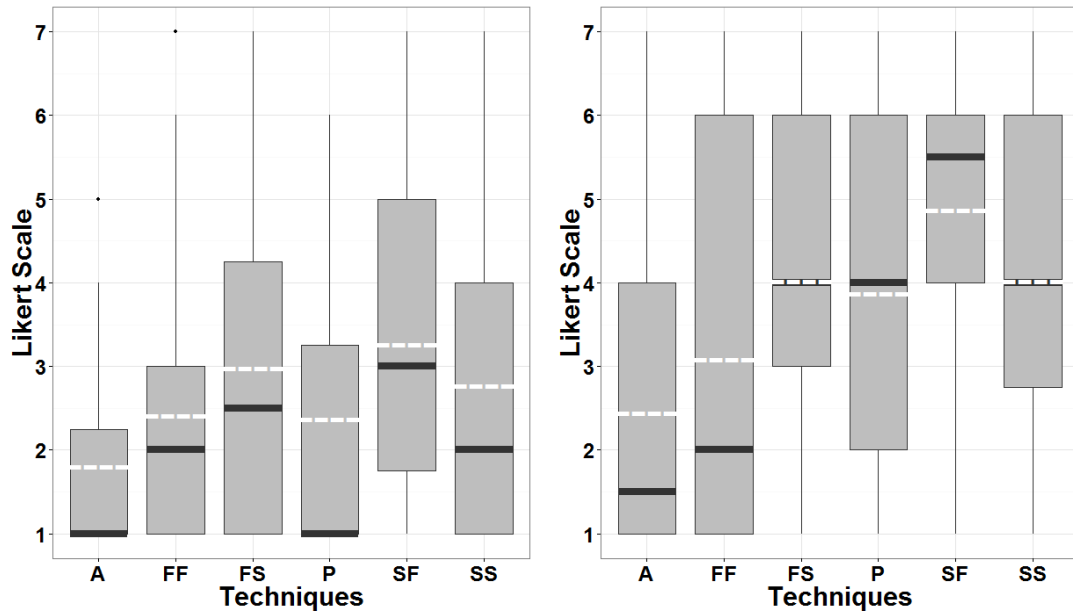


Figure 5.5: Likert results for spatial perception questions (with 1 meaning never and 7 always): Question 1 - I wanted to listen to speakers from a closer distance (Left); Question 2 - I wanted to make specific sounds louder or softer (Right). Each box represents the lower and upper quartiles (top side - 25% and bottom side - 75%), the dark horizontal line inside the box indicates the median, the white dashed line the mean, the extent of the whiskers show the minimum and maximum values that are not outliers, and the stand alone points are outliers (more or less than 1.5 times of upper or lower quartile, respectively).

seemed to be weak, presumably reflecting the fish-bowl VR ⁶.

Visual Perception For Question 8 ('I liked the visual quality of the speakers'), the effect of audio condition was not significant and the average score was 4.54, which indicated the participants' overall satisfaction with the visual content. In contrast, there was a significant effect of condition ($\chi^2(5) = 20.61, p < 0.001$) for Question 7 ('I was distracted by the visual quality'). The distribution of questionnaire responses to Questions 7 and 9 are presented in figure 5.6. Post-hoc tests for Question 7 demonstrated a significant difference between technique A and SF ($p = 0.04, r = 0.41$). For Question 9 ('I wanted to see the speakers more closely'), although there was a significant effect of technique ($\chi^2(5) = 12.56, p < 0.05$), post-hoc analysis did not reveal any specific significant differences between conditions.

Auditory Attention The distribution of questionnaire responses from Questions 10 ('I was easily distracted by other speakers, when listening to a given speaker'), 11 ('Speakers were able to communicate their stories clearly to me') and 12 ('I could pay close attention to a speaker I was interested in') are presented in figure 5.7. The effect of audio condition was not significant for Question 13 ('I was able to pay attention to multiple speakers simultaneously.') and the average score was

⁶A metaphor for creating a constrained VR within the constrained dimensions of a computer screen, as if the 3D world was placed inside a "fish-bowl".

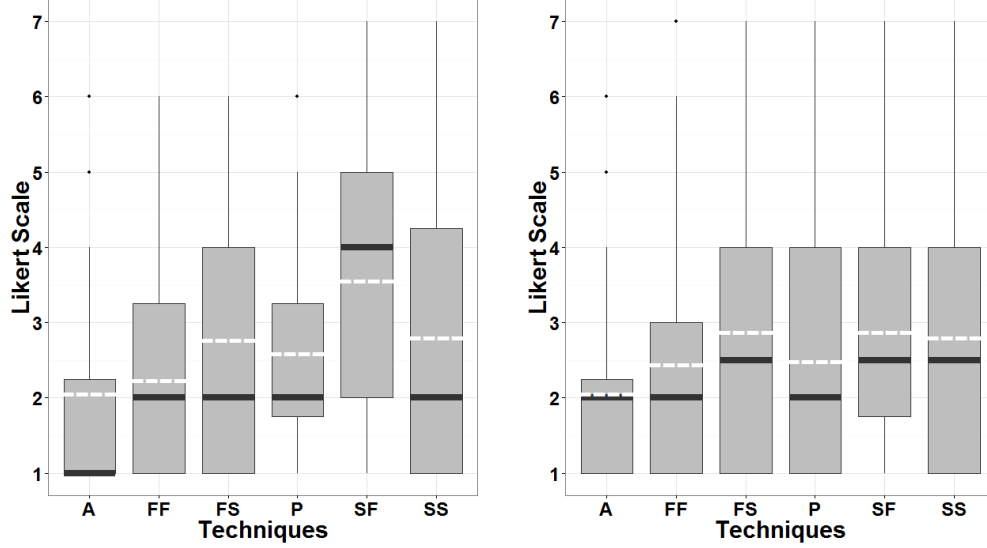


Figure 5.6: Likert results for visual perception questions (with 1 meaning never and 7 always): Question 7 - I was distracted by the visual quality (Left); Question 9 - I wanted to see the speakers more closely (Right).

2.8, which indicated that they tended not to attend to other speakers. On the other hand, Friedman tests for Questions 10, 11 and 12 showed significant effects of technique ($\chi^2(5) = 27.45, p < 0.001$; $\chi^2(5) = 18.21, p < 0.01$; and $\chi^2(5) = 18.37, p < 0.01$, respectively).

For Question 10, pairwise post-hoc tests showed significant differences (each $p < 0.001$) between techniques A and the rest of the conditions, except for condition FF. These indicated that unattended speakers were less distracting in the A condition. A post-hoc test for Question 11 showed significant differences between technique SF and each of A ($p = 0.006, r = 0.47$), FS ($p = 0.03, r = 0.42$), P ($p = 0.03, r = 0.41$) and (marginally) FF ($p = 0.05, r = 0.39$). Finally, a post-hoc test for Question 12

indicated that observers had significantly more trouble attending to the speaker in condition SF than in A, FS and P ($p < 0.01$, $p = 0.02$ and $p = 0.04$, respectively). In addition, Tukey post-hoc analysis with Bonferroni adjustment showed significant differences between conditions with slow and fast volume increase for the speaker of interest (FF and FS versus SF and SS $p = 0.019$).

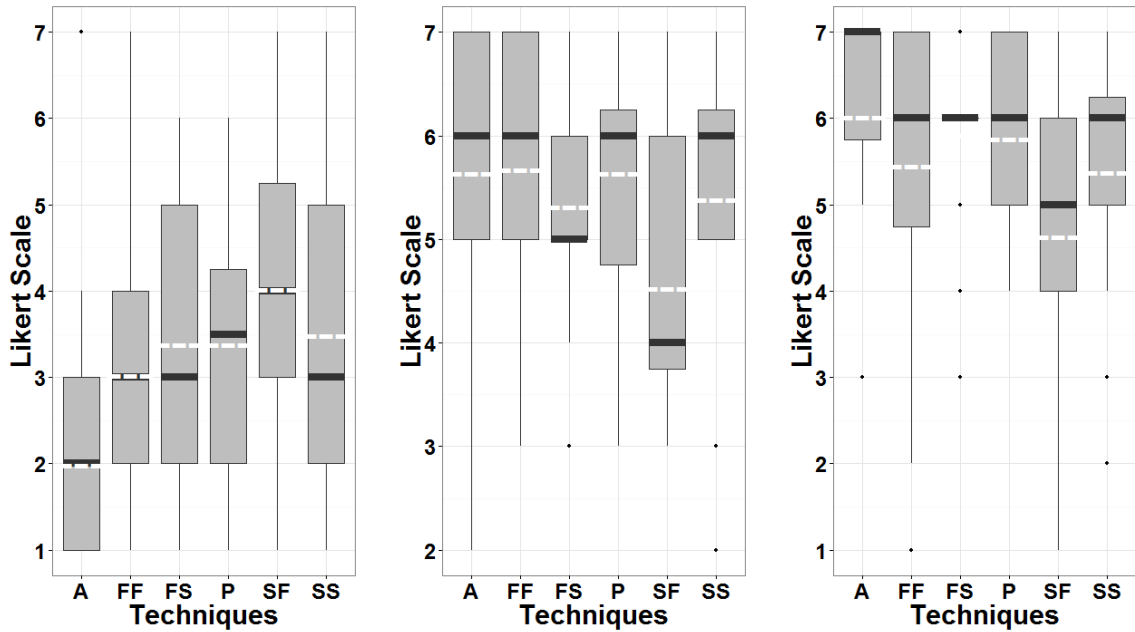


Figure 5.7: Likert results for audio attention questions (with 1 meaning never and 7 always): Question 10 - I was easily distracted by other speakers, when listening to a given speaker (Left); Question 11 - Speakers were able to communicate their stories clearly to me (Middle); Question 12 - I could pay close attention to a speaker I was interested in (Right).

Action-Reaction The distribution of questionnaire responses from Questions 14 ('I felt I knew what was going to happen next (in terms of sound)'), 15 ('Speakers

paid close attention to me’) and 16 (‘The speakers were affected by who I paid attention to’) are presented in figure 5.8. A Friedman test revealed significant effects of technique for each response ($\chi^2(5) = 12.12, p < 0.05$; $\chi^2(5) = 19.76, p < 0.01$; $\chi^2(5) = 12.50, p < 0.05$, for Questions 14, 15 and 16, respectively).

For Question 14, post-hoc comparisons did not show any significant differences between conditions, except for a marginally significant difference between condition A and SF ($p = 0.05$). For Question 15, post-hoc comparisons showed that condition SF scored significantly lower than conditions A ($p = 0.04, r = 0.40$) and marginally lower than condition FF ($p = 0.05, r = 0.40$). Also, Tukey post-hoc analysis with Bonferroni adjustment showed significant differences between conditions with slow (SS and SF) and fast (FF and FS) volume increase for the speaker of interest ($p = 0.005$), implying that in the slow volume increase conditions, participants felt that speakers paid less attention to them relative to the conditions where the foreground sound changed faster. Finally, for Question 16, post-hoc comparisons did not show any specific significant differences between conditions.

Overall Experience The distribution of questionnaire responses on Questions 17 (‘It was easy to adapt to the current audio technique’), 18 (‘The task I had to perform was difficult’), 19 (‘I enjoyed listening to the current block’) and 20 (‘The block was natural’) are presented in figure 5.9. A Friedman test for all four

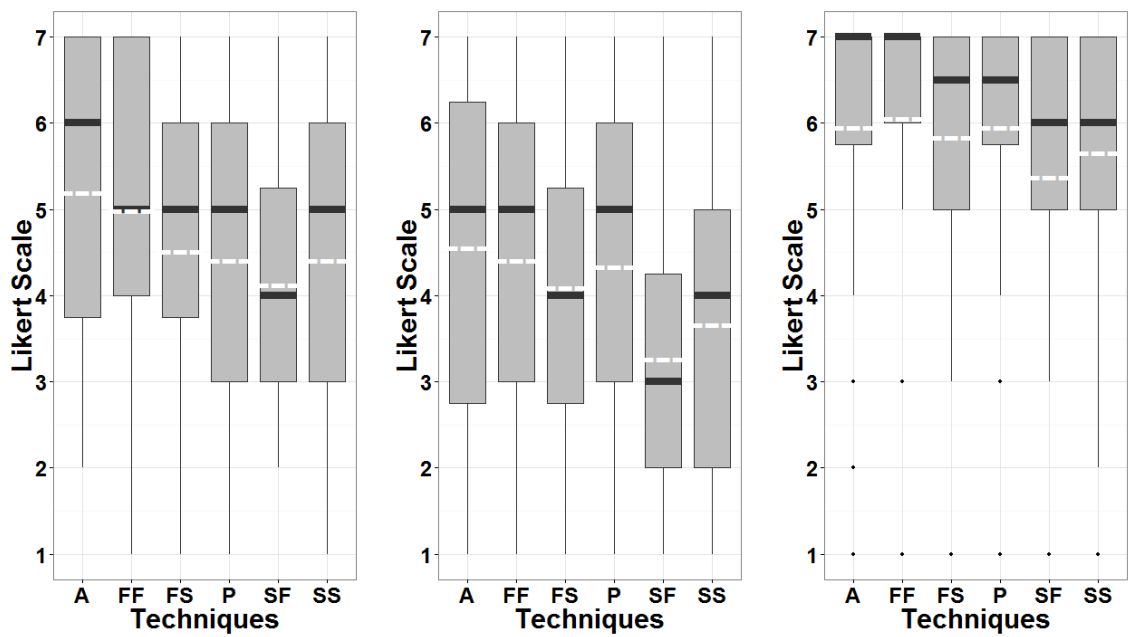


Figure 5.8: The distribution of questionnaire responses for ‘action-reaction’ questions (with 1 meaning never and 7 always): Question 14 - I felt I knew what was going to happen next (Left); Question 15 - Speakers paid close attention to me (Middle); Question 16 - The speakers were affected by who I paid attention to (Right).

questions revealed significant effects of technique (Q6: $\chi^2(5) = 34.77, p < 0.001$; Q18: $\chi^2(5) = 40.82, p < 0.001$; Q19: $\chi^2(5) = 20.0138, p < 0.01$; Q20: $\chi^2(5) = 22.29, p < 0.001$).

For Question 17, post-hoc comparisons showed that there were significantly lower ratings between slow volume increase in the foreground (SF and SS) and fast or instant switch to the voice of the speaker of interest (A, FF and P). Similarly, for Question 18, there were significant higher ratings for SF over fast or instant change in the foreground sound condition (A, FF, FS and P). Interestingly, for Questions 19 and 20, there were significantly higher ratings for no background sound (A) relative to conditions that had a slow increase in volume in the foreground voice (SF and SS). In addition, in Question 19, condition P with instant increase in the foreground volume was marginally higher than from conditions that had slow increase in volume in the foreground voice (SF and SS).

Finally, for all four questions, Tukey post-hoc analysis with Bonferroni adjustment showed that participants favoured the condition without background (A) relative to the rest of the conditions considered as a group ($p < 0.001$), except for Question 18, which was marginal ($p = 0.07$). Finally, rapid volume increase of the speaker (FS and FF) was generally preferred to slow volume increase (SS and SF); these differences were significant for Questions 17 ($p < 0.001$), 18 ($p < 0.001$), 19 ($p = 0.001$) and marginally significant for Question 20 ($p = 0.07$). Once again, in

all cases participants favoured the fast changing attended speaker conditions.

