FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Impact of Virtual Environment on User Attention in VR

Pedro Alexandre Ferreira e Silva

Master's Degree in Informatics and Computing Engineering

Supervisor: Prof. Rui Pedro Amaral Rodrigues

Second Supervisor: Prof. Teresa Carla de Canha e Matos

July 18, 2025

Impact of Virtual Environment on User Attention in VR

Pedro Alexandre Ferreira e Silva

Master’s Degree in Informatics and Computing Engineering

Approved in oral examination by the committee:

President: Prof. Alexandre Miguel Barbosa Valle de Carvalho
Referee: Prof. Bernardo José Santos Marques

July 18, 2025

Contents

1 Introduction **3**
1.1 Motivation . . . 3
1.2 Problem . . . 4
1.3 Research questions . . . 5
1.4 Objectives . . . 5
1.5 Document Structure . . . 5

2 State Of The Art **6**
2.1 Virtual Reality . . . 6
2.1.1 Immersion and Presence . . . 6
2.1.2 The Reality-Virtuality Continuum and Virtual Environments . . . 7
2.2 Visual Attention . . . 9
2.2.1 Visual Attention Components . . . 10
2.2.2 Visual Saliency: Properties and Impact on Attention . . . 11
2.2.3 Visual Attention Analysis . . . 12
2.2.4 Discussion . . . 13
2.3 Summary . . . 14

3 Approach for Assessing Lighting Impact on Attention **16**
3.1 Requirements . . . 16
3.2 Conceptual Model . . . 17
3.3 Architecture . . . 17
3.4 Proof-of-concept and experimental conditions . . . 19
3.5 Configuration and Metric Logging . . . 21
3.6 Integration of the questionnaire . . . 22
3.7 Experiment flow . . . 23
3.8 Eye Tracking Integration . . . 24
3.9 Summary . . . 25

4 Evaluation and Results **28**
4.1 Methodology . . . 28
4.1.1 Experimental Setup . . . 28
4.1.2 Design . . . 30
4.1.3 Tasks . . . 30
4.1.4 Order . . . 31
4.2 Gathered Data . . . 32
4.2.1 Demographics . . . 32
4.2.2 Metrics . . . 32

4.2.3 Follow-up Questionnaires . . . 33
4.3 Results . . . 33
4.3.1 Gaze / Fixation Metrics and Task Completion Times . . . 34
4.3.2 IPQ . . . 36
4.3.3 Comparison Questionnaire . . . 38
4.4 Discussion . . . 38

5 Conclusions **40**
5.1 Limitations and Future Work . . . 41

References **42**

A Literature Review **46**
A.1 Methodology . . . 46
A.2 Search Strings & Collection . . . 46
A.3 Screening and Selection . . . 48
A.4 Reference Chaining . . . 49
A.5 Conclusion . . . 49

B Literature Review - Search Queries **50**
B.1 Detailed Search Queries . . . 50

C Demographics and Comparison Questionnaires **53**
C.1 Demographics Questionnaire . . . 54
C.2 Comparison Questionnaire . . . 57

Resumo

Em contextos de realidade virtual (RV), especialmente em RV imersiva, é crucial dirigir a atenção do utilizador de uma forma não intrusiva para melhorar a eficiência de uma tarefa sem sacrificar os níveis de satisfação e presença do utilizador. Os trabalhos anteriores centraram-se principalmente na análise do efeito da saliência na atenção do utilizador em situações de visualização livre, deixando uma lacuna no que diz respeito à investigação do seu impacto juntamente com cenários guiados pela influência de uma tarefa. Neste estudo, tentamos atrair a atenção visual através da iluminação em duas tarefas diferentes, a visualização livre e a pesquisa visual, seguindo um design intra-sujeitos em que cada utilizador completa as duas tarefas em ambas as condições de iluminação. O seu impacto é avaliado através de métricas relevantes, tais como o tempo até à primeira fixação e os tempos de conclusão de tarefa, bem como a garantia de que a diferença nas condições de iluminação não afecta negativamente a presença do utilizador, utilizando questionários padrão da indústria como o Igroup Presence Questionnaire (IPQ). Após cada conjunto de tarefas, recolhemos também as respostas do IPQ e um questionário de comparação no final de cada experiência. Pudemos concluir que a presença de iluminação adicional não afetou significativamente a atenção do utilizador na visualização livre, nem contribuiu para uma melhoria significativa nos tempos de conclusão da tarefa de pesquisa visual, embora tenha alterado para melhor o realismo experienciado pelos utilizadores, e as respostas do questionário de comparação mostram que os participantes sentiram que a iluminação adicional afetou a sua exploração das cenas e ajudou-os a encontrar o objeto alvo. Este trabalho apresenta perspetivas importantes sobre a utilização da iluminação nos Ambientes Virtuais Imersivos para atrair a atenção do utilizador e abre oportunidades para investigação futura nesta área com outras propriedades que afectam a saliência de objectos.

Abstract

In Virtual Reality (VR) contexts, especially immersive VR, driving user attention in a non-intrusive manner is crucial to improve task efficiency without sacrificing user satisfaction and presence levels. Previous works focused mainly on analyzing the effect of saliency on user attention in free-viewing situations, leaving a gap regarding the investigation of its impact in conjunction with task-oriented scenarios. In this work, we attempt to attract visual attention through lighting in two different tasks, free-viewing and visual search, following a within-subjects design where 24 users completed both tasks in both lighting conditions, with eye-tracking integration and fixation analysis. Its impact is measured using relevant metrics such as time to first fixation and task completion times, as well as ensuring the difference in lighting conditions does not affect user presence negatively by employing industry standard questionnaires like the Igroup Presence Questionnaire (IPQ). After each set of tasks, we also collect the responses from the IPQ and a comparison questionnaire at the end of each experiment. We were able to conclude that the presence of additional lighting did not meaningfully affect user attention in free-viewing, nor did it contribute to a significant improvement in visual task completion times. Although, it did change the users' experienced realism for the better, and answers from the comparison questionnaire show that participants felt the added lighting affected their exploration of the scenes and helped them find the target object. This work presents important insights on the use of lighting inside Immersive Virtual Environments to attract user attention, and enables opportunities for future research in this area with other saliency-affecting properties.

Acknowledgements

I would like to express my gratitude to my supervisors, Prof. Teresa Matos and Prof. Rui Rodrigues, whose guidance was essential to the development of this dissertation. They never left me in the dark and always displayed immense interest in following my work's progress: all I could have asked for and more.

To my friends and family, who kept me motivated and were constantly present in the good and bad moments during my work, some of which were also busy with their dissertations. It was my pleasure to share this journey with you.

And last, but certainly not least, to my partner Iris, who despite being several countries apart managed to make me feel so close, with her invaluable support and unconditional love. Seni seviyorum, aşkım!

Pedro Silva

"You have a good rest of your life, kid."

Jesse Pinkman, *Breaking Bad*

List of Abbreviations

AOI	Area of Interest
AR	Augmented Reality
AV	Augmented Virtuality
CAVE	Cave Automatic Virtual Environment
CSV	Comma Separated Values
FOV	Field of View
GUI	Graphical User Interface
HMD	Head-mounted Display
HMM	Hidden Markov Model
IPQ	Igroup Presence Questionnaire
IVE	Immersive Virtual Environment
JSON	JavaScript Object Notation
MR	Mixed Reality
NSS	Normalized Scanpath Saliency
PMR	Pure Mixed Reality
VE	Virtual Environment
VR	Virtual Reality

List of Figures

1.1 A screenshot of the yellow paint in "Final Fantasy VII Rebirth". 4

2.1 A diagram of the reality-virtuality continuum, adapted from Flavian et al. [11] . . 8
2.2 A Cave Automatic Virtual Environment illustration, as seen in Beji et al. [4] . . . 9
2.3 Visual representation of foveal/peripheral vision and eye movements [40] 11
2.4 Kernel Density Estimate graphs of fixation coordinates in the three tasks. 14

3.1 Diagram of the conceptual model 18
3.2 Diagram of the architecture 19
3.3 Park scene in development 20
3.4 Early version of the park scene 20
3.5 Target object used during visual search tasks 21
3.6 Questionnaire page displayed with VRQuestionnaireToolkit 23
3.7 Session page inside Cognitive3D 24
3.8 Example of a task replay inside Cognitive3D 25
3.9 Final version of Park scene 26
3.10 Final version of Room scene 26
3.11 Final version of Startup scene (similar to Questionnaire scene) 27

4.1 The Vive XR Elite headset with the Full Face Tracker accessory. 29
4.2 Diagram of the experimental flow. 29
4.3 In-game description of the free-viewing task 30
4.4 In-game description of the visual search task 31
4.5 Participant performing the experiment 34

C.1 Page 1 of the demographics questionnaire 54
C.2 Page 2 of the demographics questionnaire 55
C.3 Page 3 of the demographics questionnaire 56
C.4 Questions from the comparison questionnaire 57

List of Tables

4.1 Experimental versions based on conditions 32
4.2 Visualization and performance metrics by lighting condition 35
4.3 Average time to first fixation per object under both lighting conditions 35
4.4 Time-to-first fixation under both lighting conditions in the park scene 35
4.5 Time-to-first fixation under both lighting conditions in the room scene 35
4.6 Task completion time under both lighting conditions 36
4.7 Presence score under both lighting conditions 36
4.8 Spatial presence score under both lighting conditions 37
4.9 Involvement score under both lighting conditions 37
4.10 Experienced realism score under both lighting conditions 37
4.11 Total score under both lighting conditions 37
4.12 Results from question 2 of the comparison questionnaire 38
4.13 Results from question 3 of the comparison questionnaire 38

A.1 Search Query Results 47

B.1 Detailed Search Queries for Each Database - SS1 50
B.2 Detailed Search Queries for Each Database - SS2 51
B.3 Detailed Search Queries for Each Database - SS3 51
B.4 Detailed Search Queries for Each Database - SS4 52

AI Usage Disclosure

Artificial Intelligence (hereafter AI) tools have been utilized in this dissertation. ChatGPT has been used to assist with the writing of this dissertation.
All AI-generated content was reviewed and validated for relevance, appropriateness, and accuracy before incorporation into the final manuscript to maintain scholarly integrity of this research.
The author takes full responsibility for the content of the document.