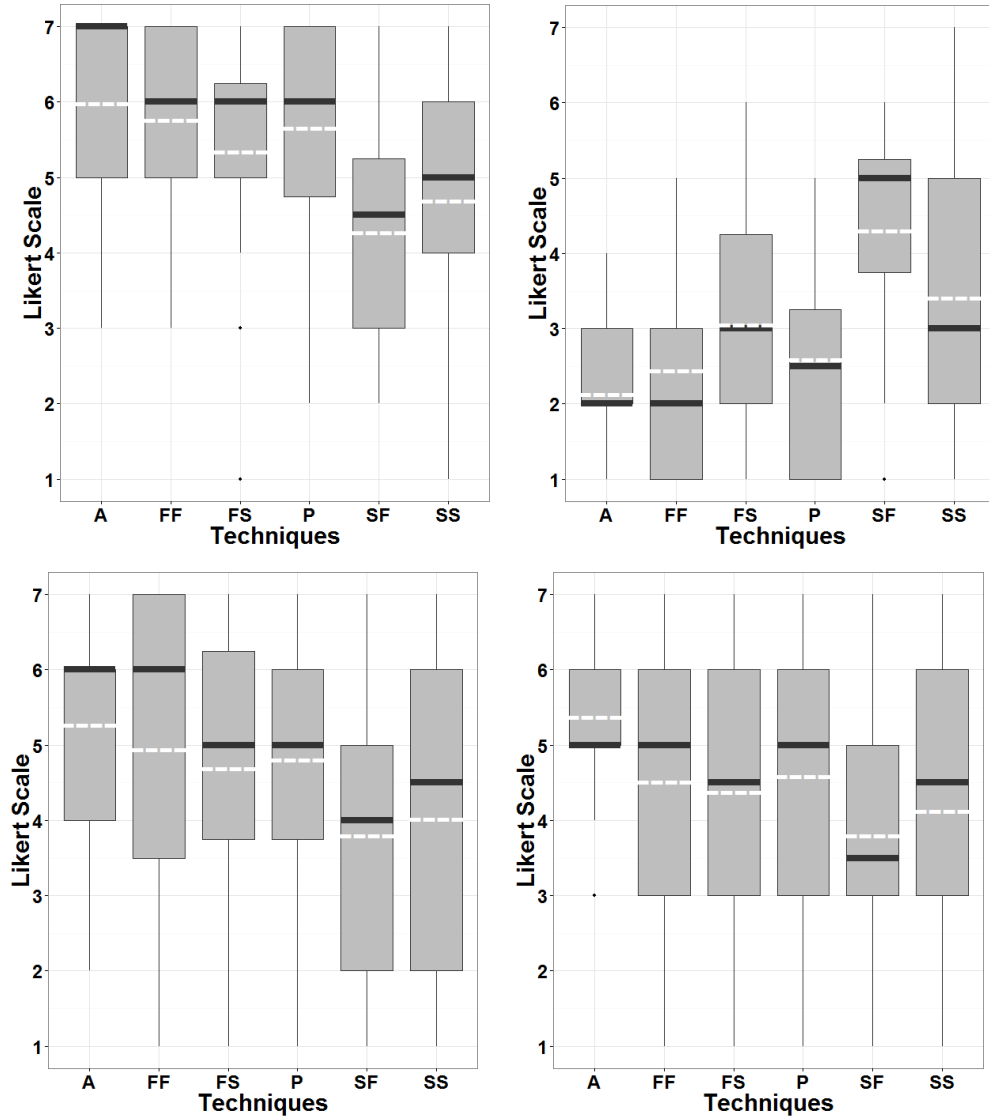


Figure 5.9: Likert results for overall experience questions (with 1 meaning never and 7 always): Question 17 - It was easy to adapt to the current audio technique (Top Left); Question 18 - The task I had to perform was difficult (Top Right); Question 19 - I enjoyed listening to the current block (Bottom Left); Question 20 - The block was natural (Bottom Right).

5.4.2.3 Gaze Analysis

To analyse the gaze data, I looked at the eye positions sampled during the experiment. In addition to the raw gaze positions relative to the display, I also recorded all the occasions that the participant’s gaze intercepted with a speaker or the pulpit in front of the speaker. This information was readily available as it was used to determine what sound sources to modulate. As a result, I could calculate the total time the participant fixated a specific speaker or when there was a fixation redirection from one speaker to another.

For the gaze analysis, I identified three categories of speakers: unclassified speakers (UA) that could be any speaker presented during a trial, speakers that the participant selected as his/her answer (SA) and speakers that were actually the correct speakers (CA). In many cases, the last two categories would refer to the same speaker, but not always. I then looked at different behaviours that the participants exhibited relative to these three types of speakers. I ran a two-way repeated measures ANOVA on the eye-data and looked at two factors — audio technique and number of avatars — and their interaction. In Tables 5.2 and 5.3, I report the significant results for all factors (I did not report interactions as they were not significant.).

Average fixation duration is presented in figure 5.10. From the figure, and con-

	Average fixation duration			Fixation redirection to a new speaker			Longest fixation time		
	$F(2, 110)$	p	η_p^2	$F(2, 275)$	p	η_p^2	$F(2, 110)$	p	η_p^2
UA	30.65	< 0.001	0.36	55.58	< 0.001	0.50	-	-	-
SA	12.38	< 0.001	0.18	6.20	0.004	0.10	3.41	0.04	0.06
CA	12.25	< 0.001	0.18	5.28	0.009	0.09	1.69	0.19	0.03

Table 5.2: A two-way repeated measures ANOVA for effect of number of avatars and participant’s visual behaviour (mean fixation duration, fixation redirection and longest fixation time) on the US, SS and CS.

	Total Time Fixating			Longest fixation time		
	$F(2, 275)$	p	η_p^2	$F(2, 275)$	p	η_p^2
SA	5.17	< 0.001	0.09	11.07	< 0.001	0.17
CA	5.30	< 0.001	0.09	10.28	< 0.001	0.16

Table 5.3: A two-way repeated measures ANOVA for effect of technique and participant’s visual behaviour (mean fixation duration, fixation redirection and longest fixation time) on the SS and CS.

firmed by the statistical analysis, one can see that the average fixation duration was highly dependent on the number of avatars but not the GCD technique. Post-hoc comparisons primarily indicated differences between conditions with three avatars compared to 5 ($p < 0.05$ for SA and CA; $p < 0.001$ for UA) or seven avatars ($p < 0.001$ for SA, CA and UA). In the general case (UA), there was also a significant difference between five and seven simultaneous speakers ($p < 0.001$).

On the other hand, the total time a participant spent fixating the selected or correct avatar (sum of all fixation durations on the target over the trial), depended more on the technique type than the number of concurrent speakers (figure 5.11) with more time spent fixating the selected/correct speaker in the SF and SS conditions. Post-hoc comparisons indicated significant differences between SS and A

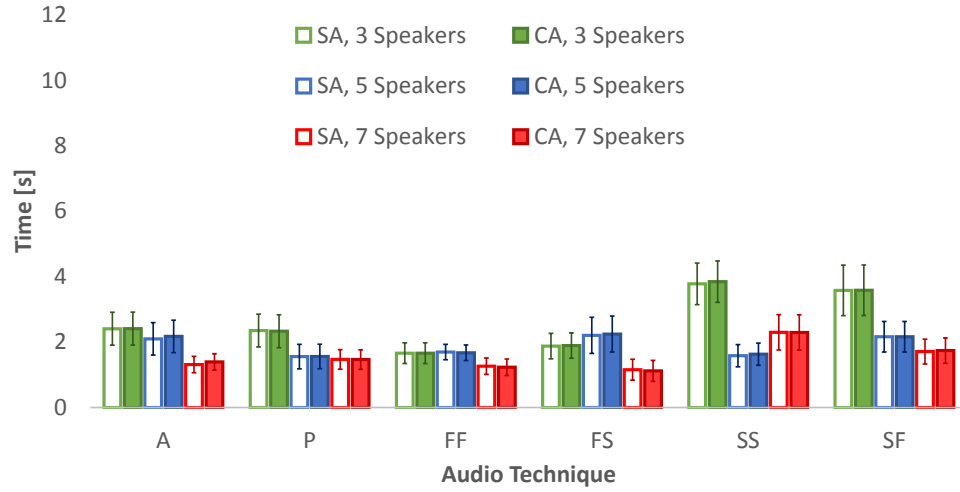


Figure 5.10: Mean fixation duration spent fixating the selected and correct speaker as a function of technique and number of concurrent speakers, averaged across participants. The error bars represent the standard error of the mean for each condition.

(SA: $p = 0.04$; CA: $p = 0.05$), P (SA: $p = 0.02$; CA: $p = 0.003$) and FF (SA and CA: $p = 0.02$) conditions with no evidence of any difference between the SS and SF conditions ($p = 1.0$ for both SA and CA).

The mean number of fixations that were redirected to a given speaker type in a trial is presented in figure 5.13. From the statistical analysis, one can conclude that the number of fixations redirected to another speaker was a function of the number of avatars. The number of fixation redirections increased with number of avatars in general ($p < 0.001$ for all conditions of three, five, and seven). On the other hand, participants performed significantly fewer redirection from the speaker

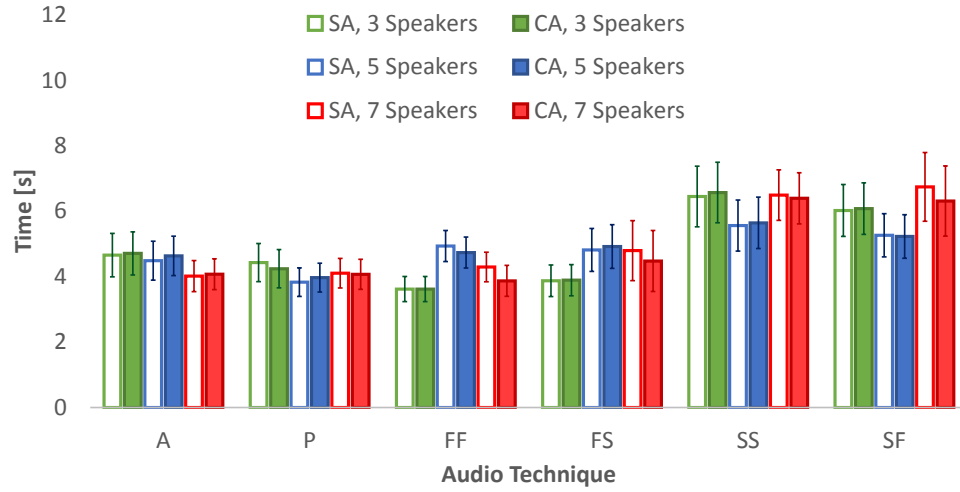


Figure 5.11: Mean total time (in a trial) that the the participant spent looking at the selected speaker as a function of technique and number of concurrent speakers, averaged across participants. The error bars represent the standard error of the mean for each condition.

when there were three speakers versus seven speakers for selected or correct avatar ($p < 0.001$).

Finally, the longest fixation duration spent on the selected or correct speaker was influenced by the audio technique used. Specifically, for the longest fixation duration spent on the correct or selected speaker, post-hoc comparisons indicated significant differences between SF and SS compared to the rest of the conditions ($p < 0.01$). Yet, there were no significant differences between SS and SF. Although there was a significant effect of number of speakers for longest fixation duration spent on selected speaker (table 5.2), post-hoc analysis did not reveal any specific

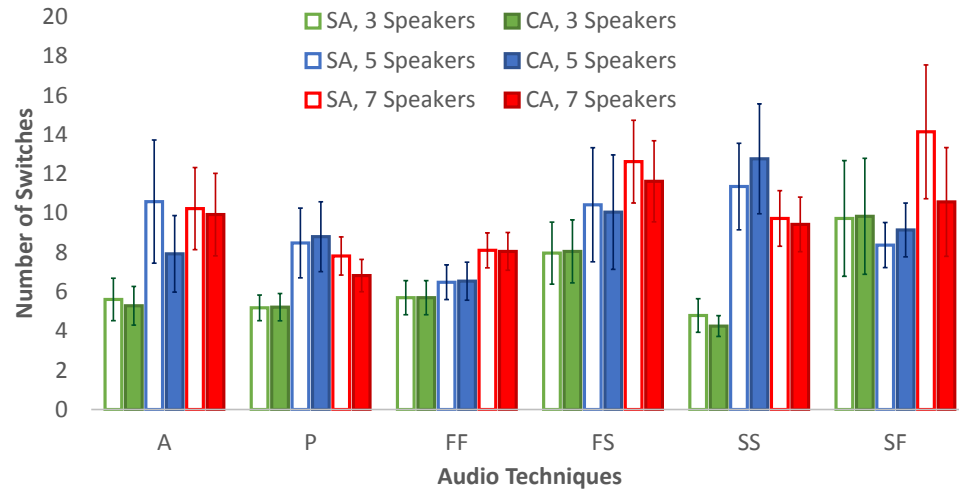


Figure 5.12: Mean number of times that fixation was redirected from speaker to another in a trial as a function of technique and number of concurrent speakers, averaged across participants. The error bars represent the standard error of the mean for each condition.

significant differences between conditions.

5.4.3 Discussion

In summary, let me highlight some important findings:

Errors: Low error rate for all conditions, with the highest rate for the FS condition.

Detection time: Significant main effects for the number of avatars and audio technique were found. There was also a significant interaction between the number of speakers and audio technique. Finally, the longest detection times were for the SS and SF conditions.

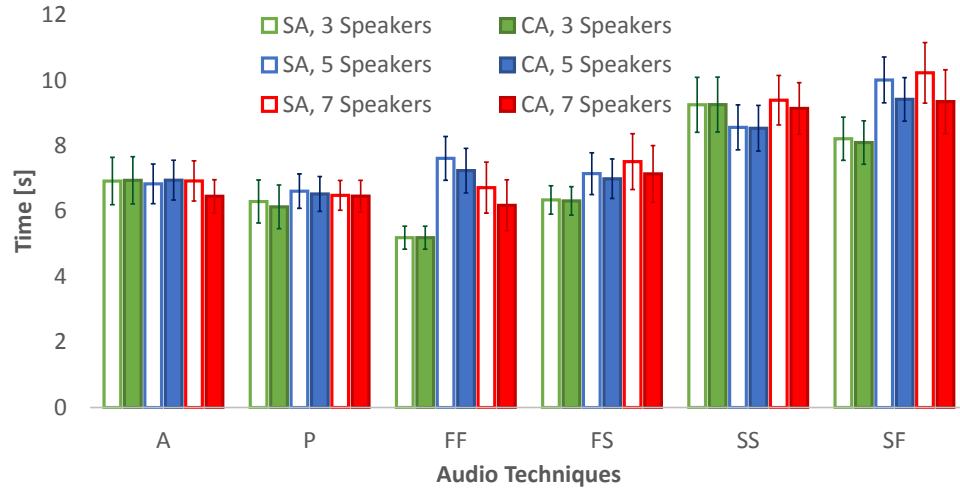


Figure 5.13: Mean of the longest duration fixation on the selected or chosen speaker in a trial as a function of technique and number of concurrent speakers. The error bars represent the standard error of the mean for each condition.