Sustainable Development Goals

The Sustainable Development Goal most aligned with the focus of this thesis is **Goal 9 – Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation**, particularly **Target 9.5**, which emphasizes enhancing scientific research and upgrading technological capabilities across industrial sectors. This thesis contributes to this objective by providing information on the effective use of immersive virtual reality (VR) properties, presenting important insights regarding the attraction of user attention and the improvement of task completion times, with applications in user-centered training, simulation, and human-computer interaction. Through the development of a VR study that integrates eye tracking and fixation analysis, the work supports innovation in both research methodologies and applied immersive systems.

To assess its alignment with Target 9.5, several performance indicators are considered. These include the applicability of the system in domains such as training or usability studies, and the engagement with public or private R&D through potential partnerships or prototype deployments. Moreover, the documentation and possible contribution to open-source tools or datasets can help promote broader innovation accessibility. Standard United Nations indicators such as R&D expenditure as a proportion of GDP and researchers per million inhabitants also contextualize the relevance of immersive VR research in national and international development goals, particularly in developing digital innovation and supporting a greater technological amplitude.

Chapter 1

Introduction

Immersive Virtual Environments (IVEs) are an application of Virtual Reality (VR) and motion tracking to create realistic environments, where users can interact with simulated objects and recreate real-life scenarios without many of the constraints it entails. The military, healthcare and entertainment sectors are a few examples of the many domains that benefit from this technology's flexibility, along with its increased immersion and presence [6]. People can learn and prepare for real-life environments inside IVEs [37], as well as enjoy themselves, more effectively than in regular 2D and traditional 3D environments such as desktop VR [26]. The previously mentioned concepts of immersion and presence are closely related: immersion enables the sensory depth of the experience, while presence reflects the user's psychological sense of being within the virtual environment [41]. Directing users toward relevant elements of the experience is essential to achieving the intended outcomes, therefore, designing IVEs that properly manage visual attention is paramount.

1.1 Motivation

To channel user attention in IVEs, we can explore the visual characteristics of elements within the environment, such as an object's appearance or behavior. Changing object properties like color, shape, texture and movement can make the objects stand out from their surroundings. These properties, as well as the idea of an object "standing out", are all related to a concept called saliency: a characteristic of an object which sets them apart from the rest of the environment, affecting user attention [33]. The effect that each of these properties have on user attention is a topic that has limited research in VR [33], unlike 2D environment alternatives. Thus, a more thorough analysis would greatly benefit applications of IVEs in the aforementioned areas. VR equipment prices have been in a downward trend over the past few years, and as this technology becomes more affordable and commonplace, the benefit of research around capturing user attention in IVEs rises.

1.2 Problem

Nowadays, developers of IVEs face a troubling dilemma. On one hand, they can preserve the realism of their solution, refusing to add immersion-breaking elements to their environments to guide users and thus increasing the chance of frustration with the lack of direction or unnecessary time spent exploring every detail to deduce the intended path. On the other hand, they can add visual instructions and exaggerated effects to relevant objects, such as unnatural glows or sudden modifications in scale and lighting, conveying their intent in an explicit albeit unsatisfying way. The problem of immersion versus user experience has been debated for a long time [1], even outside the context of VR. A well-known example is the discussion surrounding the video game "Final Fantasy VII Rebirth" and its use of uncanny yellow paint to highlight climbable rocky ledges, as seen in figure 1.1. However, with enough research, the possibility of a compromise can be achieved, where immersion is preserved and user guidance is done through manipulating existing object properties that still allow them to blend into the environment well enough.

Figure 1.1: A screenshot of the yellow paint in "Final Fantasy VII Rebirth".

Current studies [3] on the topic prove that VR improves focus during assignments compared to 2D scenarios and changes in saliency do affect user attention, more specifically between tasks. What's more, the importance of head movement inside IVEs in focusing on relevant objects, that is, moving one's head instead of the eyes to position objects of interest in their field of view, is highlighted [13]. The issue with the existing research is that the manipulated variables are insufficient to fully understand what properties can be leveraged to obtain a more definitive change in attention, with the crucial aspect of head movement in mind. As such, a broader exploration of different characteristics of objects and virtual environments themselves is needed to extract insights about how to properly create engaging and effective IVEs, balancing immersion and clarity.

1.3 Research questions

Considering the previous proposition and shortcomings, the main objectives of this dissertation are to discover which object properties are most successful at attracting user attention, as well as which combination of modifications translates into higher task success rates while keeping user immersion and satisfaction high, requiring thorough user testing and validation. In this thesis, given the constraints on time and resources, we will focus on the use of lighting in an attempt to accomplish the aforementioned goals. These can be represented by the following research questions:

- **RQ1:** Is lighting effective in drawing user attention without the influence of a task?
- **RQ2:** Does the inclusion of lighting around a target object affect the performance of its visual search without adversely affecting the level of presence?

1.4 Objectives

A user-centered study is required to properly evaluate the effect of lighting on user attention and attempt to answer the research questions. Participants should be faced with different lighting conditions (presence and absence of additional lighting), so that we can observe their behavior in both situations and register relevant metrics. This can be achieved by developing a VR user testing system using an adequate tool (e.g. Unity), with a defined flow of varied tasks performed with and without added lighting, integrating questionnaires to measure presence in both conditions. Collecting metrics such as time-to-first fixation, through the analysis of data obtained through an eye-tracker, and task completion times is essential for a meaningful analysis after the experiments.

1.5 Document Structure

Chapter 2 covers foundational concepts in virtual reality and visual attention based on relevant work, essential to the understanding of this study. Additionally, it presents the state of the art regarding attention studies and analysis methods. Details on the collection procedure of papers reviewed can be verified on Appendix A. Chapter 3 describes the implementation of the developed system used for the experimental study, from the high-level approach, requirements and conceptual model to the practical information about the development of the experimental platform. A thorough explanation of the methodology used for the experimental phase, along with an analysis of the results that originated from it, is provided in chapter 4. Finally, chapter 5 summarizes the study in its entirety and the conclusions we can extract from it, exposing limitations and suggesting future work avenues.

Chapter 2

State Of The Art

In order to better contextualize the area, a set of relevant concepts are presented in the first part of this chapter. These are then complemented with an analysis of the state of the art research found on the most relevant topics. To perform this analysis, a systematic review of scientific papers was conducted, following a mixed methods review, and whose details can be found in Appendix A. Section A.1 introduces the method being used and provides the reasoning behind its choice over other options. Section A.2 focuses on the logic behind the creation of the search queries, as well as the results obtained by querying each database. Section A.3 sheds light into the process of choosing which papers are most relevant for the task at hand and why, providing a set of factors that contribute to that decision. Finally, section A.4 refers to the search and collection of more relevant papers through the investigation of the bibliography of each previously selected paper, complementing the group of papers obtained in the aforementioned steps. In total, 33 papers were found after literature review. The objective was to gather insight of the field and identify gaps that can be addressed in this work.

2.1 Virtual Reality

VR has reshaped the way users engage with digital environments. By extending interaction beyond the flat screen, VR enables users to experience simulated spaces with a heightened sense of physical presence. Through the use of Head-Mounted Displays (HMDs) and spatial interaction, users can observe and navigate virtual worlds as though they are physically placed within them. It is a rapidly evolving field with limited research regarding more recent developments, namely how user attention can be shaped in IVEs. It is essential to comprehend key concepts in this area, such as immersion, presence and IVEs, before diving into the topic of user attention.

2.1.1 Immersion and Presence

Immersion and presence are two foundational aspects of VR experiences, both of which play critical roles in improving user experience. Understanding their meaning, what affects them and what they control is paramount.

Immersion can be defined as "*the involvement of a user in a virtual environment during which his or her awareness of time and the real world often becomes disconnected, thus providing a sense of "being" in the task environment instead*" [29]. Within IVEs, it is affected by several factors, such as isolation from the physical world, which can be achieved through noise cancelation, for example, the user's perception regarding its own inclusion and movement, and using natural motions for interaction [41]. Immersion can also facilitate the expression of emotion in a user: studies show higher emotional responses in environments with greater immersion (as a property of the system), compared to less immersive virtual reality setups [22].