Spatial Perception: (Q1) Significant effect of the technique was observed. Specifically, participants wanted to be closer to avatars for condition SF than for condition A. (Q2) A significant effect of the technique was found, where participants wanted to make the speakers louder in the SF and SS conditions than in the FF and FS conditions.

Visual Perception: (Q7) Significant effect of the technique was observed, in particular, participants were more distracted by the visual quality for condition SF than for condition A.

Auditory Attention: (Q10) Significant effect of the technique was also noted,

where unattended speakers were less distracting in the A and FF conditions than the rest of the techniques. (Q11) Observers had significantly more trouble understanding the speaker in condition SF than in the rest of the conditions. (Q12) Observers had significantly more trouble attending to the speaker in condition SF than in A, FS and P.

Action-Reaction: (Q15) Observers felt that the avatar paid significantly less attention to them in condition SF than in conditions A and FF.

Overall Experience: (Q17) Observers felt that it was significantly less easy to adapt to the SF and SS techniques than to the A, FF and P techniques. (Q18) It was significantly more difficult to perform the task with SF technique than with the A, FF, FS and P techniques. (Q19) Participants enjoyed significantly more listening to speakers with the A technique significantly more than the SF and SS techniques. (Q20) Participants enjoyed significantly more listening to speakers with the A technique significantly more than the SF and SS techniques.

I will discuss the above results in more detail in the following section.

5.4.3.1 Psychophysical results

The number of errors was on average 8.84% and participants reported that the task was relatively simple. As predicted, the time to complete each trial was proportional to the number of avatars (and hence the set of alternatives over which the participant had to search). This is congruent to the *cocktail party literature* [128, 27], which has identified seven speakers as a limiting number of concurrent speakers for differentiating the voices. For example, in his study, Miller [128] showed that eight different voices produced very a strong masking effect. Due to the fact that sounds were modulated during the trial, i.e. not uniform, and that there were never more than seven concurrent speakers, this threshold was likely not reached. In contrast, as noted above the task was very difficult to perform without any modulation or augmentation.

I also observed that the time to identify the correct speaker varied with the technique used during the block. Trials for the condition, where the unattended speakers were rapidly attenuated but the speaker of interest was slowly enhanced, took the longest to complete. In contrast, trials where the voice of the main speaker was amplified rapidly or instantaneously, took the shortest time to complete. Thus, it appears that performance was most enhanced by accentuating the attended speaker than by other manipulations.

5.4.3.2 Questionnaire Data

Perception The general subjective success of the system was demonstrated by participant responses to Question 2 ('I wanted to make specific sounds louder or softer'). As expected, since I adjusted the volume of attended and unattended speakers differently for different conditions, it was natural that the participants would want to make some sounds louder or softer. Specifically, the significant difference between conditions where the attended speaker was rapidly amplified compared to where it was slowly modulated, highlighted the perceptual impact of the techniques. What was interesting to observe was that there were significant differences between responses for Question 2, which asked about the participant's desire to come closer or move farther away from the speaker. In other words, did participants want to change their spatial distance relative to the sound sources (speakers)? As expected, the most significant difference was between the SF and A conditions, which are the extremes in terms of noise level (none versus an indistinguishable background noise immediately after a change in attention). Observers, thus, felt little need to change position in the A condition but wished to be able to do so, on average, in the SF condition.

VR Immersion and perception There was no significant difference between the conditions on questions related to VR immersion, as participants felt neutral

about the virtual speakers, watching a display or virtual space in all cases. Although the display was large and filled a significant field of view, it is perhaps not surprising that sense of immersion was modest, since nowadays people are very accustomed to animated characters on a big TV screen ⁷. Interestingly, in Question 11, participants were more aware of being in real world than immersed in the simulated VE. This might be attributed the fact that they were forced to sit in relatively fixed position, and they had to answer the question after, rather than during, the block (i.e., after returning to the ‘real world’).

When asked about the visual quality of the speakers (Questions 1, 7 and 8), participants provided ratings that indicated that they were pleased with overall quality of the experimental scene and there were no significant difference in image quality across the conditions (there were no objective differences in image quality of course). Yet, when participants were asked about their satisfaction with visual quality in general, their satisfaction differed significantly between condition A and FS, indicating that they did associate the quality of sound with visual quality. This phenomenon has been previously reported [160, 61, 226, 119, 159, 39], specifically it has been established that quality perception in one modality, for example audio, can impact the perception of another modality, for example visual quality. Finally, when asked, if they wanted to see the avatars more closely, participants had sig-

⁷This is similar to what the theory behaind utilizing VR for fighting different phobias [140], but as part of a more natural and slower process.

nificant differences in their replies for condition A and SF, indicating again that they associated visual and auditory quality and also supporting my hypotheses that people’s natural spatial association of distance with volume would promote a desire to move closer when speech recognition was difficult.

Audio attention and action-reaction As discussed, it was very important to see participants’ ratings of techniques to the audio-attention and action-reaction, as these categories touch upon factors crucial in evaluating GCD applications. In particular, it was important to ensure that the behavioural consequences with GCD were as predicted, in other words, did not cause any side effects due to latency or other limitations [150]. Specifically, I observed that instantaneous or fast modulation of sound, as expected, supported the participants’ ability to pay close attention to speakers of interest, while delayed modulation led to the opposite effect. Similarly, participants found that background speakers were distracting in the SF condition and they had a harder time acquiring clear stories with the techniques that had slow modulation of the foreground sound. These observations emphasize the importance of building GCD systems with minimal, if any, latency [208]. Similarly, the different techniques controlling the sounds influenced how well participants could predict what will happen next. For example, in conditions, where the foreground sound changed instantaneously or rapidly, participants felt better

in control and were immersed more in the environment. Nevertheless, overall, the responses to the questions on audio attention and action-reaction confirm that my techniques indeed produced the desired cocktail party effect even in the SF and SS conditions.

Overall Experience As expected, participants found the SS and SF techniques to be the most challenging to adapt to, the most difficult to operate and the least enjoyable. Nevertheless, the responses were leaning more towards positive than negative rating. Finally, when asked how natural the techniques were, all techniques were rated as natural. The most highly rated technique was the A technique which is not in reality most natural as in the real world people cannot mute background speakers. Once again SF was rated as the most unnatural technique. It is possible that people operationalize natural in terms of how easy it is to interact with each technique. I believe that this requires further investigation. A more detailed study regarding what people think is easy and natural warranted(section 5.5).

5.4.3.3 Gaze-Data

In terms of gaze data, an interesting pattern emerged. Averaged fixation duration on the selected and correct speaker in each trial was the highest when there were only three concurrent speakers; this was particularly evident in the SS and SF cases.

This could be explained by the fact that participants had to divide their attention less between speakers when there were fewer speakers. In term of the total time spent on each speaker, participants fixated selected or correct speakers the longest in the SS and SF cases. This is because the participant had to wait longer to hear the speaker of interest at the maximum volume. This was not reflected in the average fixation time per speaker, but in the longest fixation they made to the correct or selected speakers. It is possible to assume that the average time was low because of the large number of attention switches or that some fixations were very short. Yet, the total number of attentional switches depended on the number of speakers in the scene rather than technique used. The search space depended on the number of speakers, and thus, participants made one and two fixations per avatar on average. It is possible that if a more complex sound modulation was implemented, the participant would have been able to listen to several speakers at the same time, hence reducing the number of switches. In addition, there was a difference in the number of fixations redirected to the correct versus the selected speaker. Specifically, the number of redirected fixations was higher for the selected speaker than the correct speaker for the five or seven concurrent speakers. It is likely, that when unsure, participants selected the speaker that they attended to the most.

5.5 Experiment 2

Following the first experiment, I was interested in comparing and assessing the realism and ease of use of the proposed techniques (Questions 17, 18 and 20) more directly. Specifically, I had participants explicitly compare techniques, while discriminating between five simultaneous speakers, as I have established that this number of speakers is already a challenging condition. I added a base case of no gaze-contingent augmentation, where I did not modulate the sound level of attended or background speakers. To compare the effectiveness of the techniques more precisely, I controlled the duration of each trial. The time constraint helped to make sure speech discrimination task would not reach the ceiling effect in terms of performance time. Thus, allowing to use error rate as a measure of task performance. I hypothesized that when directly asked to compare techniques, participants will be able to accurately identify which techniques are closer to what they would hear in the real world and disassociate the complexity of the test they had to perform from the realism question. Specifically, I expected that participants will find the non-gaze-contingent technique the most realistic and the technique without speakers in the background as the most unreal technique. I also predicted that they would find conditions with fast attenuation of the background and rapid enhancement of the speaker as the easiest to use, and this would be reflected in reduced error rate.

5.5.1 Methods

5.5.1.1 Task

The speech discrimination task from Experiment 1 was used. As before, participants had to listen to a short discussion session and identify the speaker who was discussing the identified topic. The only difference was that, while trial duration was unlimited in the first experiment (trials ended when the participant was ready to respond), trials in the current experiment always ended after 50 seconds. In addition, participants were asked to compare consecutively presented conditions. As a result, after every two trials, participants answered two questions:

- In which listening condition did the conversation feel more natural/realistic?
- In which listening condition was it easiest to attend to a speaker?

Figure 5.14 shows the sequence for each pair of trials. This sequence was followed through the entire experiment. To ensure properly maintained calibration, the experiment was divided into four blocks of six comparison trials (with two discrimination tasks per trial). Therefore, each participant was exposed to 24 trials (48 speech discriminations) in total. Assignment of trials to blocks was randomized for each participant, as was the order of the conditions.

5.5.1.2 Procedure

As in Experiment 1, each session started with participants filling in a short demographics questionnaire, which was followed by an online audio test⁸ and a short English comprehension tests. After completing the tests, participants were shown a demo trial and were given instruction about how to respond during the trial. This was followed by the experimental blocks. At the end of the experiment participants were debriefed and all of their questions were answered as well as any additional comments about the experiment recorded.

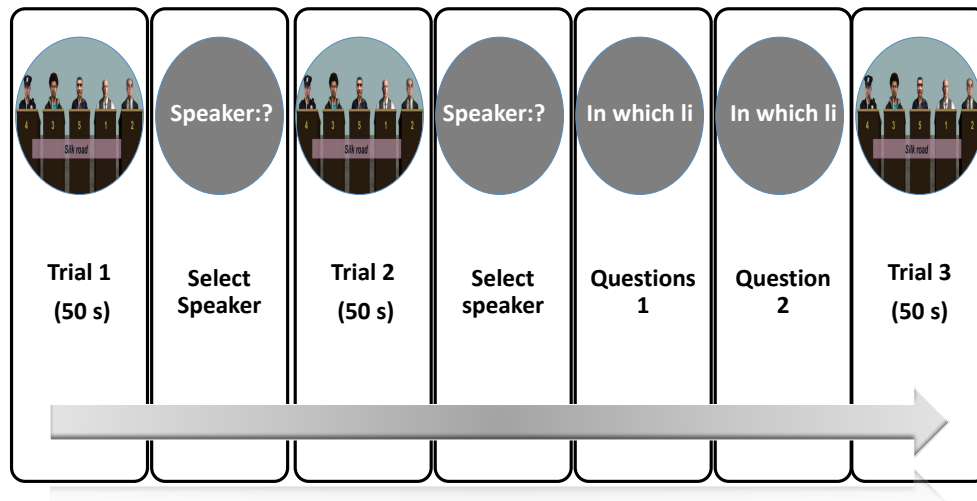


Figure 5.14: Procedure for experiment 2.

⁸The test required participants to adjust volume of tunes at different frequencies, same as in Experiment 1.

5.5.1.3 Stimuli

In this experiment, I used the same set of animated avatars and spatial layout as in Experiment 1. However, this time, there were always five concurrent speakers. I used the same corpus of audio clips as well, because a different set of participants was recruited for this study. Two groups of participants were used to compare different sets of techniques from the previous study. In addition, I also added a non-gaze-contingent condition (N). It was possible to introduce this condition since we limited time for each trial and could now compare this condition with the other conditions based on an error rate.:

1. Group 1: participants had to compare (1) non-gaze-contingent condition (N), (2) absolute unattended sound cut-off (A), (3) partial unattended sound cut-off (P) and (4) low volume amplification for speaker of interest and fast unattended volume drop (SF).
2. Group 2: participants had to compare (1) fast volume amplification for speaker of interest and slow unattended volume drop (FS), (2) slow volume amplification for speaker of interest and slow unattended volume drop (SS), (3) fast volume amplification for speaker of interest and fast unattended volume drop (FF) and (4) slow volume amplification for speaker of interest and fast unattended volume drop (SF).

For the first the group, I selected four techniques to compare, each of which varied in the degree of relative enhancement of the speaker and showed significant differences between each other for Experiment 1. Specifically, the two techniques that had most different responses were A and SF. I decided to include N as it was the nominal base case for a typical virtual environment (VE) and I observed that it was the most difficult case during a pilot study. Finally, P was conceptually a different case from A as it maintained speakers at the background while A did not. On the other hand, in Experiment 1, I observed that there was a significant difference between techniques based on sound modulation in the background and foreground and between the different speeds. Therefore, I used FS, SS, FF and SF trials in the second group to compare these dynamics. Finally, the repeatability of the experiment could be verified, since FF was very close to P condition and FS was used in both groups.

5.5.1.4 Participants

I recruited two groups of participants from an Undergraduate Research Participant Pool at York University. The first group included 19 participants (18 females and 1 male, ranging in age from 17 to 32, average age 18.89). Their English proficiency scores ranged from (85.71% to 100.00% with a mean a score of 98.12%). The second group included 21 participants (14 females and 7 males, ranging in age from 17 to

22, average age 18.94). Their English proficiency scores were ranged from (85.71% to 100.00% with mean score of 98.02%). All participants had uncorrected distance visual acuity of 20/30 or better, with good hearing in both ears. Written informed consent was obtained from all participants in accordance with a protocol approved by the York University Ethics Board.

5.5.2 Results

5.5.2.1 Error Rate

Unlike the first experiment, where the unlimited task execution time resulted in very few errors, this experiment had a limited trial duration, and hence, I expected to see a significant difference in error rate between conditions. Indeed, with one-way repeated-measures ANOVA, I found a significant main effect of technique used for both group 1 ($F(3, 54) = 45.18, p < 0.001, \eta_p^2 = 0.72$) and 2 ($F(3, 60) = 9.52, p < 0.001, \eta_p^2 = 0.32$). The distribution of percentage correct responses under different conditions is shown in figure 5.15. Overall, one can see that participants had the most accurate performance with A, P, FF, and FS techniques and the worst performance with N. Post-hoc paired comparisons using a Wilcoxon signed rank test with a Bonferroni correction were used to confirm differences between the techniques.

In group 1, the A and P techniques had the lowest error rates and were not significantly different from each other ($p = 1.0$). The error rate in the N condition was very high and significantly worse than the other three conditions (all $p < 0.01$). The SF condition had an intermediate error rate that was significantly higher than the A and P conditions ($p = 0.04$ and $p = 0.006$, respectively). Thus, all the augmentation techniques improved recognition performance compared to no augmentation but the condition with gradual attended speaker enhancement was more error prone than the instantaneous A and P conditions.