Presence is described as "*the subjective experience of being in one place or environment, even when one is physically situated in another*" [41]. It is influenced by multiple factors, such as the degree of immersion - where higher immersion leads to a stronger sense of presence - update rates and field of view [41], the latter of which will be explored in a different section. Additionally, diegetic elements, which belong to the virtual narrative and will be the utilized type of element going forward, enhance the sense of presence, while non-diegetic elements, integrated outside the narrative framework, diminish it [39]. Presence also varies across different dimensions, the physical world and psychological world, with virtual environments affecting the mental dimension: the more focused a user is on its content, the higher their sense of presence [41]. Virtual environments are a key concept of this thesis, and the following subsection will explore them in depth alongside user engagement.

2.1.2 The Reality-Virtuality Continuum and Virtual Environments

Virtual environments, and more specifically IVEs, will be the type of area where the tests conducted by users will take place. Therefore, a thorough analysis of the reality-virtuality continuum, as well as a deeper analysis of virtual environments and some properties related to user engagement in them, is indispensable.

Before diving into virtual environments, we must establish what can be described as a "*reality-virtuality continuum*" [11]. This spectrum measures the balance between real-life aspects and virtual elements in an environment. On one end, we have the **Real Environment (RE)**, which represents reality itself, perceived directly or through an indirect method such as video. On the other end, A **Virtual Environment (VE)** can be described as a "*computer-generated environment in which objects that do not exist are displayed on a device and where users can interact in real time*" [11]. Between these two extremes is **Mixed Reality (MR)**, further subdivided into three different classifications of experiences, in increasing order of virtuality: **Augmented Reality (AR)**, **Pure Mixed Reality (PMR)** and **Augmented Virtuality (AV)**. Instead of isolating the user from the real world, AR complements it with virtual components and provides a more natural sense of presence, due to the user being able to fully perceive the real world [20]. AV opts for the opposite: a virtual environment as a basis for the experience and the incorporation of real-life elements into it. PMR serves as a middle ground between AR and AV, maintaining the real world as the primary environment while allowing dynamic interactions between virtual and physical objects. In the context of this thesis, the main focus is on IVEs, and as such, further exploration will be related to VEs.

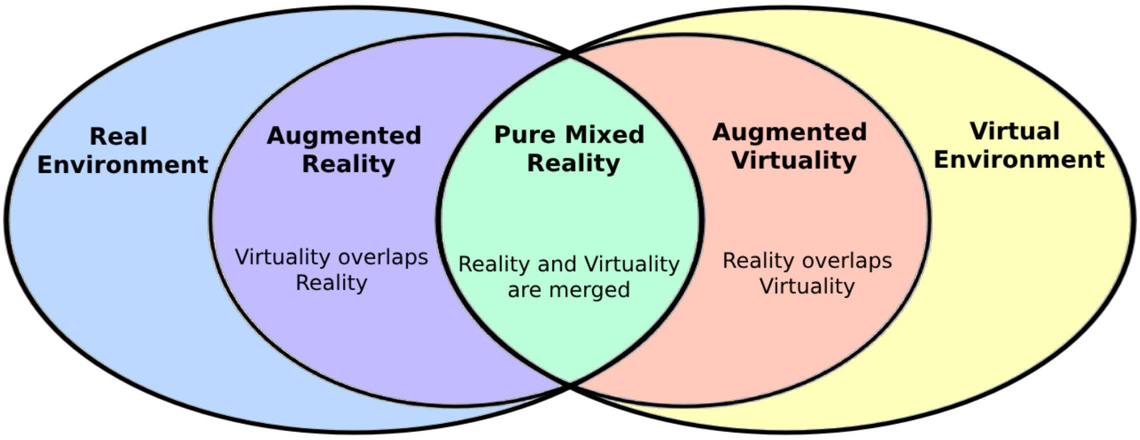


Figure 2.1: A diagram of the reality-virtuality continuum, adapted from Flavian et al. [11]

In the context of VR, IVEs, facilitated by HMDs, provide fully immersive experiences. These environments significantly enhance immersion and presence by fully engaging the user's visual and auditory senses. When combined with advanced interactivity, such as real-time sensory feedback, IVEs create the sensation of physical presence within the virtual world [20]. A less portable alternative to HMDs, which will not be used in the context of this thesis, are Cave Automatic Virtual Environments (CAVEs), a technology that projects stereoscopic 3D images onto room surfaces [7]. While CAVEs also allow users to interact with the VE, they differ from HMDs in the sense that they do not completely isolate users from the physical world [7]. Figure 2.2 shows a labeled illustration of a CAVE.

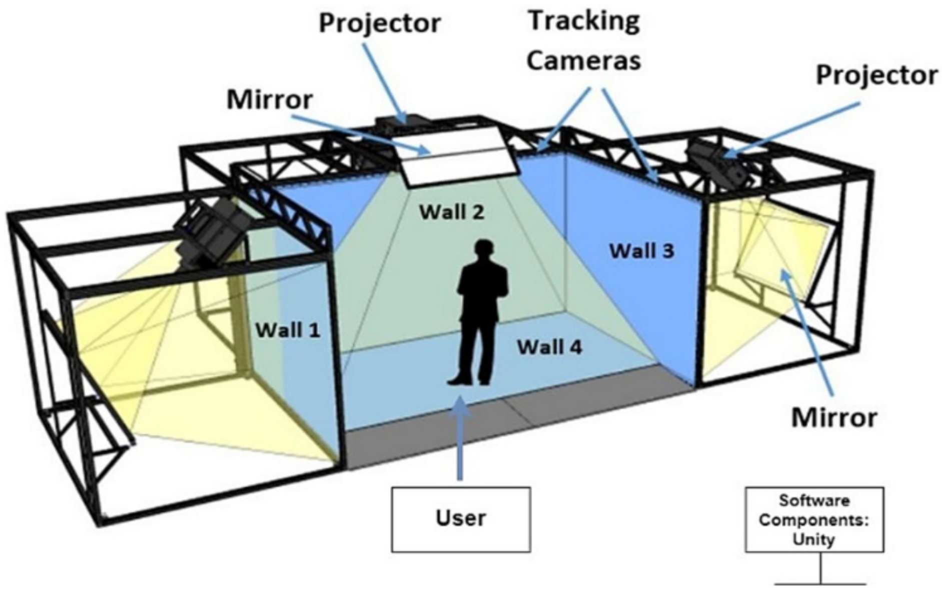


Figure 2.2: A Cave Automatic Virtual Environment illustration, as seen in Beji et al. [4]

Interactivity in virtual environments dictates the level of agency a user has in changing the behavior of a given scenario. Various types of interaction mechanisms in IVEs, such as graphical user interfaces (GUI), gestures, and voice commands, have been studied in relation to factors like development complexity and functional integration (interaction with objects) [24]. User interaction is one of the factors that can impact **task performance**: how the user behaved during the execution of the task. It is typically measured using metrics such as completion time, accuracy, error rates, or gaze and fixation data, depending on the specific type of task [33] [39]. In addition to influencing task performance, user interaction also plays a role in **spatial awareness** within IVEs.

Spatial awareness, an essential aspect of user experience in virtual environments, refers to a user's ability to relate their movements to objects in the environment, such as maintaining appropriate distances to avoid collisions. Research has shown that IVEs enhance spatial awareness compared to traditional desktop-based VR setups, offering users a more intuitive understanding of their surroundings [2]. This formation of the environment's mental map is affected by user attention: by focusing on landmarks or other salient features, users can better understand the layout of the virtual space, ultimately improving their orientation and navigation capabilities [8]. Visual attention is a component of user attention that will be heavily studied in this work, requiring a deeper analysis in the following section.

2.2 Visual Attention

Visual attention is the procedure where "*humans select a portion of the available visual information for localization, identification, and understanding of objects in an environment*" [5], and it is commonplace in real-life situations, such as the shift of our gaze to direct attention and take decisions [8]. Vision itself can be divided into foveal and peripheral vision, representing distinct areas inside the visual field, and both obtain different types of information.

2.2.1 Visual Attention Components

The main distinctions between **foveal** and **peripheral** vision are clear, as they refer to different parts of the visual field. The former pertains to vision inside the center of the visual field, the fovea, where the image is high resolution, and the latter to the region outside it, the periphery, which is wider than foveal vision, although lower resolution [40]. Although they do not have similarities in their characteristics, these two areas of vision interact with each other. For example, one common interpretation on how the visual field is so homogeneous despite the stark differences between both processing types is that foveal information such as texture or brightness can be extrapolated to the periphery. This extrapolation can be compared to the "filling-in of gaps in the visual field" such as blind spots [40]. Foveal vision is relevant for the concept of Field Of View (FOV), which represents the viewable area that a user or an optical device can observe. Studies show that while a higher FOV during VR training does not necessarily correlate with better real-world performance of that task, it contributes to better task performance in that training, which is the relevant scenario for this work [30]. Both types of vision are represented in figure 2.3.

Along with the different areas of vision, **eye movements** can also be separated by distinct types, which accomplish separate goals for the user and can be identified as fixations, saccades or blinks. A **fixation** refers to the pauses of the human gaze over regions of interest, with an arbitrary minimum amount of time (for example, 150 ms) [17], where users can process visual information. Conversely, a **saccade** is the transition between fixations, a quick eye movement that represents the aforementioned changes in visual attention [17]. To analyze eye position data with respect to salient positions, we must be able to distinguish between those caused by fixations and saccades, as an object cannot be considered potentially salient by mere eye positioning: it must have multiple fixations centered on it. Moreover, interpreting raw eye position data can give us basic information regarding where users usually look at, but only through separating fixations and saccades can we truly discover, at a high-level, why they gaze at those different locations and what the differences mean [36]. **Blinking** is another useful eye movement, as tracking these can allow dynamic changes to the scene, for example, moving an object that freezes whenever the user is looking at it, that will not be perceived by the user - a phenomenon called **change blindness** [42]. Lasting approximately between 100 to 200 milliseconds, blinking can be voluntary when used in a social interaction, or involuntary, as every human does every few seconds or due to an external factor [42].