Conditions for group 2 were arranged to look at the effects of temporal dynamics on the augmentation. While the SF condition showed enhancement whereby participants were not as error prone as N (from group 1), it resulted in significantly more errors than the FS ($p = 0.04$, compared to SF) or FF conditions ($p = 0.008$ and $p = 0.01$, compared to SS and SF, respectively). Combined with the results for group 1, which showed low error in the instantaneous A and P cases, this suggests that the most important aspect of the augmentation is rapid enhancement of the attended speaker. The speed of attenuation of the background seemed to be of lesser importance, since there were no significant differences between the FS and FF conditions ($p = 0.51$) or between the SF and SS conditions ($p = 1.00$). The statistical analyses therefore confirm that the most important factor in determining the number of errors was the speed of modulation of the speaker of interest.

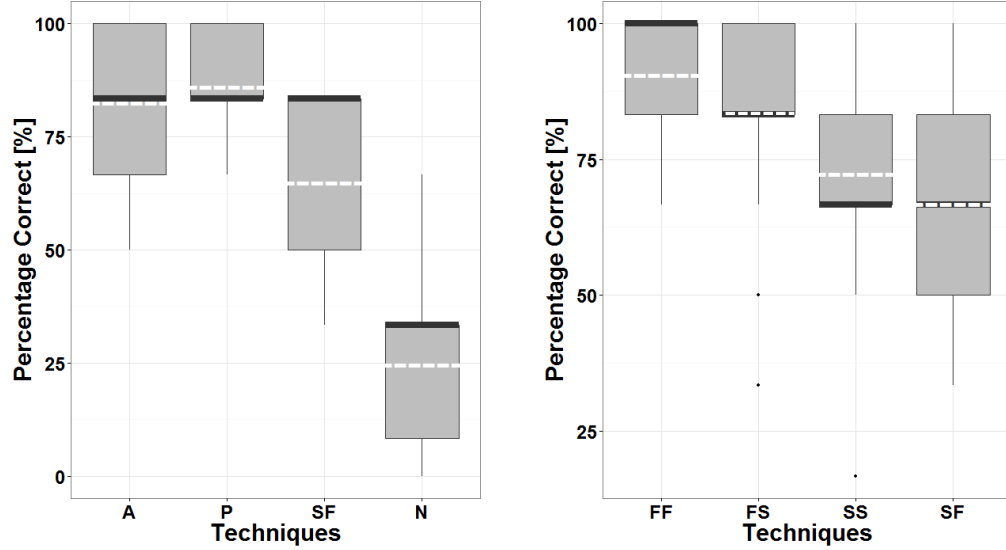


Figure 5.15: Percentage correct for speaker identification under different GCD techniques for two test groups: Group 1 (Left) and Group 2 (Right).

5.5.2.2 Preferences

To better understand the relationship between conditions, I performed a paired comparison analysis of the choice data. I used a Turner and Firth Model model [197, 198] that fits generalised non-linear models (gnm) using an over-parameterized representation.

The weight estimates for both groups are presented in Tables 5.4 (Q1) and 5.5 (Q2). The tables show the estimated weights from the fitted Poisson (log) model with the analysis repeated four times with each condition treated, in turn, as the base case (in separate rows). The estimated weights show the contribution of the augmentation technique to preference relative to the base case; a positive estimated

weight implies that the case is more likely to be preferred in comparisons than the base case.

Figure 5.16 shows preference responses for combination of conditions, polled over all participants. The y axis represents the percentage of times the choice listed first in the pair of conditions was selected (each pair was presented twice during the experiment, but in reverse order). The score of 50% indicated no preference between the conditions on average.

From the figure and the table of weights, one can see that, overall, very strong preferences between conditions in term of realism are lacking. The only significant difference was in group 1, between the augmented audio techniques and the base case (N). However, contrary to expectations, participants thought the augmented audio techniques were more realistic than the base case. The weight estimates and the preference ratings for the techniques with partially muted speakers (P and SF) were larger (more realistic) than for the A condition but this difference was not significant. There were no significant differences in realism between the different temporal dynamic conditions compared in group 2.

On the other hand, participants had strong preferences between audio techniques when asked which of the conditions made the task easier. Group 1 compared various levels of augmentation and the A condition was judged easiest followed by the P, then SF and then the N condition. Thus, the task was judged significantly

easier as the degree of enhancement of the attended speaker increased (from none, N, to absolute, A). This is reflected in significant differences in weights between all of these conditions. In terms of the temporal dynamics, the conditions with the rapid enhancement of the attended speaker (FS, FF) were reported to be easier than those with slow enhancement (SS and SF). For conditions differing only in the rate of attenuation of the background, when the enhancement of the attended stimulus was slow there were no significant differences (SF versus SS) and when the attended stimulus was rapidly enhanced there was only a marginally significant preference for a faster offset (FF versus FS). Thus, perceived task performance seemed to be mainly affected by the rate of enhancement of the attended speaker and the effect of background modulation on perceived task performance was small.

Group 1					Group 2				
	N	SF	P	A		SF	SS	FS	FF
N	0	0.33**	0.35 **	0.30*	SF	0	0.02	0	0.12
SF	-0.33**	0	0.03	0.03	SS	-0.02	0	-0.02	0.10
P	-0.35**	-0.03	0	-0.05	FS	0	0.02	0	0.12
A	-0.30*	0.03	0.05	0	FF	-0.12	-0.10	-0.12	0

Table 5.4: The chart shows estimated weights of audio conditions in determining preferences for Q1 (realism) for both groups. Each row shows the contribution of each audio technique (columns) to preference relative to the base case (1st column); a positive estimated weight implies that the case is more likely to be preferred. Significance of each contribution is shown by stars ($.p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

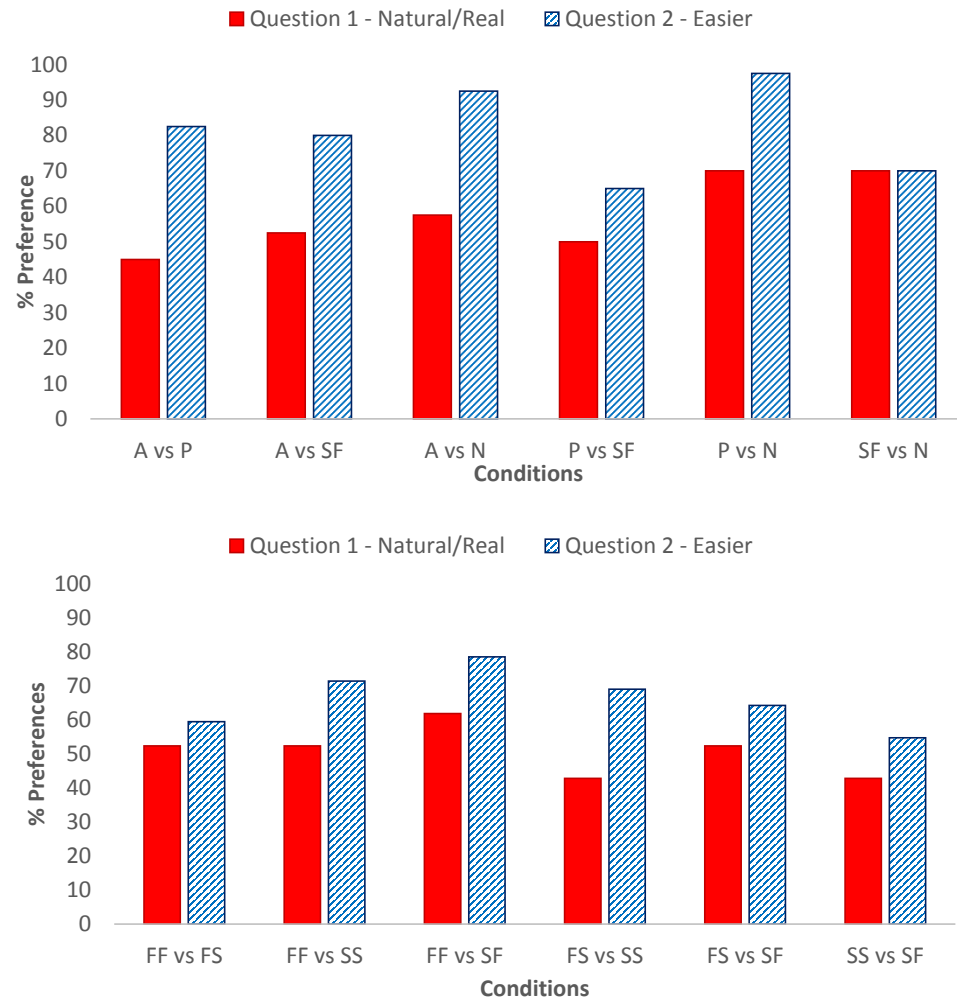


Figure 5.16: The mean preference pooled over observers for group 1 (top) and group 2 (bottom). Preferences are coded as the percentage of times the first condition in the pair listed was preferred over the second (note that first and second do not correspond to presentation order, which was randomized)

Group 1					Group 2				
	N	SF	P	A		SF	SS	FS	FF
N	0	0.64***	0.96***	1.47***	SF	0	0.07	0.38**	0.58***
SF	-0.64***	0	0.32*	0.83***	SS	-0.07	0	0.32**	0.51***
P	-0.96***	-0.32*	0	0.51***	FS	-0.38**	-0.32**	0	0.20.
A	-1.47***	-0.83***	-0.51***	0	FF	-0.58***	-0.51***	-0.20.	0

Table 5.5: The chart shows estimated weights of audio conditions in determining preferences for Q2 (easier) for both groups. Each row shows the contribution of each audio technique (columns) to preference relative to the base case (1st column); a positive estimated weight implies that the case is more likely to be preferred. Significance of each contribution is shown by stars ($p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

5.5.3 Discussion

The results showed that accuracy in identifying the correct speaker was highly related to the rate of sound modulation of the speaker of interest. This was evident in both the error measures and the subjective responses that indicated better performance for conditions with rapid or instantaneous enhancement of the currently fixated speaker. Perceived and objective task performance was primarily affected by the rate of enhancement of the attended speaker and the effect of background modulation on perceived task performance was small.

Rapid emphasis of the attended speaker is presumably beneficial, since it allows for the immediate analysis of the selected audio stream with higher signal to noise ratio. Therefore a performance improvement in these conditions was to be expected as the participant could pick up the keywords that would have helped them to identify the speaker earlier. I was concerned that a rapid onset would either be

(a) distracting if one made a chance glance to an avatar they were not actually interested in listening to (e.g., the Midas touch problem) or (b) be perceived as unnatural and distracting. Although a slower volume adjustment was arguably a more natural condition, it was not perceived as such and did not produce more accurate identification.

The synchronous phase of modulation of background and foreground sounds also played an important factor, especially in the slow modulation case. I expected that participants would have the hardest time with the SS case, where there was slow increase in foreground speaker volume and slow volume decrease for previously heard speakers. However, this technique was not as detrimental as the SF condition where the foreground volume was slowly increased and the background was dropped down rather quickly. I suspect that this was a more difficult case as it did not provide the participant the ability to relate previously attended sounds relative to the newly enhanced voice. As a result, the participant has to start the sound discrimination task from scratch.

Contrary to expectations, people thought that the augmented audio techniques were more realistic than the base case (N). The N case is the most realistic in the sense that the audio streams were unmodified. However, the scenario is unrealistic in that the richer spatial audio cues the participant could normally rely on in the real world were not present and there was no evidence of a cocktail party effect

(observers could not attend well to a single conversation in this condition). Thus, it may be that participant felt the augmented cases were more realistic since they provided the ability to attend to a single speaker in a cocktail party scenario that the participant expected from everyday life. I believed that the A condition would be judged as very unrealistic since the competing speakers were completely muted which does not correspond to the real world scenario. Although preferences were slightly smaller in the A condition than the SF and P conditions, this difference was small and non-significant. Similarly, there were no differences in realism among the different dynamic conditions. Thus, it appears that a participant's sense of realism can be maintained despite fairly unrealistic GCD audio enhancement.

5.6 General Discussion

In the two studies presented above, I examined several factors that are very important for simulating realistic VR for simulations and games - multiple virtual speakers and different sound modulation techniques. The ability to correctly and realistically simulate multiple avatars is important for studying social interaction mechanisms [225] or when building games [39]. Poor fidelity auditory cues can mislead the user about the layout of the environment in general and to pay attention to speakers that one would not attend to during a real life scenario. Furthermore, in the real world, there are many occasions similar to a CP scenario,

however, simulating the full range of spatial sound cues present in a realistic CP is very computationally intensive, requires obtaining individualized HRTFs, and is not always possible, hence simplifications are necessary. For instance, many virtual applications are limited to generated voices that are fairly similar or come from a limited range of variations. It was established that similar voices impair content discrimination [54, 28]. Indeed, in the experiment, participants had a hard time discriminating between speakers without any audio augmentation. However with augmentation, they performed exceptionally well with performance approaching expectations from real world experience. This evidence for behavioural improvement can be supported by Brungart and Simpson [27], who observed improvements in speaker segregation by amplifying the volume of a speaker. The results presented here are also consistent with Best et al. [17] who looked at spatial (*‘where’*) and temporal predictability (*‘when’*) of events that can significantly improve discrimination of the sound sources. Specifically, they showed that the spatial knowledge of where the object is located can help in enhance source identification. This was also supported by other studies that looked at the effect of gaze direction on sound source localization [106, 105, 114]. Gaze-contingent systems can help with this aspect as the participant always knows which speaker they are looking at each instance and can verify what they perceive by observing the speaker’s lip animation. Although I did not include questions about the speaker’s facial animation, on many occasions

during the debriefing session, participants reported that they used animation for verification and as an additional cue to help them to successfully complete an identification task. Finally, temporal information, according to Best et al. [17], can help in sound discrimination when targets are extremely similar or unfamiliar. The speakers were purposely selected to be similar to each-other (all male voices) and hence gaze-contingent techniques with instant or fast audio modulations were the most accurate as participants could associate the source with their actions through close temporal coupling. Furthermore, directing attention to a particular sensory channel not only helps in localizing and discriminating between sound sources, but also helps to overcome errors or imprecision in the other modality [160]. Hence, it is expected that since visual attention was involved, participants could better tolerate some audio inconsistencies, such as slow speaker modulation. Conversely, when attending to audio channels they likely could better tolerate any imprecision associated with visual display and gaze tracking.

5.7 Conclusions

In this chapter, I demonstrated the use of gaze-contingent audio display to augment user experience in social VEs. Specifically, I looked at how background modulation as well as change in volume for the speaker of interest improved speaker identification as well as how these techniques are rated by the users along different aspects

of realism and immersion. Overall, it seems rapid enhancement of the sound of the source one is attending to and a slower drop of the background sources can both improve system realism and simultaneously improve the quality and speed of auditory information acquisition.

6 Conclusions

In the course of my work I have developed three gaze-contingent applications that supported innovative studies of three different aspects of multi-modal display and multi-sensory perception. I developed new experimental test-beds and setup multiple experiments to verify how incorporating a gaze-contingent (gc) system can improve usability or facilitate the study of natural behaviour. Overall, GCDs proved to be a valuable tool to be incorporated in research and development. The specifics and contribution of each gaze-contingent system are outlined below.