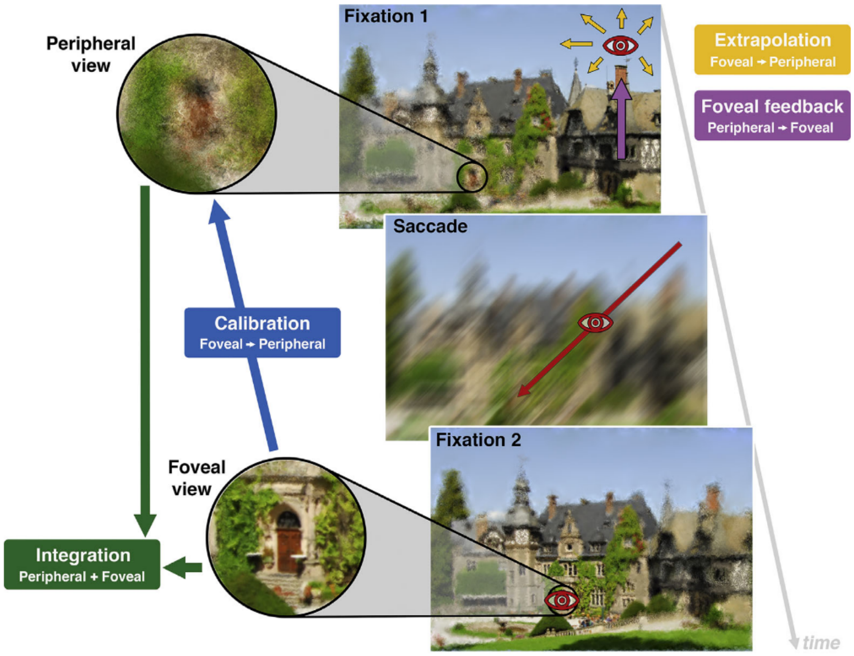


Figure 2.3: Visual representation of foveal/peripheral vision and eye movements [40]

Taken from Stewart et al. [40]. The diagram illustrates differences between foveal and peripheral vision, fixations and saccades, and how these interact

Along with the two vision regions and different types of eye movements, there are also two categories that can be attributed to visual attention. **Overt** attention refers to the physical adjustment of the gaze to a given stimulus and is the most common human behavior in scenarios similar to real life. This, along with the fact that overt attention is easily recorded with the help of eye trackers, is the reason why most recent models of visual attention target this variant [31]. On the other hand, **covert** attention comes before eye movement and can target many objects during fixation: a subconscious change in attention that precedes physical action [31]. The main motivation behind these attention changes is also a relevant topic that will be explored in the next subsection.

2.2.2 Visual Saliency: Properties and Impact on Attention

Visual saliency plays a major role in controlling attention [38] and is the main focus of this thesis. By manipulating object properties to modify saliency, the goal is to discover how these changes correlate to attention allocation and task outcomes in IVEs.

Visual saliency is a characteristic of objects that makes them stand out from others around them, drawing visual attention [23]. Saliency can be affected in many ways, namely by modifying basic object properties such as color, shape, and texture; but also object movement and behavior. Some of these are evidenced by studies regarding object-based saliency [25] [27] and others require more extensive research. This thesis proposes to expand on this research through user testing to discover if lighting can impact visual attention effectively.

In VR environments, the impact of saliency in user attention has already been studied, but mostly in free-viewing conditions, where the user is not assigned a specific task [14] [19] [18]. Even in the case of task-oriented investigations, they are not completely focused on the effects of saliency over attention [12] or lack depth when it comes to the properties of objects relevant to saliency that are being manipulated [33]. As such, a study focused on evaluating which object properties related to saliency affect visual attention in task-oriented scenarios and how these change task completion rates is yet to be accomplished. Knowing how to capture and analyze visual attention is paramount to performing such studies, a topic which will be discussed in the next and final subsection.

2.2.3 Visual Attention Analysis

In order to draw conclusions with respect to the aforementioned correlation between saliency, visual attention, and task performance, proper methods to register and analyze attention are required. This last subsection will delve into two main attention models that have been previously introduced, top-down and bottom-up, alongside possible tools and methods to collect fixation data and interpret it.

Bottom-up attention considers basic object properties such as color, shape, and texture to predict visual attention [16], but most visual attention is thought to be top-down and is motivated by high-level features such as assigned goals in a given environment or scene context, thus being associated with task-oriented scenarios [34] [15].

Several works have been published presenting different models of bottom-up attention, based on global contrast alone (the saliency of a region in comparison to the entire image) [16] or a combination of factors such as color and orientation [21], but not specifically in IVEs. On the other hand, research on top-down models and scenarios is not as extensive as on their counterparts, since they are more complex than bottom-up and free-viewing contexts [15]. Using task-dependent influences (such as the task/scene context) has been explored by some researchers [9] [28]. Attention model analysis does not have to be limited to those applied in IVEs, though, as a study shows that gaze data and saliency for VR scenarios are sufficiently similar to those in regular 2D displays. Moreover, models suited for these conventional modes can be adapted to function in VR environments, by following several steps such as 2D projection of the spherical panorama (patch-based projection was deemed the most accurate but computationally intensive) and taking equator bias into account (users focusing more on the middle of the panorama by default) [38]. Given this information, it might be possible to study 2D bottom-up saliency models and use the features they prioritize as a starting point when crafting the environments for VR studies, to figure out which object properties are most impactful on user attention.

As one would assume, an overwhelming amount of articles utilize eye-tracking data to analyze user visual attention [15] [9] [28] [17]. Despite the immediate instinct of capturing eye fixation data to measure visual attention, some studies argue that head movement is sufficient to properly interpret attention [13] [38], reducing experiment costs, but it is preferable to perform a more in-depth analysis. Before, the importance of separating fixation from saccade eye positions was highlighted, and there are various methods to accomplish this distinction [31] [35]:

- **Velocity-based methods**: Classify eye movements through point-to-point velocity analysis. Fixations are characterized by low velocities and saccades by higher velocities. Although simple, it is a method prone to be affected by noise;
- **Hidden Markov Models (HMM)-based methods**: Use HMMs to identify transitions between fixations and saccades based on probability. More robust against noise compared to velocity-based methods, as they rely on modeled state transitions and velocity changes;
- **Dispersion-based methods**: Identify fixations by measuring the spatial dispersion of gaze points within an arbitrary time-frame. Fixations show small dispersions, while saccades display larger dispersions. Of course, the performance of these methods is shaped by the chosen dispersion thresholds, and a poor decision can be extremely prejudicial to the results obtained [36];
- **Area of interest (AOI)-based methods**: Classify fixations by mapping gaze points onto predefined AOIs, and label everything outside of them as saccade points. Fixation points are then collapsed into groups represented as a single fixation. A duration threshold is used to distinguish true fixations from saccades going through those areas. These methods capture **dwell times** (the time a user spends looking inside an AOI) and can be used to deduce more information.

In regards to prediction models, this data can then be interpreted, for example, using normalized scanpath saliency (NSS) - a metric used to compare model-predicted positions with actual eye positions, represented by the following formula:

$$\text{NSS} = \frac{1}{\sigma_S}\left(S(x_{\text{human}}, y_{\text{human}}) - \mu_S\right)$$

where μ_S and σ_S^2 represent the mean and variance of S, the unnormalized predicted gaze density map [28]. In simpler terms, an NSS value of 0 indicates that the model prediction is comparable to a random eye position, meaning that there is no correlation between the real and predicted value, while an NSS value of 1 proves that the observer's eye position is in a position where the predicted saliency is one standard deviation above the average saliency. Many other comparison metrics exist, such as AUC-Borji and KL-Div [32], and the work of Riche et. al. suggests that a single metric is insufficient to fairly evaluate a model, based on a comparison of several metrics using Kendall's coefficient of concordance.

2.2.4 Discussion

Many visual attention studies focus on exploring how it can be affected in VR through characteristics of the VEs, or properties of the objects that compose them. Rivu et al. [33] used color changes of an object in the background to evaluate the impact of saliency on visual attention, and concluded that it was ineffective at diverting attention during concrete tasks or free-viewing conditions, suggesting the addition of illumination as future work. Hadnett-Hunter et al. [12] explored the effects of task itself on visual attention between free-viewing, visual search and navigation, discovering a significant difference in fixation distribution between certain tasks. For example, free-viewing and visual search had a bias towards the upper visual field, while navigation was more centralized, as displayed in figure 2.4.

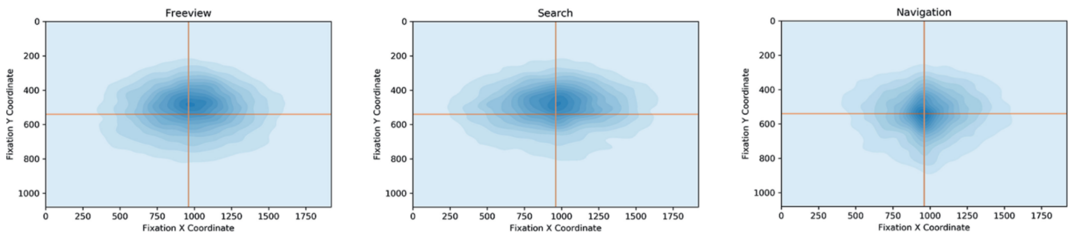


Figure 2.4: Kernel Density Estimate graphs of fixation coordinates in the three tasks.