6.1 Heading and driving

6.1.1 Summary

I developed a real-time gaze-contingent system for simulating visual defects in the VE application, such as a driving simulator. The system supported a wide range of defect simulation ranging from different sized occlusions to complex spatial perturbations that imitated metamorphopsia. I used this platform to examine the effects

of gaze-contingent alteration of regions of the visual field on heading estimation and online steering performance. This topic is of interest to the research community, as it is unclear to what degree peripheral visual information can compensate for central field loss/impairment. I conducted two experiments examining heading perception in the presence of these simulated gaze-contingent visual defects and also made a demonstration of the utility of the system in studying steering behaviour in the presence of these same defects (compared to control conditions with no distortions or scotomas). The results showed that the ability to accurately detect heading declined significantly in the presence of both types of visual field defects simulated (visual perturbations/distortions as well as scotomas). My results also showed that gaze-contingent simulation of visual defects is very useful for exploring different visual mechanisms associated with complex behaviour such as driving as well as studying the effects of different visual defects on everyday life activities in a reliable and repeatable manner.

6.1.2 Lessons Learned and Future Studies

As a possible extension for the current work, the GCD system could be used for collision detection and avoidance research. These aspects could be explored for both VE locomotion metaphors and to study real world collision avoidance. During the work I noted that it is hard to plan short term navigation paths especially when

one has to avoid certain obstacles on the way, particularly in the presence of visual occlusion or distortion. As the current work was based on very schematic visual defects, it will be interesting to extend it to more organic defects that would imitate what is viewed by a real patient.

6.2 Depth-of-Field Simulations

6.2.1 Summary

I developed a real-time GCD system for DOF, which simulates blur resulting from accommodation at the virtual depth of the point of regard. This was done in order to evaluate the benefits of adding gaze-contingent depth-of-field simulation to stereoscopic displays for VR applications. I proposed this would be beneficial because 3D displays are often associated with fatigue and major usability issues (such as visual discomfort, eye-strain, and even nausea). Consequently, I hypothesized that some of the discomfort and fatigue symptoms exhibited from viewing stereoscopic displays may result from a mismatch between stereopsis and blur, rather than between sensed accommodation and vergence. To test this hypothesis, I conducted a series of experiments that evaluated the impact of DOF on people of different age groups (younger versus older adults). The difference between short duration discomfort and fatigue due to prolonged viewing was also examined. Re-

sults indicate that age may be a determining factor for a user's experience of DOF. There was also a major difference in a user's perception of viewing comfort during short-term exposure and prolonged viewing. Specifically, people did not find that the presence of DOF enhanced short-term viewing comfort, while DOF alleviated some symptoms of visual fatigue but not all.

6.2.2 Lessons Learned and Future Studies

In future studies, I would like to examine an even longer exposure time, as half an hour may not be sufficient to produce a pronounced visual fatigue; under these conditions, I expect that differences between the two techniques will be more apparent. I would also like to extend the prolonged exposure study with stereoscopic displays to both the elderly and young populations and determine whether there are significant differences between the two groups and whether prolonged viewing improves depth perception for as elderly population. In term of more extended studies, it will also be interesting to look at more extreme cases of blur with gaze-contingent DOF or with DOF that is dynamically adapted to the user's instantaneous pupil size. I am also interested in studying more extreme disparity ranges, perhaps by combining our methods with other methods such as gaze-contingent disparity adjustment. The purpose of such simulations will be to improve user experience with S3D television and computer games, where disparities are sometimes exaggerated by separating

the cameras by very large (or very small) distances to enhance (reduce) depth which sometimes produces artefacts such as miniaturization or gigantism. Finally, I would like to incorporate DOF into other interfaces such as visualization applications or even in photo and movie editing applications.

6.3 Cocktail Party

6.3.1 Summary

I developed gaze-contingent display that allowed tracking of a user's gaze in real-time and modifications of the volume of the speakers' voices contingent on the current region of overt attention. This was done to evaluate the benefits of adding gaze-contingent audio modulation for VR applications. I then looked at six different techniques for sound modulation in two studies that compared the sound discrimination and user preferences of these techniques. The techniques were compared in terms of source recognition and preference in a set of user studies. I hypothesized that visual attention could be used as a tool to modulate the quality and intensity of sounds from multiple sources to efficiently and naturally select spatial sound sources. Overall, I observed that users liked the ability to control the sounds with their eyes. They felt that a rapid change in attenuation with attention but not elimination of competing sounds (partial rather than absolute selection) was most

natural. In conclusion, audio gaze-contingent displays offer potential for simulating rich, natural social and other interactions in virtual environments. They should be considered for improving both performance and fidelity in applications related to social behaviour scenarios or when the user needs to work through multiple audio sources of information.

6.3.2 Lessons Learned and Future Studies

During the two presented studies, users were positioned in one place and could not move within the environment. Yet when asked if they would like to come closer to the speakers they expressed an interest in doing so. As a result, I would be interested to see how users can perform sound discrimination tasks with or without attentional modulation while walking in the virtual space. In addition I would like to look at more complex algorithms for sound modulation. Specifically, it would be interesting to develop gaze-contingent techniques with different levels of background drop-down incorporating attentional history. Such a technique could, for example have the voice of speaker of interest to be at the maximum volume in foreground and the voice of speaker one attended to previously at the mid-ground, while the rest are at the lower background level. Furthermore, it would be interesting to incorporate not only volume modulation but also pitch modulation as some studies suggested that pitch can also significantly alter sound discrimination [172]. Finally,

it would be interesting to use the current techniques in social game design or teleconferencing application.

6.4 Experimental Sample Size

Before I conclude this chapter, I would like to discuss the sample size, I have used in my experiments. Specifically, in Chapter 3, I had eight participants in Experiment 1, four participants in Experiment 2, where in Chapter 4, I had 2 groups consisting of nine participants each in the first experiment and 7 for each group in the second experiment; Finally, in Chapter 5 I had 28 participants for experiment 1 and a total of 40 participants for the second experiment. The small groups of participants, of course are a limiting factor for my studies and it also outlines the difficulty of recruiting participants for human based perceptual experiments. Especially when there is an age or other restrictive inclusion criteria as in Experiment 1, Chapter 4, where the experiment required recruiting participants age 50 years or older. However, such a modest number of participants is not unusual within studies similar in design to those that I have presented in this dissertation. For example, Virzi [211] noted that 80% of the usability studies can reveal important trends with four or five participants. Furthermore, a small number of participants is very common with locomotion studies, especially for heading judgement studies. For instance,

Cornelissen et al. [36] and Crowell and Banks [38] used only 3 participants for their studies. To reiterate, I made a significant effort to use large number of participants in my studies where possible and also tried to compensate the small number of participants with large number of experimental trials and within subjects analysis.

6.5 Final Words

In summary, in this dissertation, I presented several examples of gaze-contingent multi-modal and multi-sensory displays for VE. Although, each modality or sense was studied separately, I believe that gaze information should be used for many, if not all aspects, of the virtual simulation, as it can both help to augment or simplify different processes and computations. Furthermore, GCD can help to augment both the visual and auditory senses, as people rely on multiple senses when they interact with the world around them. Hence, system developers should try to benefit from GCD for all types of outputs. Ideally, kinesthetic information should also be addressed and implemented. As I have demonstrated with visual field defect simulation (Chapter 3) and audio augmentation (Chapter 5), GCD systems can both provide missing information but also to reduce information ambiguity. Therefore, GCD system designers should address different computational tradeoffs

and use gaze to either improve cues or to reduce cue quality. This would replicate what happens in the real world, where people compensate for the lack of cues of one sense with cues from a different sense. For example, if one uses many computational resources on visual realism, it might allow one to simplify the computational resources for other senses. Similarly, it is important to use GCD to simulate both 2D as well as 3D cues. As it was addressed in Chapter 4, I speculate that to achieve realistic blur one had to incorporate blur due to accommodation but also due to different properties of the different regions of the visual field. To conclude, I believe that GCD will be more commonly used in many VR and game applications once clear benefits of such system are established and demonstrated building on the results of this dissertation.

Bibliography

- [1] *Development of a Head and Eye Slaved Display System for a Rotary Wing Research Simulator.*
- [2] ISO 9241-302: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 3: Visual display requirements. Tech. rep., 1992.
- [3] AI, Z., GUPTA, B. K., RASMUSSEN, M., LIN, Y. J., DECH, F., PANKO, W., AND SILVERSTEIN, J. C. Simulation of eye diseases in a virtual environment. In *Hawaii International Conference on System Sciences* (2000), IEEE, pp. 1–5.
- [4] AKELEY, K., WATT, S. J., GIRSHICK, A. R., AND BANKS, M. S. A stereo display prototype with multiple focal distances. In *Transactions on Graphics* (2004), vol. 23, ACM Press, pp. 804–813.
- [5] ALLISON, R. S., SCHUMACHER, J., SADR, S., AND HERPERS, R. Apparent motion during saccadic suppression periods. In *Experimental Brain Research* (2010), vol. 202, Springer, pp. 155–69.
- [6] APFELBAUM, H., PELAH, A., AND PELI, E. Heading assessment by “tunnel vision” patients and control subjects standing or walking in a virtual reality environment. In *Transactions on Applied Perception* (2007), vol. 4, ACM Press, pp. 1–16.
- [7] ATCHISON, D. A., AND WOODS, R. L. Subjective depth-of-focus of the eye. In *Optometry and Vision Science* (1997), vol. 74, LWW, pp. 511–520.
- [8] BACH, M. The freiburg visual acuity test-variability unchanged by post-hoc re-analysis. In *Graefe’s Archive for Clinical and Experimental Ophthalmology* (2007), vol. 245, pp. 965–971.
- [9] BANKS, M. S., EHRLICH, S. M., BACKUS, B. T., AND CROWELL, J. A. Estimating heading during real and simulated eye movements. In *Vision Research* (1996), vol. 36, Elsevier, pp. 431–443.

- [10] BARISIC, I., TIMMERMANS, B., VOGLEY, K., PFEIFFER, U., BENTE, G., AND SCHILBACH, L. Using dual eye tracking to investigate real time social interactions. In *CHI 2013 workshop on "Gaze Interaction in the Post-WIMP World"* (2013).
- [11] BARTH, E., DORR, M., BÖHME, M., GEGENFURTNER, K., AND MARTINETZ, T. *Guiding the mind's eye: Improving communication and vision by external control of the scanpath*, vol. 6057. 2006, pp. 116–123.
- [12] BASU, A., AND WIEBE, K. Enhancing videoconferencing using spatially varying sensing. In *International Conference on Systems, Man and Cybernetics* (1998), vol. 28, IEEE, pp. 137–148.
- [13] BATTEAU, D. W. The role of the pinna in human localization. In *Royal Society of London B: Biological Sciences* (1967), vol. 168, The Royal Society, pp. 158–180.
- [14] BELLMANN, C., FEELY, M., CROSSLAND, M. D., KABANAROU, S. A., AND R., G. S. Fixation stability using central and pericentral fixation targets in patients with age-related macular degeneration. In *Ophthalmology* (2004), vol. 111, Elsevier, pp. 2265–2270.
- [15] BENTE, G., ESCHENBURG, F., AND KRÄMER, N. C. Virtual gaze. a pilot study on the effects of computer simulated gaze in avatar-based conversations. In *Virtual Reality* (2007), vol. 4563, IEEE, pp. 185–194.
- [16] BERNHARD, M., DELL'MOUR, C., HECHER, M., STAVRAKIS, E., AND WIMMER, M. The effects of fast disparity adjustment in gaze-controlled stereoscopic applications. In *Symposium on Eye Tracking Research and Applications* (2014), ACM Press, pp. 111–118.
- [17] BEST, V., OZMERAL, E. J., AND SHINN-CUNNINGHAM, B. G. Visually-guided attention enhances target identification in a complex auditory scene. In *Journal for the Association for Research in Otolaryngology* (2007), vol. 8, Springer, pp. 294–304.
- [18] BIOCCA, F., HARMS, C., AND BURGOON, J. Toward a more robust theory and measure of social presence: Review and suggested criteria. In *Presence* (2003), vol. 12, MIT press, pp. 456–480.
- [19] BLAKE, B. *Complexity in Natural Scenes: A Viewer Centered Metric for Computing Adaptive Detail*. PhD thesis, London University (Queen Mary College), 1989.

- [20] BLAKEMORE, C., AND JULESZ, B. Stereoscopic depth aftereffect produced without monocular cues. In *Science* (1971), vol. 171, American Association for the Advancement of Science, pp. 286–288.
- [21] BLUM, T., WIECZOREK, M., AICHERT, A., TIBREWAL, R., AND NAVAB, N. The effect of out-of-focus blur on visual discomfort when using stereo displays. In *International Symposium on Mixed and Augmented Reality* (2010), IEEE, pp. 13–17.
- [22] BOER, E. R. Tangent point oriented curve negotiation. In *Intelligent Vehicles Symposium* (1996), no. 617, IEEE, pp. 7–12.
- [23] BÖHME, M., DORR, M., MARTINETZ, T., AND BARTH, E. Gaze-contingent temporal filtering of video. In *Symposium on Eye Tracking Research and Applications* (2006), ACM Press, pp. 109–115.
- [24] BOLT, R. A. Gaze-orchestrated dynamic windows. In *SIGGRAPH Computer Graphics* (1981), vol. 15, ACM, pp. 109–119.
- [25] BOSMAN, S., GROENENDAAL, B., FINDLATER, J. W., VISSER, T., DE GRAAF, M., AND MARKOPOULOS, P. Gentleguide: An exploration of haptic output for indoors pedestrian guidance. In *Human-Computer Interaction with Mobile Devices and Services* (2003), Springer, pp. 358–362.
- [26] BROOKER, J. P., AND SHARKEY, P. M. Operator performance evaluation of controlled depth of field in a stereographically displayed virtual environment. In *Stereoscopic Displays and Virtual Reality Systems VIII* (2001), A. J. Woods, M. T. Bolas, J. O. Merritt, and S. A. Benton, Eds., vol. 4297, SPIE, pp. 408–417.
- [27] BRUNGART, D. S., AND SIMPSON, B. D. Cocktail party listening in a dynamic multitalker environment. In *Perception and Psychophysics* (2007), vol. 69, Springer, pp. 79–91.
- [28] BRUNGART, D. S., SIMPSON, B. D., ERICSON, M. A., AND SCOTT, K. R. Informational and energetic masking effects in the perception of multiple simultaneous talkers. In *The Journal of the Acoustical Society of America* (2001), vol. 110, Acoustical Society of America, pp. 2527–2538.
- [29] BURGE, J., AND GEISLER, W. S. Optimal defocus estimation in individual natural images. In *Proceedings of the National Academy of Sciences* (2011), vol. 108, pp. 16849–16854.