Taken from Hadnett-Hunter et al. [12]

Finally, Sitzmann et al. [38] analyzed a gaze and head orientation dataset for users visualizing "omni-directional stereo panoramas in VR", concluding that having a smaller amount of salient regions in a scene contributes to quicker attention allocation towards them. As briefly mentioned before, the work by Rivu et al. only manipulates color, one of the many existing object properties, so performing a similar study with a new property can provide more varied insights. There is also another aspect worth exploring regarding which objects are affected by the changed properties. In the case of visual search, instead of changing an object in the background, we can affect the target object itself and others around it, studying its positive contribution for the task instead of how much it diverts attention. We decided to follow one of properties stated in their future work avenues and opt for the variation of object illumination through environment lighting.

Additionally, the reasoning behind implementing free-viewing and visual search tasks is backed by the findings provided in Hadnett-Hunter et al.'s work. While visual attention in free-viewing is more affected by bottom-up saliency, the top-down influence of having a specific task prevails in goal-oriented settings, especially visual search, where that influence was proved to be the strongest. This allowed us to explore whether or not lighting can impact visual attention by targeting bottom-up saliency, leading to faster fixations on certain objects, as well as having an effect on top-down factors, improving task completion times.

2.3 Summary

In this chapter, some fundamental concepts of virtual reality and visual attention were explored, namely immersion, virtual environments, fixations, saccades and saliency, providing a solid foundation for understanding the upcoming study of the effects of IVE characteristics on attention. Despite the existing body of knowledge on visual attention in VR analyzed in section 2.2.4, gaps remain in understanding the effects of different virtual environment properties in task-oriented scenarios within IVEs, particularly lighting, and their correlation with visual attention and task completion times. While previous studies have investigated the influence of other VE properties such as task or object color, the effects of lighting as a modifiable element remain to be discovered, especially in experimental settings where both objective and subjective metrics can be gathered. Addressing these gaps through user-focused testing and will contribute to advancing the field of virtual reality and unlocking new possibilities for immersive technology applications, with clearer objective paths and improved user satisfaction.

Chapter 3

Approach for Assessing Lighting Impact on Attention

Therefore, we propose to explore the impact of lighting on attention and task completion. By applying it to objects that compose the scenes, including task-relevant ones such as the target object and others around it, we can measure how fixation speed and task performance change in environments with and without the additional lighting. Moreover, this modification of lighting conditions must not negatively affect user presence.

3.1 Requirements

To assess the influence of lighting, it is necessary to expose users to specific conditions, collecting and analyzing a series of data. Since we want to evaluate the impact of lighting on task-oriented settings, but also establish its performance in free-viewing conditions, we must support different types of tasks. To preserve user presence between tasks, questionnaires have to be integrated in the VR experience, avoiding the premature removal of the HMD. We also need a way to interpret raw eye-tracker data into gaze and fixations. The variation of independent variables (lighting, for example) across conditions must be counterbalanced: otherwise, results that seem to be derived from the presence of lighting can actually be caused by another variable that did not get properly distributed across conditions.

Now, we can establish the necessary conditions to implement into our system. We chose to develop a free-viewing task and a visual search task, details about each one are present on section 4.1.3. In an attempt to avoid confounding variables, we change three independent variables in each condition: the presence or absence of lighting, the environment used and the positioning of the target object in the visual search task. These environments can vary between an open park, with a large city around it, and a small living room with a more limited set of props. This also allows us to evaluate if the complexity and size of each scene can affect the success of lighting variation. The target object can appear in multiple locations in each scene. Regarding the data to be collected, we register eye-tracker data in a way that allows us to interpret it into gaze and fixation data, as well as subjective data related to user presence, obtained from the Igroup Presence Questionnaire (IPQ), a standard among many user studies.

3.2 Conceptual Model

Figure 3.1 showcases a diagram of the conceptual model. It consists of a central block, the **session manager**, orchestrating the different managers based on the condition, which is decided depending on the participant order. The idea is that the only manual setup work required is specifying all of the possible conditions that originate from combinations of the independent variables, and the system is able to choose one of them in sequential order of participants. This way, the session manager stores condition-related variables and provides them to the remaining managers. Loading the scenes with the appropriate variables in place is carried by the **scene manager**, handling the environment order and transitions between them, ensuring the correct application of lighting conditions and object placements. The **task manager** is responsible for the logic of each task, which encompasses displaying instructional information, toggling the visibility of the target object, if applicable, and tracking the end of the task (timer-based in the free-viewing task and dependent on interaction with the target object in the visual search task). The **eye-tracking manager** interfaces with the eye-tracker hardware accessory of the headset and collects information about where the user is looking at during each task. Displaying the questionnaires inside VR between tasks is handled by the **questionnaire manager**. Finally, the **data manager** collects the information from the eye-tracking, the task and the questionnaire manager and transforms it into a format ready for analysis.

3.3 Architecture

Figure 3.2 represents a diagram for the system architecture. Based on the previous conceptual model, we designed an architecture that would satisfy the requirements for this user study. The final application was built using Unity v2021.3.33f and tested with the VIVE XR Elite headset [1], equipped with the official eye-tracking accessory: the VIVE Full Face Tracker [2]. A standalone build was chosen to eliminate connection cables during experiments and to enable access to the headset's eye-tracking capabilities from the application. Nonetheless, through Wi-Fi, we mirror the screen present in the HMD onto a PC, so that it was possible to follow the progress of each session. We used the VRQuestionnaireToolkit Unity Package [10] to handle questionnaire display and answer storage in Comma Separated Values (CSV) files. Cognitive3D [3], a 3D spatial analytics platform, was used to store and analyze eye-tracking data: its Unity SDK was integrated into the application to record this type of data and store it in a remote project we set up prior to the user experiments. This way, we had access to a plethora of insights from the data collected, and even

[1] https://www.vive.com/us/product/vive-xr-elite/overview/
[2] https://www.vive.com/us/accessory/vive-full-face-tracker/
[3] https://cognitive3d.com/

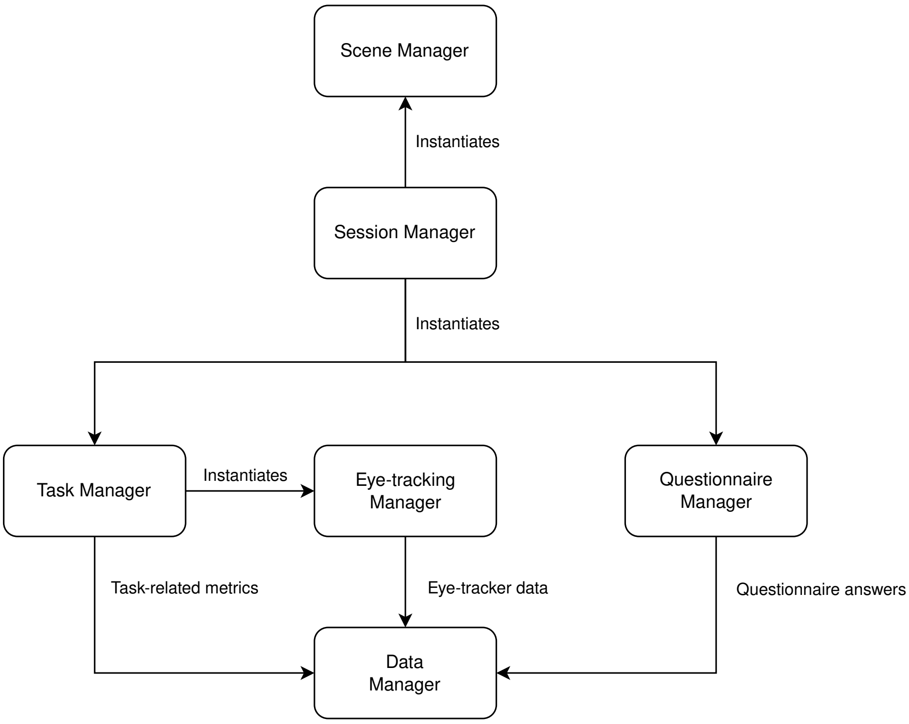


Figure 3.1: Diagram of the conceptual model

This diagram illustrates the interaction between system components during a session. The Session Manager orchestrates the experiment by instantiating core modules: Scene Manager, Task Manager (which in turn instantiates the Eye-tracking Manager), and Questionnaire Manager. These modules collect data, such as task metrics, eye-tracker data, and questionnaire responses, which are processed and stored by the Data Manager.

session replays, as we discuss in section 3.8. After finishing the user testing phase, we were able to download gaze and fixation metrics in CSV format. Task-related metrics such as completion times were stored in a JSON file. Lastly, all of the generated files from metrics and answers were submitted to a preprocessing step, accomplished via a simple Python script, to prepare them for analysis.

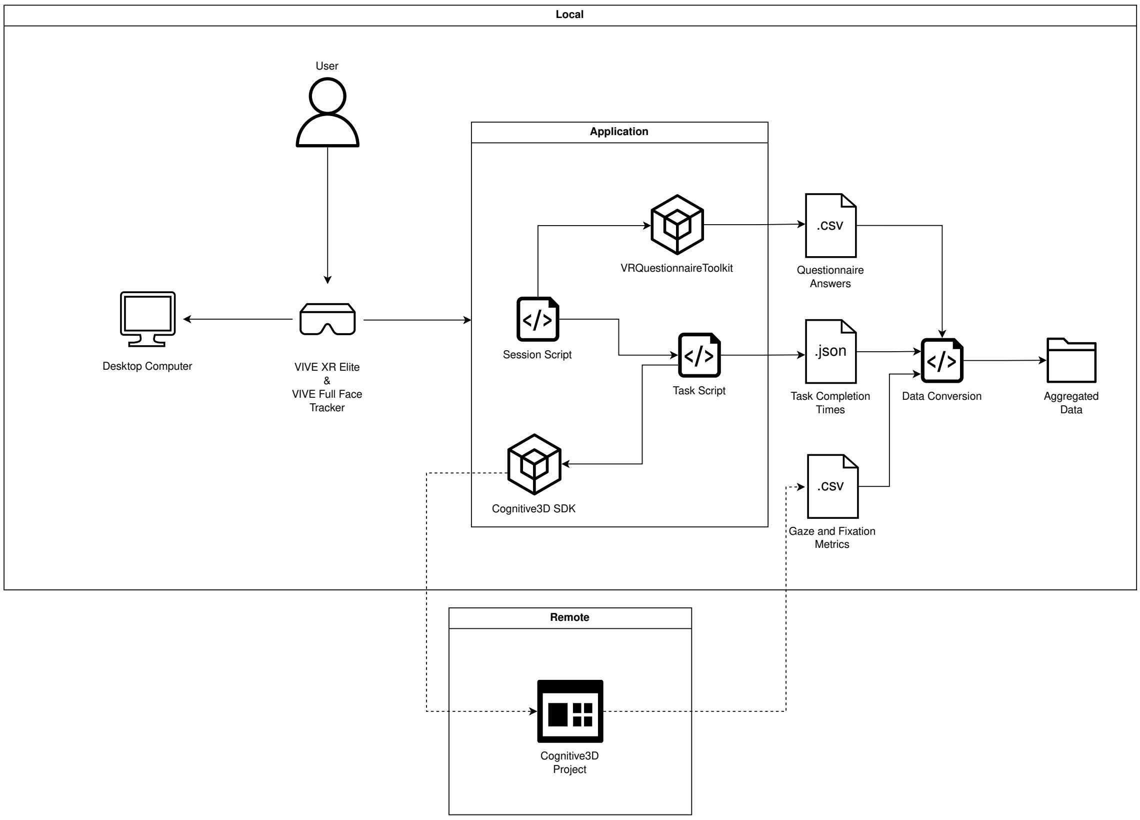


Figure 3.2: Diagram of the architecture

This diagram shows how the local VR application, running on a VIVE XR Elite headset, collects eye-tracking and task data through Cognitive3D and custom scripts. Questionnaire responses and task metrics are exported locally, while gaze and fixation data is obtained via a remote Cognitive3D project, which receives and processes the raw data from the eye-tracker. All data is then converted and aggregated for analysis.

3.4 Proof-of-concept and experimental conditions

The initial development phase focused on creating a functional prototype for the visual search task since integration of a posterior free-viewing task would be trivial. This proof of concept was implemented using Unity 6000.0.38f1 with the official VR Project Template as a starting point, configured for PCVR to facilitate rapid iteration and debugging during early development stages. The first experimental scene created was a simple park environment, using free models of trees, bushes, lamp posts, and benches from the Unity Asset Store to enrich it and make the visual search task more challenging. Later, the four benches would be moved to a preset scene from the Unity Asset Store, vastly improving the level of detail. Figures 3.3 and 3.4 show the development process before the transition to the preset.

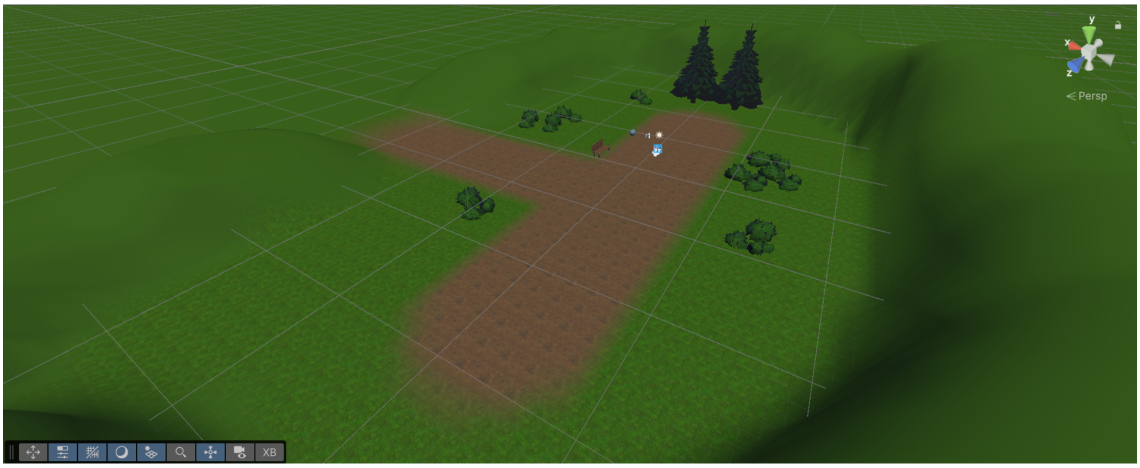


Figure 3.3: Park scene in development

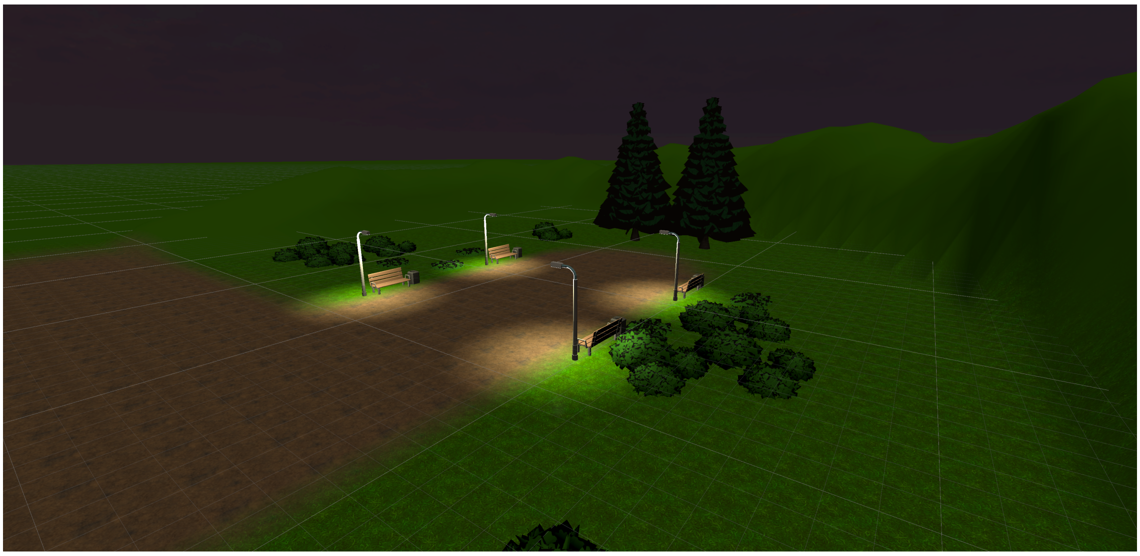


Figure 3.4: Early version of the park scene

Within this scene, a basic flow for the aforementioned task was implemented involving a single target object: the scene would load and the object would automatically spawn in a predefined location, which the user could then search and interact by pointing at it with the controller and pressing the trigger button. This target object consisted of a uniform pink sphere, in place of a more complex item such as a pencil or a book. The motive behind this decision was to circumvent potential cognitive biases related to the nature of the object: for example, a user might instinctively expect a pencil to be on top of a table or a flower to be on the ground. By using a neutral object the users would be more inclined to search everywhere in the scene for the target, with the only clue

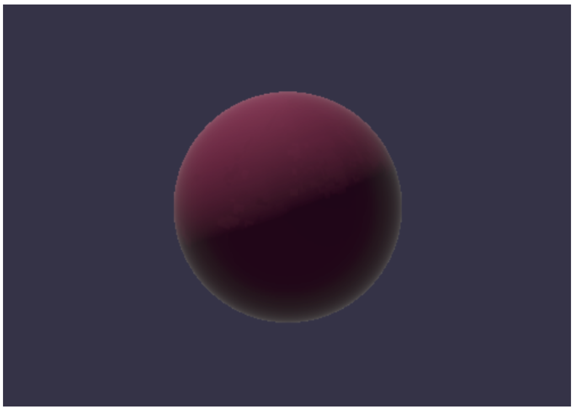


Figure 3.5: Target object used during visual search tasks

available being the presence of the additional lighting near the object in the applicable conditions. Figure 3.5 shows the sphere in question.

One aspect worth highlighting is the positioning of the additional lighting. In the visual task scenarios, this lighting is always limited to one source and is placed near the target object. No other light sources are present outside of the target object area since the goal is to evaluate how much improvement the lighting would provide in terms of task completion times. By adding light sources irrelevant to the visual search task, the performance could be negatively impacted.