- [30] CHANG, E. C., AND YAP, C. K. A wavelet approach to foveating images. In *Symposium on Computational Geometry* (1997), ACM Press, pp. 397–399.
- [31] CHERRY, E. C. Some experiments on the recognition of speech, with one and with two ears. In *The Journal of the Acoustical Society of America* (1953), vol. 25, Acoustical Society of America, pp. 975–979.
- [32] CHESSFOREVA. Unity3D chess project, 2010.
- [33] CLARK, J. H. Hierarchical geometric models for visible surface algorithms. In *Communications of the ACM* (1976), vol. 19, pp. 547–554.
- [34] ÇÖLTEKIN, A. *Foveation for 3D Visualization and Stereo Imaging*. PhD thesis, Helsinki University of Technology, 2006.
- [35] ÇÖLTEKIN, A., AND HAGGRÉN, H. Stereo foveation. In *The Photogrammetric Journal of Finland* (2006), vol. 20, IEEE, pp. 45–54.
- [36] CORNELISSEN, F. W., AND VAN DEN DOBBELSTEEN, J. J. Heading detection with simulated visual field defects. In *Visual Impairment Research* (1999), vol. 1, Informa UK Ltd UK, pp. 71–84.
- [37] COWAN, B., AND KAPRALOS, B. Spatial sound for video games and virtual environments utilizing real-time gpu-based convolution. In *Conference on Future Play: Research, Play, Share* (2008), ACM Press, pp. 166–172.
- [38] CROWELL, J. A., AND BANKS, M. S. Perceiving heading with different retinal regions and types of optic flow. In *Perception and Psychophysics* (1993), vol. 53, Springer, pp. 325–337.
- [39] CULLEN, B., GALPERIN, D., COLLINS, K., KAPRALOS, B., AND HOGUE, A. The effects of audio on depth perception in S3D games. In *Audio Mostly Conference: A Conference on Interaction with Sound* (2012), ACM Press, pp. 32–39.
- [40] DAGNELIE, G., AND MASSOF, R. W. Toward an artificial eye. In *Spectrum* (1996), vol. 33, IEEE, pp. 20–29.
- [41] DEAS, L., WILCOX, L. M., KAZIMI, A., AND ALLISON, R. S. Audio-visual integration in stereoscopic 3D. In *Symposium on Applied Perception* (2013), ACM Press, pp. 83–90.

- [42] DORR, M., JARODZKA, H., AND BARTH, E. Space-variant spatio-temporal filtering of video for gaze visualization and perceptual learning. In *Symposium on Eye-Tracking Research and Applications* (2010), ACM Press, pp. 307–314.
- [43] DORR, M., VIG, E., GEGENFURTNER, K. R., MARTINETZ, T., AND BARTH, E. Eye movement modelling and gaze guidance. In *International Workshop on Human-Computer Conversation* (2008).
- [44] DUCHOWSKI, A. T. Incorporating the viewer’s point of regard (por) in gaze-contingent virtual environments. In *The Engineering Reality of Virtual Reality* (San Jose, CA, USA, 1998), M. T. Bolas, S. S. Fisher, and J. O. Merritt, Eds., vol. 3295, SPIE, pp. 332–343.
- [45] DUCHOWSKI, A. T. Acuity-matching resolution degradation through wavelet coefficient scaling. In *Image Processing* (2000), vol. 9, IEEE, pp. 1437–1440.
- [46] DUCHOWSKI, A. T. A breadth-first survey of eye-tracking applications. In *Behavior Research Methods* (2002), vol. 34, Springer, pp. 455–470.
- [47] DUCHOWSKI, A. T. *Eye Tracking Methodology*. Springer Science & Business Media, 2007.
- [48] DUCHOWSKI, A. T., AND ÇÖLTEKIN, A. Foveated gaze-contingent displays for peripheral lod management, 3D visualization, and stereo imaging. In *Multimedia Computing, Communications, and Applications* (2007), vol. 3, ACM, pp. 1–18.
- [49] DUCHOWSKI, A. T., AND EADDY, T. D. A gaze-contingent display compensating for scotomata. In *Eurographics* (2009), pp. 9–12.
- [50] DUCHOWSKI, A. T., HOUSE, D. H., GESTRING, J., CONGDON, R., ŚWIRSKI, L., DODGSON, N. A., KREJTZ, K., AND KREJTZ, I. Comparing estimated gaze depth in virtual and physical environments. In *Symposium on Eye Tracking Research and Applications* (2014), ACM Press, pp. 103–110.
- [51] DUCHOWSKI, A. T., HOUSE, D. H., GESTRING, J., WANG, R. I., KREJTZ, K., KREJTZ, I., MANTIUK, R., AND BAZYLUK, B. Reducing visual discomfort of 3D stereoscopic displays with gaze-contingent depth-of-field. In *Symposium on Applied Perception* (2014), ACM Press, pp. 39–46.
- [52] DUCHOWSKI, A. T., AND MCCORMICK, B. H. Gaze-contingent video resolution degradation. In *Human Vision and Electronic Imaging III* (1998), vol. 3299, SPIE, pp. 318–329.

- [53] EADDY, M., BLASKO, G., BABCOCK, J., AND FEINER, S. My own private kiosk: Privacy-preserving public displays. In *Wearable Computers* (2004), vol. 1, IEEE, pp. 132–135.
- [54] EGAN, J. P., CARTERETTE, E. C., AND THWING, E. J. Some factors affecting multi channel listening. In *The Journal of the Acoustical Society of America* (1954), vol. 26, Acoustical Society of America, pp. 774–782.
- [55] EIZENMAN, M., AND GUESTIN, E. D. Covert monitoring of the point-of-gaze. In *Science and Technology for Humanity* (2009), IEEE, pp. 551–556.
- [56] ELMER-DEWITT, P. Technology: Into the wild blue (digital) yonder, August 1988. [Online; posted 01-August-1988].
- [57] FORTENBAUGH, F., HICKS, J., HAO, L., AND TURANO, K. High-speed navigators: Using more than what meets the eye. In *Journal of Vision* (2006), vol. 6, Association for Research in Vision and Ophthalmology, pp. 565–579.
- [58] FORTENBAUGH, F. C., HICKS, J. C., HAO, L., AND TURANO, K. A. Losing sight of the bigger picture: Peripheral field loss compresses representations of space. In *Vision Research* (2007), vol. 47, Elsevier, pp. 2506–2520.
- [59] FORTENBAUGH, F. C., HICKS, J. C., HAO, L., AND TURANO, K. A. A technique for simulating visual field losses in virtual environments to study human navigation. In *Behavior Research Methods* (2007), vol. 39, Springer, pp. 552–560.
- [60] FOULSHAM, T., AND KINGSTONE, A. Modelling the influence of central and peripheral information on saccade biases in gaze-contingent scene viewing. In *Visual Cognition* (2012), vol. 20, Taylor & Francis, pp. 546–579.
- [61] FRATER, M. R., ARNOLD, J. F., AND VAHEDIAN, A. Impact of audio on subjective assessment of video quality in videoconferencing applications. In *Circuits and Systems for Video Technology* (2001), vol. 11, pp. 1059–1062.
- [62] FUKUSHIMA, T., TORII, M., AND UKAI, K. The relationship between ca/c ratio and individual differences in dynamic accommodative responses while viewing stereoscopic images. In *Journal of Vision* (2009), vol. 9, Association for Research in Vision and Ophthalmology, pp. 1–13.
- [63] GEISLER, W. S., AND PERRY, J. S. A real-time foveated multiresolution system for low-bandwidth video communication. In *Human Vision and Electronic Imaging III* (1998), vol. 3299, International Society for Optics and Photonics, pp. 294–305.

- [64] GEISLER, W. S., AND PERRY, J. S. Real-time simulation of arbitrary visual fields. In *Symposium on Eye Tracking Research and Applications* (2002), ACM Press, pp. 57–69.
- [65] GEISLER, W. S., AND PERRY, J. S. Real-time simulation of arbitrary visual fields. In *Symposium on Eye Tracking Research and Applications* (2002), ACM Press, pp. 83–87.
- [66] GIBSON, J. J. Visually controlled locomotion and visual orientation in animals. In *British Journal of Psychology* (1958), vol. 49, pp. 182–194.
- [67] GROSSMANN, P. Depth from focus. In *Pattern Recognition Letters* (1987), vol. 5, Elsevier, pp. 63–69.
- [68] GUENTER, B., FINCH, M., DRUCKER, S., DESNEY, T., AND SNYDER, J. Foveated 3D graphics. In *Transactions on Graphics* (2012), vol. 31, ACM Press, pp. 164:1—164:10.
- [69] GULLSTRAND, A. The optical system of the eye. In *Helmholtz’s Treatise on Physiological Optics* (1910), H. von Helmholtz, Ed., vol. 1, J. P. C. Southall, pp. 350–358.
- [70] HALLE, M. Autostereoscopic displays and computer graphics. In *SIGGRAPH Computer Graphics* (1997), vol. 31, ACM Press, pp. 58–62.
- [71] HAYASHI, K., HAYASHI, H., NAKAO, F., AND HAYASHI, F. Aging changes in apparent accommodation in eyes with a monofocal intraocular lens. In *American Journal of Ophthalmology* (2003), vol. 135, Elsevier, pp. 432–436.
- [72] HELD, R. T., COOPER, E. A., AND BANKS, M. S. Blur and disparity are complementary cues to depth. In *Current Biology* (2012), vol. 22, Elsevier, pp. 426–431.
- [73] HILLAIRES, S., LÉCUYER, A., COZOT, R., AND CASIEZ, G. Depth-of-field blur effects for first-person navigation in virtual environments. In *Computer Graphics and Applications* (2008), vol. 28, IEEE, pp. 47–55.
- [74] HILLAIRES, S., LÉCUYER, A., COZOT, R., AND CASIEZ, G. Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments. In *Virtual Reality* (2008), IEEE, pp. 47–50.
- [75] HOFFMAN, D. M., AND BANKS, M. S. Focus information is used to interpret binocular images. In *Journal of Vision* (2010), vol. 10, Association for Research in Vision and Ophthalmology, pp. 1–17.

- [76] HOFFMAN, D. M., GIRSHICK, A. R., AKELEY, K., AND BANKS, M. S. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. In *Journal of Vision* (2008), vol. 8, Association for Research in Vision and Ophthalmology, pp. 1–30.
- [77] HOPPE, H. View-dependent refinement of progressive meshes. In *Conference on Computer Graphics and Interactive Techniques* (1996), ACM Press, pp. 99–108.
- [78] HOWARD, I. P. *Seeing in Depth, Vol. 1: Basic Mechanisms*. University of Toronto Press, 2002.
- [79] HOWARD, I. P., AND TEMPLETON, W. B. *Human Spatial Orientation*. John Wiley & Sons, 1966.
- [80] HYRSKYKARI, A., MAJARANTA, P., AND RÄIHÄ, K. J. From gaze control to attentive interfaces. In *International Conference on Human-Computer Interaction* (2005), vol. 2.
- [81] INOUE, T., AND OHZU, H. Accommodative responses to stereoscopic three-dimensional display. In *Applied Optics* (1997), vol. 36, Optical Society of America, pp. 4509–4515.
- [82] JACOB, R. J. The use of eye movements in human-computer interaction techniques: What you look at is what you get. In *Information Systems* (1991), vol. 9, ACM Press, pp. 152–169.
- [83] JACOB, R. J. K. What you look at is what you get: Eye movement-based interaction techniques. In *Human Factors in Computing Systems: Empowering People* (1990), ACM Press, pp. 11–18.
- [84] JARODZKA, H. *In the Eye of an Expert: Conveying Perceptual Skills in Biological and Medical Domains via Eye Movement Modeling Examples*. PhD thesis, Universität Tübingen, 2010.
- [85] JONES, A., MCDOWALL, I., YAMADA, H., BOLAS, M., AND DEBEVEC, P. Rendering for an interactive 360° light field display. In *Transactions on Graphics* (2007), vol. 26, ACM Press, pp. 40–1–40–2.
- [86] KELLY, D. H. Spatial frequency selectivity in the retina. In *Vision Research* (1975), vol. 15, Elsevier, pp. 665–672.

- [87] KIM, D., CHOI, S., PARK, S., AND SOHN, K. Stereoscopic visual fatigue measurement based on fusional response curve and eye-blinks. In *Digital Signal Processing* (2011), IEEE, pp. 1–6.
- [88] KISTLER, F., WISSNER, M., AND ANDRÉ, E. Level of detail based behavior control for virtual characters. In *Intelligent Virtual Agents* (2010), Springer-Verlag, pp. 118–124.
- [89] KOLASINSKI, E. M. Simulator sickness in virtual environments. Tech. rep., U.S Army Research Institute for the Behaviour and Social Sciences, Alexandria, 1995.
- [90] KOOI, F. L., AND TOET, A. Visual comfort of binocular and 3D displays. In *Displays* (2004), vol. 25, Elsevier, pp. 99–108.
- [91] KORETZ, J. F., COOK, C. A., AND KAUFMAN, P. L. Aging of the human lens: Changes in lens shape upon accommodation and with accommodative loss. In *Journal of the Optical Society of America* (2002), vol. 19, pp. 144–151.
- [92] KORETZ, J. F., KAUFMAN, P. L., NEIDER, M. W., AND GOECKNER, P. A. Accommodation and presbyopia in the human eye - aging of the anterior segment. In *Vision Research* (1989), vol. 29, Elsevier, pp. 1685–1692.
- [93] KORTUM, P., AND GEISLER, W. S. Implementation of a foveated image coding system for image bandwidth reduction. In *Human Vision and Electronic Imaging* (1996), vol. 2657, SPIE, pp. 350–360.
- [94] KUNKA, B., AND KOSTEK, B. Exploiting audio-visual correlation by means of gaze tracking. In *International Journal of Computer Science and Applications* (2010), vol. 7, pp. 104–123.
- [95] KUNKA, B., KOSTEK, B., KULESZA, M., SZCZUKO, P., AND CZYZEWSKI, A. Gaze-tracking-based audio-visual correlation analysis employing quality of experience methodology. In *Intelligent Decision Technologies* (2010), vol. 4, IOS Press, pp. 217–227.
- [96] KUZE, J., AND UKAI, K. Subjective evaluation of visual fatigue caused by motion images. In *Displays* (2008), vol. 29, Elsevier, pp. 159–166.
- [97] LAMBOOIJ, M. Measuring visual discomfort associated with 3D displays. In *Stereoscopic Displays and Applications XX* (2009), vol. 7237, SPIE, pp. 72370K–1–72370K–12.