This minimal setup helped validate the correctness of the visual search task implementation and future features essential for the experiment, before adding a second scene. To avoid false correlations between lighting conditions and other task-related variables such as target object positioning and scene order, these were counterbalanced across participants: a more detailed analysis of the testing procedure will be done on chapter 4. This procedure prompted the implementation of an automatic way to configure each experiment's set of values for each variable, as well as a more detailed log of each experiment which extended beyond recording the task completion time.

3.5 Configuration and Metric Logging

After defining the variable ordering for the experiments, that is, which independent variables (such as presence/absence of lighting or target object positioning) each participant will face, we needed to find a way to enforce this order in the code. As such, a decision was made to use each participant's unique sequential ID number to choose which of the experimental conditions they would see, storing these conditions in a JavaScript Object Notation (JSON) file. This JSON file allowed the definition of experimental parameters: in this case, the presence or absence of lighting in specific areas, the scene order, and the specific positions of objects and corresponding lights within each environment. In addition, the participant ID and a condition ID were automatically assigned for each session, minimizing the need for manual setup and reducing the risk of human error. This level of automation facilitated repeatable experiments between the eight different setups. These setups are based on combinations of three independent variables (lighting, target object position and scene order) and will be detailed further on Chapter 4. In addition to the aforementioned ease of configuration, the project records key performance and behavioral metrics relevant to this research's objectives. These included:

- Free-viewing and Visual search:
  - Total Gaze Count;
  - Average Time To First Gaze;
  - Total Fixation Count;
  - Average Fixation Length;
  - Average Time To First Fixation.
- Only visual search:
  - Task completion time.

A logging mechanism was implemented to register each participant's visual search task completion time, associating this data with both the automatically assigned participant ID and the specific condition ID for that trial. This ensured that all recorded metrics could be mapped to individual sessions and experimental configurations, facilitating the data analysis step.

3.6 Integration of the questionnaire

To maintain user immersion and minimize context-switching during the experiment, questionnaires were integrated directly within the VR environment. This approach allowed participants to complete subjective assessments, in this case, the IPQ, without removing the headset, preserving the continuity of the virtual experience and speeding up the process as a whole. The integration was achieved using the open source VRQuestionnaireToolkit Unity package [10], which provides an easy-to-use solution for managing industry-standard questionnaires in VR or even custom ones, with only a few adjustments needed to ensure functionality on a more modern version of Unity. In addition to already including many frequently used questionnaires, such as NASA-TLX and IPQ, it also offers automatic export of responses in '.csv' format, making it compatible with common data analysis software like SPSS, the program chosen during this research. In this study, after each questionnaire submission, the package saved the answers in a '.csv' file for each user and aggregated every answer in a global file to facilitate further investigation.

The users were asked to fill out each questionnaire in a desert environment, as shown in figure 3.6. The reasoning behind creating a different scene was to separate the previous task from the questionnaire itself, as it is meant to be related to both tasks (free-viewing and visual search) and not just the latter. As the number of scenes being developed grew, a well-defined flow to manage the loading/unloading of environments and the transitions between them was required.

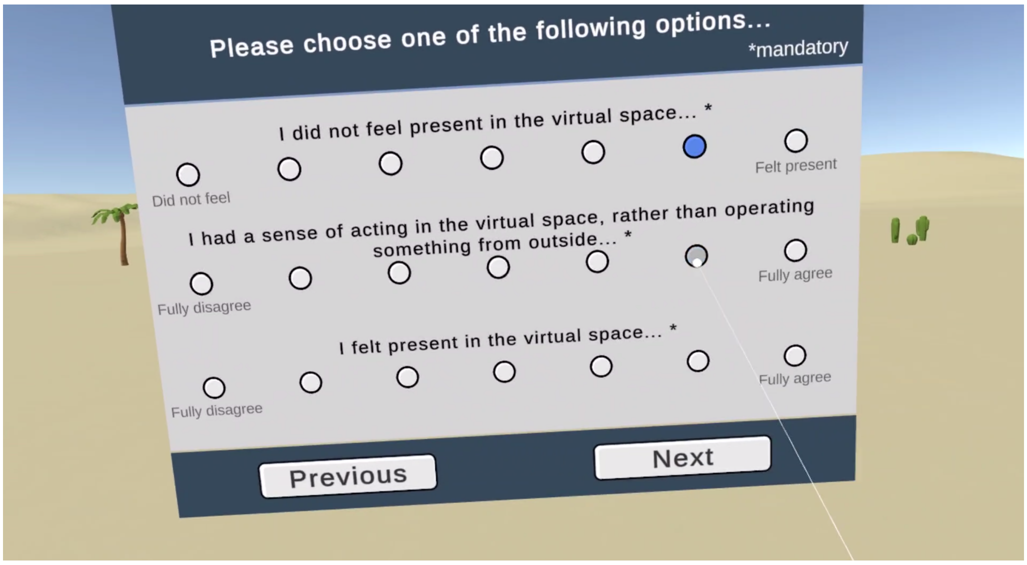


Figure 3.6: Questionnaire page displayed with VRQuestionnaireToolkit

3.7 Experiment flow

The experimental application was structured around a modular scene-based architecture to facilitate a controlled testing experience. Four primary scenes were developed:

- StartupScene: Set in a desert, it introduced participants to the experiment and provided initial instructions related to interaction and questionnaires;
- RoomScene and ParkScene: Served as the main environments for free-viewing and visual search task execution, each one offering distinct spatial contexts: while the room presented a more contained, simpler environment, the park was a fully-fledged preset taken from the Unity Asset Store, slightly modified to improve performance in a standalone build and to incorporate the intended four-bench setup shown previously in 3.4;
- QuestionnaireScene: With an environment similar to the StartupScene, it allowed participants to complete assessments after each task set within the VR environment.

Upon launching the application, the program loads the ‘StartupScene‘, reads the configuration settings, and directs the participant through the predefined sequence of scenes, according to the configuration file, loading the questionnaire scene whenever a free-viewing and a visual search task are completed in sequence. Fade-in and fade-out animations were implemented between scene changes, preventing abrupt visual shifts that could disrupt the user’s experience. Initially, the experimental design employed a between-subjects approach, assigning different participants to a condition of presence/absence of light. However, as it requires a smaller sample size and prevents individual factors from affecting only one of the conditions, the design was transitioned to a within-subjects framework, in which each participant experienced both the presence and the absence of light: more details in the next chapter. The aforementioned configuration file allowed for this change with minimal adjustments, highlighting the importance of its inclusion in enabling faster iteration between experimental requirements.

3.8 Eye Tracking Integration

Integrating eye tracking into the VR experiment was essential for capturing detailed insights into participants' visual attention and behavior. By recording gaze direction and fixation patterns, it becomes easier to analyze how lighting conditions influence attention allocation during free-viewing tasks. To facilitate comprehensive data collection via the eye-tracker accessory and its analysis, especially regarding the spatial aspect of fixation locations, the Cognitive3D platform was employed. Cognitive3D is a spatial analytics solution designed for VR and AR applications, enabling the capture of user interactions within 3D environments. Its Unity SDK allows for the recording of various metrics, including gaze data, fixation points, and session durations, becoming accessible through a dashboard shown in figure 3.7. Additionally, Cognitive3D provides visualization tools such as heatmaps and replay features, which are invaluable for interpreting user behavior in spatial contexts.

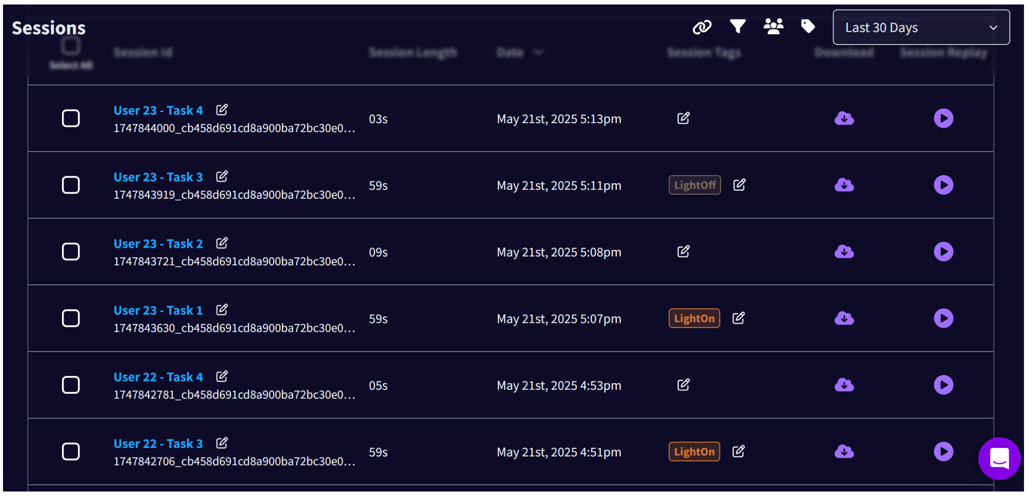


Figure 3.7: Session page inside Cognitive3D

As seen in figure 3.8, eye fixations are displayed as blue semitransparent spheres. In Section 2.2, we have mentioned dispertion-based methods to identify fixations and distinguish them from saccades, and Cognitive3D uses a dispertion threshold technique. This approach groups sequential gaze points that fall within a defined spatial threshold and persist for a minimum temporal duration, in our case. If the gaze points remain spatially clustered (directions within 1º) and meet the minimum time threshold (60 ms), they are considered a fixation. Otherwise, the system classifies them as part of a saccade or transient movement. During a fixation, as soon as the user deviates from its direction for 10ms or longer, the fixation ends. This method allows for real-time fixation detection in immersive 3D environments.