- [98] LAMBOOIJ, M., FORTUIN, M., HEYNDERICKX, I., AND IJSSELSTEIJN, W. Visual discomfort and visual fatigue of stereoscopic displays: A review. In *Journal of Imaging Science and Technology* (2009), vol. 53, pp. 030201–1–030201–14.
- [99] LANCE, B., AND MARSELLA, S. Glances, glares, and glowering: How should a virtual human express emotion through gaze? In *Autonomous Agents and Multi-Agent Systems* (2010), vol. 20, Springer, pp. 50–69.
- [100] LAND, M. Focusing by shape change in the lens of the eye: A commentary on Young (1801) 'On the mechanism of the eye'. In *Philosophical Transactions of the Royal Society of London B: Biological Sciences* (2015), vol. 370, The Royal Society, pp. 1–6.
- [101] LAND, M. F., AND LEE, D. N. Where we look when we steer. In *Nature* (1994), vol. 369, Nature Publishing Group, pp. 742–744.
- [102] LANGER, M. S., AND SICILIANO, R. A. Are blur and disparity complementary cues to depth? In *Vision Research* (2014), vol. 107, Elsevier, pp. 15–21.
- [103] LAPPI, O. Future path and tangent point models in the visual control of locomotion in curve driving. In *Journal of Vision* (2014), vol. 12, Association for Research in Vision and Ophthalmology, pp. 1–22.
- [104] LEROY, L., F. P., AND MOREAU, G. Real-time adaptive blur for reducing eye strain in stereoscopic displays. In *Transactions on Applied Perception* (2012), vol. 9, ACM Press, pp. 1–18.
- [105] LEWALD, J. The effect of gaze eccentricity on perceived sound direction and its relation to visual localization. In *Hearing Research* (1998), vol. 115, Elsevier, pp. 206–216.
- [106] LEWALD, J., AND EHRENSTEIN, W. H. The effect of eye position on auditory lateralization. In *Experimental Brain Research* (1996), vol. 108, Springer, pp. 473–485.
- [107] LI, L., P. E., AND WARREN, W. H. Heading perception in patients with advanced retinitis pigmentosa. In *Optometry and Vision Science* (2002), vol. 79, LWW, pp. 581–589.
- [108] LINDSTROM, P., KOLLER, D., RIBARSKY, W., HODGES, L. F., FAUST, N., AND TURNER, G. A. Real-time, continuous level of detail rendering of height fields. In *Conference on Computer Graphics and Interactive Techniques* (1996), ACM Press, pp. 109–118.

- [109] LINTERN, G., AND LIU, Y. T. Explicit and implicit horizons for simulated landing approaches. In *Human Factors and Ergonomics Society* (1991), vol. 33, SAGE Publications, pp. 401–417.
- [110] LIU, S., AND HUA, H. Spatialchromatic foveation for gaze contingent displays. In *Symposium on Eye Tracking Research and Applications* (2008), ACM Press, pp. 139–142.
- [111] LONG, N., AND OVER, R. Stereoscopic depth aftereffects with random-dot patterns. In *Vision Research* (1973), vol. 13, Elsevier, pp. 1283–1287.
- [112] LOSING, V., PFEIFFER, T., ROTTKAMP, L., AND ZEUNERT, M. Guiding visual search tasks using gaze-contingent auditory feedback. In *International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (2014), ACM Press, pp. 1093–1102.
- [113] LUEBKE, D. P. *Level of Detail for 3D Graphics*. Morgan Kaufmann, 2003.
- [114] MADDOX, R. K., P. D. A. S. G. C., AND LEE, A. K. Directing eye gaze enhances auditory spatial cue discrimination. In *Current Biology* (2014), vol. 24, Elsevier, pp. 748–752.
- [115] MANTIUK, R., B. B., AND TOMASZEWSKA, A. Gaze-dependent depth-of-field effect rendering in virtual environments. In *Serious Games Development and Applications* (2011), M. Ma, F. O. M., and M. P. J., Eds., vol. 6944 of *Lecture Notes in Computer Science*, Springer, pp. 1–12.
- [116] MARMITT, G., AND DUCHOWSKI, A. T. Modeling visual attention in VR: Measuring the accuracy of predicted scanpaths. In *Eurographics* (2002), pp. 217–226.
- [117] MARSHALL, J. A., BURBECK, C. A., ARIELY, D., ROLLAND, J. P., AND MARTIN, K. E. Occlusion edge blur: A cue to relative visual depth. In *Journal of the Optical Society of America* (1996), vol. A, pp. 681–688.
- [118] MARTIN-GONZALEZ, A., LANZL, I., KHORAMNIA, R., AND NAVAB, N. Simulation and modeling of metamorphopsia with a deformable amsler grid. In *Medicine Meets Virtual Reality* (2011), vol. 18, pp. 336–343.
- [119] MASTOROPOULOU, G., DEBATTISTA, K., CHALMERS, A., AND TROSCIANKO, T. The influence of sound effects on the perceived smoothness of rendered animations. In *Symposium on Applied Perception in Graphics and Visualization* (2005), ACM Press, pp. 9–15.

- [120] MATHER, G. The use of image blur as a depth cue. In *Perception* (1997), vol. 26, Pion Ltd, pp. 1147–1158.
- [121] MATHER, G., AND SMITH, D. Depth cue integration: Stereopsis and image blur. In *Vision Research* (2000), vol. 40, Elsevier, pp. 3501–3506.
- [122] MAUDERER, M., CONTE, S., AND NACENTA, M. Utility of gaze-contingent dof blur for depth perception. In *ITS 2013 Workshop on Visual Aadaptation of Interfaces* (2013).
- [123] MAUDERER, M., CONTE, S., NACENTA, M. A., AND DHANRAJ, V. Depth perception with gaze-contingent depth of field. In *Conference on Human Factors in Computing Systems* (2014), ACM Press, pp. 217–226.
- [124] MAURIN, H., SONNENWALD, D. H., CAIRNS, B., MANNING, J. E., FREID, E. B., AND FUCHS, H. Exploring gender differences in perceptions of 3d telepresence collaboration technology: An example from emergency medical care. In *Nordic Conference on Human-Computer Interaction* (New York, New York, USA, 2006), ACM Press, pp. 381–384.
- [125] MCCAULEY, M. E., AND SHARKEY, T. J. Cybersickness: Perception of self-motion in virtual environments. In *Presence: Teleoperators and Virtual Environments* (1992), vol. 1, MIT Press, pp. 311–318.
- [126] MCCONKIE, G. W., AND CURRIE, C. B. Visual stability across saccades while viewing complex pictures. In *Journal of Experimental Psychology. Human Perception and Performance* (1996), vol. 22, American Psychological Association, pp. 563–81.
- [127] MCGURK, H., AND MACDONALD, J. Hearing lips and seeing voices. In *Nature* (1976), vol. 264, Nature Publishing Group, pp. 746–748.
- [128] MILLER, G. A. The masking of speech. In *Psychological Bulletin* (1947), vol. 44, American Psychological Association, pp. 105–129.
- [129] MITCHELL, D. E. Retinal disparity and diplopia. In *Vision Research* (1966), vol. 6, Elsevier, pp. 441–451.
- [130] MODEL, D., AND EIZENMAN, M. User-calibration-free remote gaze estimation system. In *Symposium on Eye-Tracking Research and Applications* (2010), ACM Press, pp. 29–36.

- [131] MON-WILLIAMS, M., AND WANN, J. P. Binocular virtual reality displays: When problems do and don't occur. In *Human Factors: The Journal of the Human Factors and Ergonomics Society* (1998), vol. 40, SAGE Publications, pp. 42–49.
- [132] MON-WILLIAMS, M., WANN, J. P., AND RUSHTON, S. Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display. In *Ophthalmic and Physiological Optics* (1993), vol. 13, Wiley Online Library, pp. 387–391.
- [133] MORDI, J. A., AND CIUFFREDA, K. J. Static aspects of accommodation: Age and presbyopia. In *Vision Research* (1998), vol. 38, Elsevier, pp. 1643–1653.
- [134] MURPHY, H., AND DUCHOWSKI, A. T. Hybrid image/model based gaze-contingent rendering. In *Symposium on Applied Perception in Graphics and Visualization* (2007), ACM Press, pp. 107–114.
- [135] MURPHY, H. A., DUCHOWSKI, A. T., AND TYRRELL, R. A. Hybrid image/model-based gaze-contingent rendering. In *Transactions on Applied Perception* (2009), vol. 5, ACM Press, pp. 1–21.
- [136] NEIL, M. A., PAIGE, E. G., AND SUCHAROV, L. O. Spatial-light-modulator-based three-dimensional multiplanar display. In *Stereoscopic Displays and Virtual Reality Systems IV* (San Jose, CA, USA, 1997), SPIE, pp. 337–341.
- [137] NGUYEN, V. A., HOWARD, I. P., AND ALLISON, R. S. Detection of the depth order of defocused images. In *Vision Research* (2005), vol. 45, Elsevier, pp. 1003–1011.
- [138] NIKOLOV, S. G., BULL, D. R., CANAGARAJAH, C. N., JONES, M. G., AND GILCHRIST, I. D. Multi-modality gaze-contingent displays for image fusion. In *Conference on Information Fusion* (2002), vol. 2, IEEE, pp. 1213–1220.
- [139] NIKOLOV, S. G., NEWMAN, T. D., BULL, D. R., CANAGARAJAH, N. C., JONES, M. G., AND GILCHRIST, I. D. Gaze-contingent display using texture mapping and opengl: System and applications. In *Symposium on Eye Tracking Research and Applications* (2004), ACM Press, pp. 11–18.

- [140] NORTH, M. M., NORTH, S. M., AND COBLE, J. Virtual reality therapy: An innovative paradigm. In *Encyclopedia of Psychotherapy* (1996), T. Pyle and A. Wilson, Eds., vol. 2, pp. 889–893.
- [141] O’HARE, L., AND HIBBARD, P. B. Visual discomfort and blur. In *Journal of Vision* (2013), vol. 13, Association for Research in Vision and Ophthalmology, pp. 1–12.
- [142] O’HARE, L., ZHANG, T., NEFS, H. T., AND HIBBARD, P. B. Visual discomfort and depth-of-field. In *i-Perception* (2013), vol. 4, Pion Publications, pp. 156–169.
- [143] OHSHIMA, T., YAMAMOTO, H., AND TAMUR, H. Gaze-directed adaptive rendering for interacting with virtual space. In *Virtual Reality Annual International Symposium* (1996), vol. 96, IEEE, pp. 103–110.
- [144] OKADA, Y., UKAI, K., WOLFFSOHN, J. S., GILMARTIN, B., IJIMA, A., AND BANDO, T. Target spatial frequency determines the response to conflicting defocus-and convergence-driven accommodative stimuli. In *Vision Research* (2006), vol. 46, Elsevier, pp. 475–484.
- [145] O’SULLIVAN, C., AND DINGLIANA, J. Collisions and perception. In *Transactions on Graphics* (2001), vol. 20, ACM Press, pp. 151–168.
- [146] O’SULLIVAN, C., DINGLIANA, J., AND HOWLETT, S. Eye-movements and interactive graphics. In *The Mind’s Eyes: Cognitive and Applied Aspects of Eye Movement Research* (2003), Elsevier, pp. 555–571.
- [147] PALMISANO, S., AND GILLAM, B. Visual perception of touchdown point during simulated landing. In *Journal of Experimental Psychology: Applied* (2005), vol. 11, American Psychological Association, pp. 19–32.
- [148] PAPAGEORGIOU, E., HARDIESS, G., MALLOT, H. A., AND SCHIEFER, U. Gaze patterns predicting successful collision avoidance in patients with homonymous visual field defects. In *Vision Research* (2012), vol. 65, Elsevier, pp. 25–37.
- [149] PARKHURST, D., AND CULURCIELLO, E. Evaluating variable resolution displays with visual search: Task performance and eye movements. In *Symposium on Eye Tracking Research and Applications* (2000), ACM Press, pp. 105–109.

- [150] PARKHURST, D. J., AND NIEBUR, E. Variable-resolution displays: A theoretical, practical, and behavioral evaluation. In *Human Factors: The Journal of the Human Factors and Ergonomics Society* (2002), vol. 44, SAGE Publications, pp. 611–629.
- [151] PAUSCH, R., CREA, T., AND CONWAY, M. A literature survey for virtual environments: Military flight simulator visual systems and simulator sickness. In *Presence: Teleoperators and Virtual Environments* (1992), vol. 1, MIT Press, pp. 344–363.
- [152] PELECHANO, N., STOCKER, C., ALLBECK, J., AND BADLER, N. Being a part of the crowd: Towards validating VR crowds using presence. In *International Joint Conference on Autonomous Agents and Multiagent Systems* (2008), vol. 1, International Foundation for Autonomous Agents and Multiagent Systems, pp. 136–142.
- [153] PELI, E. The visual effects of head-mounted display (hmd) are not distinguishable from those of desk-top computer display. In *Vision Research* (1998), vol. 38, Elsevier, pp. 2053–2066.
- [154] PENTLAND, A. P. A new sense for depth of field. In *Pattern Analysis and Machine Intelligence* (1987), no. 4, IEEE, pp. 523–531.
- [155] RAMSDALE, C., AND CHARMAN, W. N. A longitudinal study of the changes in the static accommodation response. In *Ophthalmic and Physiological Optics* (1989), vol. 9, Wiley Online Library, pp. 255–263.
- [156] RAYNER, K. Eye movements in reading and information processing: 20 years of research. In *Psychological Bulletin* (1998), vol. 124, American Psychological Association, pp. 372–422.
- [157] REGAN, D., AND BEVERLEY, K. I. Visual fields for frontal plane motion and for changing size. In *Vision Research* (1983), vol. 23, Elsevier, pp. 673–676.
- [158] REINGOLD, E. M., AND LOSCHKY, L. C. Saliency of peripheral targets in gaze-contingent multiresolutional displays. In *Behavior Research Methods, Instruments, and Computers* (2002), vol. 34, Springer, pp. 491–499.
- [159] RIMELL, A. N., MANSFIELD, N. J., AND HANDS, D. The influence of content, task and sensory interaction on multimedia quality perception. In *Ergonomics* (2008), vol. 51, Taylor & Francis, pp. 85–97.

- [160] RIMELL, A. N., AND OWEN, A. The effect of focused attention on audio-visual quality perception with applications in multi-modal codec design. In *Conference on Acoustics, Speech, and Signal Processing* (2000), vol. 4, IEEE, pp. 2377–2380.
- [161] ROGERS, S. D., KADAR, E. E., AND COSTALL, A. Drivers’ gaze patterns in braking from three different approaches to a crash barrier. In *Ecological Psychology* (2005), vol. 17, Taylor & Francis, pp. 39–53.
- [162] ROGERS, S. D., KADAR, E. E., AND COSTALL, A. Gaze patterns in the visual control of straight-road driving and braking as a function of speed and expertise. In *Ecological Psychology* (2005), vol. 17, Taylor & Francis, pp. 19–38.
- [163] ROYDEN, C. S., BANKS, M. S., AND CROWELL, J. A. The perception of heading during eye movements. In *Nature* (1992), vol. 360, Nature Publishing Group, pp. 583–585.
- [164] ROYDEN, C. S., CROWELL, J. A., AND BANKS, M. S. Estimating heading during eye movements. In *Vision Research* (1994), vol. 34, Elsevier, pp. 3197–3214.
- [165] RUSHTON, S. K., AND ALLISON, R. S. Biologically-inspired heuristics for human-like walking trajectories toward targets and around obstacles. In *Displays* (2013), vol. 34, Elsevier, pp. 105–113.
- [166] RUSHTON, S. K., AND HARRIS, J. M. The utility of not changing direction and the visual guidance of locomotion. In *Optic Flow and Beyond* (2004), L. M. Vaina, S. A. Beardsley, and S. K. Rushton, Eds., vol. 324 of *Synthese Library*, Springer, pp. 363–381.
- [167] RUSHTON, S. K., HARRIS, J. M., LLOYD, M. R., AND WANN, J. P. Guidance of locomotion on foot uses perceived target location rather than optic flow. In *Current Biology* (1998), vol. 8, Elsevier, pp. 1191–4.
- [168] SALVUCCI, D. D., AND GRAY, R. A two-point visual control model of steering. In *Perception* (2004), vol. 33, Pion Ltd, pp. 1233–1248.
- [169] SAXBY, G. *Practical holography*. CRC Press, 2003.
- [170] SCHOR, C. M. The relationship between fusional vergence eye movements and fixation disparity. In *Vision Research* (1979), vol. 19, Elsevier, pp. 1359–1367.

- [171] SCHUMACHER, J., ALLISON, R. S., AND HERPERS, R. Using saccadic suppression to hide graphic updates. In *Eurographics Symposium on Virtual Environments* (2004), pp. 17–24.
- [172] SHAMMA, S. A., ELHILALI, M., AND MICHEYL, C. Temporal coherence and attention in auditory scene analysis. In *Trends in Neurosciences* (2011), vol. 34, Elsevier, pp. 114–23.
- [173] SHEEDY, J., AND BERGSTROM, N. Performance and comfort on near-eye computer displays. In *Optometry and Vision Science* (2002), vol. 79, LWW, pp. 306–312.
- [174] SHEEDY, J. E., HAYES, J. N., AND ENGLE, J. Is all asthenopia the same? In *Optometry and Vision Science* (2003), vol. 80, LWW, pp. 732–739.
- [175] SHIWA, S., OMURA, K., AND KISHINO, F. Proposal for a 3-d display with accommodative compensation: 3DDAC. In *Journal of the Society for Information Display* (1996), vol. 4, pp. 255–261.
- [176] SPEAR, P. D. Neural bases of visual deficits during aging. In *Vision Research* (1993), vol. 33, Elsevier, pp. 2589–2609.
- [177] STARKER, I., AND BOLT, R. A. A gaze-responsive self-disclosing display. In *SIGCHI Conference on Human Factors in Computing Systems* (1990), ACM Press, pp. 3–10.
- [178] STELMACH, L. B., AND TAM, W. J. Processing image sequences based on eye movements. In *International Symposium on Electronic Imaging: Science and Technology* (1994), vol. 2179, International Society for Optics and Photonics, SPIE, pp. 90–98.
- [179] STERN, J. A., BOYER, D., AND SCHROEDER, D. Blink rate: A possible measure of fatigue. In *The Journal of the Human Factors and Ergonomics Society* (1994), vol. 36, Sage Publications, pp. 284–297.
- [180] STEVENS, S. S. The measurement of loudness. In *The Journal of the Acoustical Society of America* (1955), vol. 27, Acoustical Society of America, pp. 815–829.
- [181] STRUTT, J. W. On our perception of sound direction. In *Philosophical Magazine* (1907), vol. 13, Taylor & Francis, pp. 214–232.
- [182] SUCHAROV, L. *An investigation of new methods of creating three-dimensional multiplanar displays*. University of Oxford, 1998.

- [183] SUCHAROV, L. O. D., PAIGE, E. G. S., AND NEIL, M. A. A. Three-dimensional image display, 1999. CA Patent App. CA 2,300,047.
- [184] SUYAMA, S., TAKADA, H., UEHIRA, K., SAKAI, S., AND OHTSUKA, S. Novel direct-vision 3-d display using luminance-modulated two 2-d images displayed at different depths. In *SID Symposium Digest of Technical Papers* (2000), vol. 31, Wiley Online Library, pp. 1208–1211.
- [185] TAKAKI, Y., TANAKA, K., AND NAKAMURA, J. Super multi-view display with a lower resolution flat-panel display. In *Optics Express* (2011), vol. 19, OSA, pp. 4129–4139.
- [186] TANRIVERDI, V., AND JACOB, R. J. K. Interacting with eye movements in virtual environments. In *Human Factors in Computing Systems* (2000), ACM Press, pp. 265–272.
- [187] TARITA-NISTOR, L., GONZALEZ, E. G., MARKOWITZ, S. N., AND STEINBACH, M. J. Fixation characteristics of patients with macular degeneration recorded with the mp-1 microperimeter. In *Retina* (2008), vol. 28, LWW, pp. 125–133.
- [188] TARITA-NISTOR, L., GONZÁLEZ, E. G., MARKOWITZ, S. N., AND STEINBACH, M. J. Plasticity of fixation in patients with central vision loss. In *Visual Neuroscience* (2010), vol. 26, Cambridge Univ Press, pp. 1–8.
- [189] THIEBAUX, M., LANCE, B., AND MARSELLA, S. Real-time expressive gaze animation for virtual humans. In *International Conference on Autonomous Agents and Multiagent Systems* (2009), International Foundation for Autonomous Agents and Multiagent Systems, pp. 321–328.
- [190] THOMPSON, W., WILLEMSSEN, P., GOOCH, A., CREEM-REGEHR, S., LOOMIS, J., AND BEALL, A. Does the quality of the computer graphics matter when judging distances in visually immersive environments? In *Presence: Teleoperators and Virtual Environments* (2004), vol. 13, MIT Press, pp. 560–571.
- [191] TLAPALE, E., BERNARD, J. B., CASTET, E., AND KORNPORST, P. The solaire project: A gaze-contingent system to facilitate reading for patients with scotomas. Tech. rep., ODYSSEE - INRIA Sophia Antipolis, 2006.
- [192] TOET, A., VAN DER HOEVEN, M., KAHRIMANOVIĆ, M., AND DELLEMAN, N. J. Effects of field of view on human locomotion. In *Defense and Secu-*

- ity Symposium (Orlando, FL, USA, 2008), vol. 6955, SPIE, pp. 69550H1–69550H11.
- [193] TOYAMA, T., KIENINGER, T., SHAFAIT, F., AND DENGEL, A. Gaze guided object recognition using a head-mounted eye tracker. In *Symposium on Eye Tracking Research and Applications* (2012), ACM Press, pp. 91–98.
 - [194] TSIRLIN, I., WILCOX, L. M., AND ALLISON, R. S. The effect of crosstalk on the perceived depth from disparity and monocular occlusions. In *Broadcasting* (2011), vol. 57, pp. 445–453.
 - [195] TSUMURA, N., ENDO, C., HANEISHI, H., AND MIYAKE, Y. Image compression and decompression based on gazing area. In *Human Vision and Electronic Imaging* (1996), vol. 2657, SPIE, pp. 361–367.
 - [196] TURANO, K. A., YU, D., HAO, L., AND HICKS, J. C. Optic-flow and egocentric-direction strategies in walking: Central vs peripheral visual field. In *Vision Research* (2005), vol. 45, Elsevier, pp. 3117–3132.
 - [197] TURNER, H., AND FIRTH, D. Generalized nonlinear models in r: an overview of the gnm package. (r package version 1.0-6). Tech. rep., 2007.
 - [198] TURNER, H., AND FIRTH, D. Generalized nonlinear models in r: An overview of the gnm package. Tech. rep., 2012.
 - [199] TWARDON, L., KOESLING, H., FINKE, A., AND RITTER, H. Gaze-contingent audio-visual substitution for the blind and visually impaired. In *International Conference on Pervasive Computing Technologies for Healthcare* (2013), Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, pp. 129–136.
 - [200] UKAI, K., AND HOWARTH, P. A. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. In *Displays* (2008), vol. 29, Elsevier, pp. 106–116.
 - [201] UKAI, K., AND KATO, Y. The use of video refraction to measure the dynamic properties of the near triad in observers of a 3-d display. In *Ophthalmic and Physiological Optics* (2002), vol. 22, Wiley Online Library, pp. 385–388.
 - [202] VAN DER LINDE, I. Multiresolution image compression using image foveation and simulated depth of field for stereoscopic displays. In *Electronic Imaging* (San Jose, CA, USA, 2004), A. J. Woods, J. O. Merritt, S. A. Benton, and M. T. Bolas, Eds., vol. 5291, SPIE, pp. 71–80.

- [203] VAN ERP, A. N., JAN, B., VELTMAN, J. A., VAN VEEN, H. A., AND OIVING, A. B. Tactile torso display as countermeasure to reduce night vision goggles induced drift. In *RTO HFM Symposium on Spatial Disorientation in Military Vehicles: Causes, Consequences and Cures* (2003), vol. 49, pp. 49–149–8.
- [204] VAN ERP, J. B., AND VAN VEEN, H. A. H. C. Vibro-tactile information presentation in automobiles. In *Eurohaptics* (2001), vol. 2001, pp. 99–104.
- [205] VILLARRUEL, C. *Computer Graphics and Human Depth Perception with Gaze-Contingent Depth of Field*. PhD thesis, Mount Holyoke College, 2006.
- [206] VINNIKOV, M. *Gaze-contingent Real-time Simulation of Impaired Vision in Naturalistic Displays and Effects of Visual Impairments on Everyday Tasks*. Master’s thesis, York University, 2009.
- [207] VINNIKOV, M., AND ALLISON, R. S. Contingency evaluation of gaze-contingent displays for real-time visual field simulations. In *Symposium on Eye-Tracking Research and Applications* (Austin, Texas, 2010), ACM Press, pp. 263–266.
- [208] VINNIKOV, M., AND ALLISON, R. S. Gaze-contingent simulations of visual defects in virtual environment: Challenges and limitations. In *CHI 2013 workshop on "Gaze Interaction in the Post-WIMP World"* (2013).
- [209] VINNIKOV, M., AND ALLISON, R. S. Gaze-contingent depth of field in realistic scenes: The user experience. In *Symposium on Eye Tracking Research and Applications* (2014), ACM Press, pp. 119–126.
- [210] VINNIKOV, M., ALLISON, R. S., AND SWIERAD, D. Real-time simulation of visual defects with gaze-contingent display. In *Symposium on Eye Tracking Research and Applications* (Savannah, Georgia, 2008), ACM Press, pp. 127–130.
- [211] VIRZI, R. A. Refining the test phase of usability evaluation: How many subjects is enough? In *Human Factors and Ergonomics Society* (1992), vol. 34, SAGE Publications, pp. 457–468.
- [212] VÕ, M. L. H., SMITH, T. J., MITAL, P. K., AND HENDERSON, J. M. Do the eyes really have it? dynamic allocation of attention when viewing moving faces. In *Journal of Vision* (2012), vol. 12, Association for Research in Vision and Ophthalmology, pp. 1–14.

- [213] WANG, B., AND CIUFFREDA, K. J. Depth-of-focus of the human eye: Theory and clinical implications. In *Survey of Ophthalmology* (2006), vol. 51, Elsevier, pp. 75–85.
- [214] WANN, J., AND LAND, M. Steering with or without the flow: Is the retrieval of heading necessary? In *Trends in Cognitive Sciences* (2000), vol. 4, Elsevier Current Trends, pp. 319–324.
- [215] WANN, J. P., RUSHTON, S., AND MON-WILLIAMS, M. Natural problems for stereoscopic depth perception in virtual environments. In *Vision Research* (1995), vol. 35, Elsevier, pp. 2731–2736.
- [216] WARREN, W. H. Perception of heading is a brain in the neck. In *Nature Neuroscience* (1998), vol. 1, Nature America, pp. 647–649.
- [217] WARREN, W. H., AND HANNON, D. J. Eye movements and optical flow. In *Journal of the Optical Society of America* (1990), vol. 7, pp. 160–9.
- [218] WARREN, W. H., AND KURTZ, K. J. The role of central and peripheral vision in perceiving the direction of self-motion. In *Perception and Psychophysics* (1992), vol. 51, Springer, pp. 443–54.
- [219] WARREN, W. H., MORRIS, M. W., AND KALISH, M. Perception of translational heading from optical flow. In *Journal of Experimental Psychology. Human Perception and Performance* (1988), vol. 14, American Psychological Association, pp. 646–60.
- [220] WATSON, B., WALKER, N., AND HODGES, L. F. A user study evaluating level of detail degradation in the periphery of head-mounted displays. In *Interactive Virtual Environments Conference* (1995), pp. 18–19.
- [221] WATSON, B. W., WALKER, N., HODGES, L., AND WORDEN, A. Managing level of detail through peripheral degradation: Effects on search performance with a head-mounted display. In *Computer-Human Interaction* (1997), vol. 4, ACM Press, pp. 323–346.
- [222] WATT, S. J., AKELEY, K., ERNST, M. O., AND BANKS, M. S. Focus cues affect perceived depth. In *Journal of Vision* (2005), vol. 5, Association for Research in Vision and Ophthalmology, pp. 834–862.
- [223] WETZSTEIN, G., LANMAN, D., HEIDRICH, W., AND RASKAR, R. Layered 3D. In *Transactions on Graphics* (2011), vol. 30, ACM Press, pp. 95:1–95:12.

- [224] WILKIE, R. M., WANN, J. P., AND ALLISON, R. S. Active gaze, visual look-ahead and locomotor control. In *Journal of Experimental Psychology: Human Perception and Performance* (2008), vol. 34, American Psychological Association, pp. 1150–1164.
- [225] WILMS, M., SCHILBACH, L., PFEIFFER, U., BENTE, G., FINK, G. R., AND VOGLEY, K. It’s in your eyes—using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. In *Social Cognitive and Affective Neuroscience* (2010), vol. 5, Oxford University Press, pp. 98–107.
- [226] WINKLER, S., AND FALLER, C. Audiovisual quality evaluation of low-bitrate video. In *Human Vision and Electronic Imaging X* (2005), B. E. Rogowitz, T. N. Pappas, and S. J. Daly, Eds., vol. 5666, SPIE, pp. 139–148.
- [227] WINN, B., WHITAKER, D., ELLIOTT, D. B., AND PHILLIPS, N. J. Factors affecting light-adapted pupil size in normal human subjects. In *Investigative Ophthalmology and Visual Science* (1994), vol. 35, The Association for Research in Vision and Ophthalmology, pp. 1132–1137.
- [228] WITKIN, H. A., AND LEVENTHAL, T. Sound localization with conflicting visual and auditory cues. In *Journal of Experimental Psychology* (1952), vol. 43, American Psychological Association, pp. 58–67.
- [229] YANG, R., HUANG, X., LI, S., AND JAYNES, C. Toward the light field display: Autostereoscopic rendering via a cluster of projectors. In *Visualization and Computer Graphics* (2008), vol. 14, IEEE, pp. 84–96.
- [230] YANG, S., AND SHEEDY, J. E. Effects of vergence and accommodative responses on viewer’s comfort in viewing 3D stimuli. In *Stereoscopic Displays and Applications XXII* (2011), International Society for Optics and Photonics, pp. 78630Q–78630Q.
- [231] YANO, S., EMOTO, M., AND MITSUHASHI, T. Two factors in visual fatigue caused by stereoscopic hdtv images. In *Displays* (2004), vol. 25, Elsevier, pp. 141–150.
- [232] YOUNG, T. The bakerian lecture : On the mechanism of the eye. In *Philosophical Transactions of the Royal Society of London* (1801), vol. 91, The Royal Society, pp. 23–88.

- [233] ZHANG, T., O'HARE, L., HIBBARD, P. B., NEFS, H. T., AND HEYNDER-
ICKX, I. Depth of field affects perceived depth in stereographs. In *Transac-
tions on Applied Perception* (2014), vol. 11, ACM Press, pp. 1–18.
- [234] ZONE, R. *Stereoscopic cinema and the origins of 3-D film*. University Press
of Kentucky, 2007.