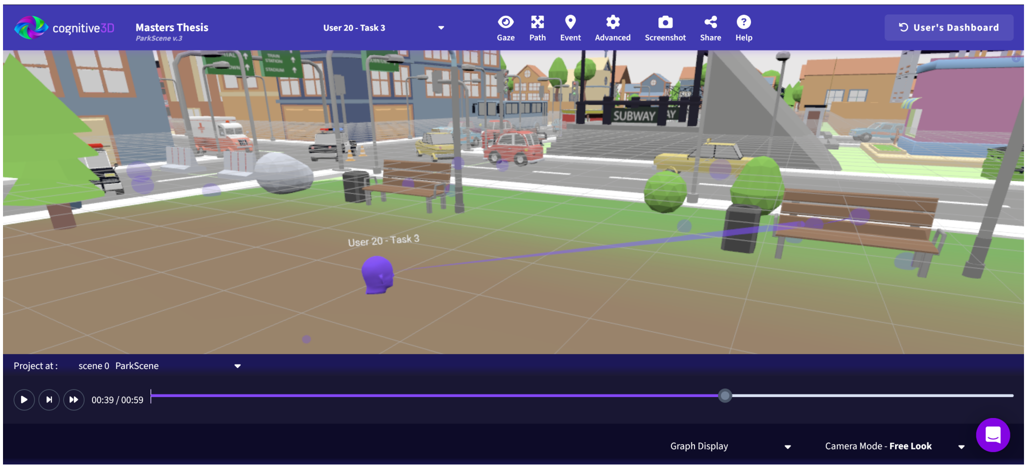


Figure 3.8: Example of a task replay inside Cognitive3D

The development was, at first, conducted using a PCVR setup with the VIVE OpenXR plugin. However, to utilize the eye-tracking capabilities of the VIVE XR Elite headset and provide an untethered user experience, a transition to a standalone Android build was necessary. This shift required moving from the VIVE OpenXR plugin to the VIVE Wave SDK, as the latter offered better support for eye-tracking features on standalone devices and was the only VIVE SDK compatible with Cognitive3D at the moment of development. During this transition, compatibility issues arose between the latest Unity version and the VIVE Wave SDK. Specifically, the SDK had not yet been updated to support Unity 6, leading to unavoidable compilation errors. To resolve these issues, the development environment was downgraded to Unity 2021.3.33, which provided stable support for both the VIVE Wave SDK and the Cognitive3D Unity SDK, at the cost of a lengthy transition process which involved debugging many aspects of the conversion from OpenXR to Wave SDK. With the appropriate setup in place, the system captured detailed eye-tracking data. This data was then integrated with Cognitive3D's analytics tools, allowing for the recording and visualization of fixation points and gaze paths throughout the VR sessions. The combination of these technologies enabled an informed evaluation of how lighting conditions affected user attention within virtual environments, and the storage of more complex information such as session replays has proved useful in the data analysis stage.

3.9 Summary

The final experimental setup was a standalone Unity application deployed on the VIVE XR Elite with integrated eye tracking, enabling the execution of tasks and questionnaires without headset removal. It fulfilled all research requirements by collecting gaze, fixation, and visual search execution time data using Cognitive3D, allowing analysis of user attention under varying lighting conditions. Essential features included configurable experimental conditions and in-headset questionnaires, which improved development efficiency and maintained user presence. A recording of a simulated session is available on YouTube [4]. Several key development milestones contributed to the success of the final solution. The initial prototype with basic interaction logic, the integration of the VRQuestionnaireToolkit and the modifications made to accomodate eye-tracking data collection via the Cognitive3D SDK were the most defining moments of the development phase. These efforts resulted in a platform ready to be applied in the user study, collecting all of the necessary data for further analysis and then allowing us to tackle the research questions defined in section 1.3. The final versions of each scene developed can be seen on figures 3.9, 3.10 and 3.11, with the red circles present in the first two figures representing the possible target object positions. In the next chapter, an in-depth exploration of the testing procedure will be conducted to describe exactly how these were obtained, and to prove the absence of unintended biases or erroneous procedures that could have tainted the data, followed by an analysis of the results.

Figure 3.9: Final version of Park scene

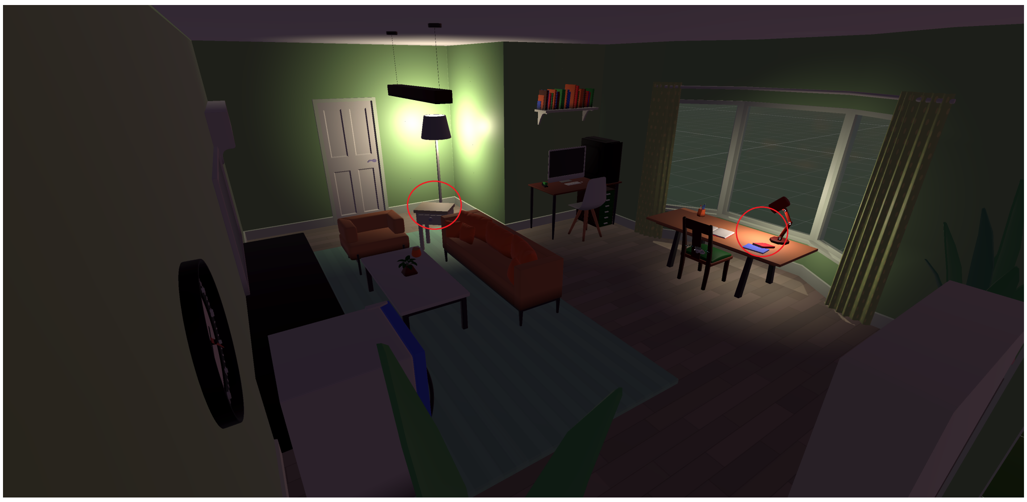


Figure 3.10: Final version of Room scene

[4]https://www.youtube.com/watch?v=s6ud7mZrLbg

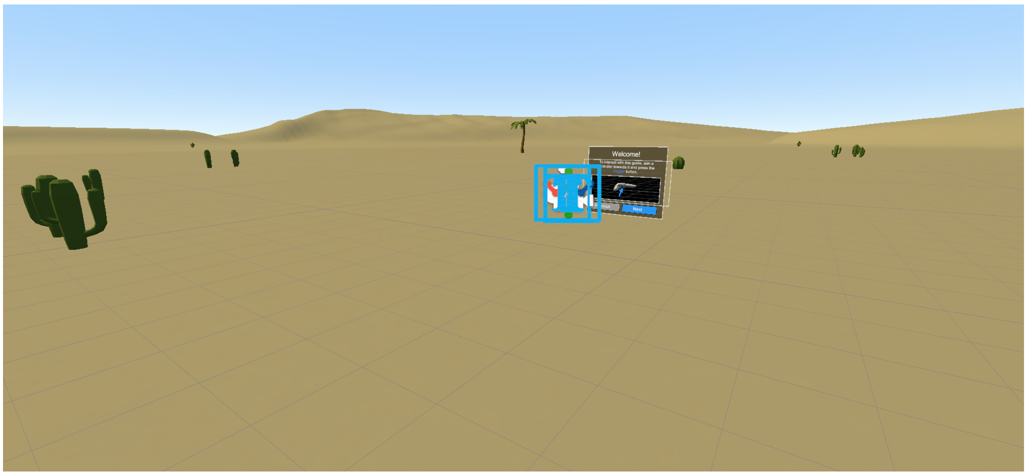


Figure 3.11: Final version of Startup scene (similar to Questionnaire scene)

Chapter 4

Evaluation and Results

At this stage, it is important to remember the research questions:

- **RQ1:** Is lighting effective in drawing user attention without the influence of a task?
- **RQ2:** Does the presence of lighting around a target object affect the performance of its visual search without adversely affecting the level of presence?

To address the aforementioned questions, performing tests with users and analyzing the practical impact this target variable has is mandatory. Ideally, the tasks completed inside a scene with the additional lighting on must result in a decreased time to first fixation in objects near the lights, for the free-viewing task, and in a significantly faster completion time for the visual search task.

4.1 Methodology

This section outlines the methodology used to conduct the experimental phase of the study, detailing the technical setup, study design, task structure, and ordering of independent variables. Each of these aspects is explained in detail in the following subsections, and will clarify how the data described in section 4.2 was gathered.

4.1.1 Experimental Setup

The VIVE XR Elite headset with the Full Face Tracker accessory, seen in figure 4.1, was linked to the PC to mirror its display, keeping the supervisor aware of the experiment's progress and allowing them to provide aid as a last resort. The experiments were conducted in the Graphics, Interaction, and Games Laboratory at FEUP, where a large open area was available to ensure participant safety and freedom of movement. A vast majority of the users were contacted directly to participate in this study, with only a few of them being informed via email. Each experiment started with a brief explanation of the procedure, followed by a demographics questionnaire with some targeted questions about previous VR experience, answered on a laptop. Upon submission, the VR portion of the experiment began, starting with a calibration of the eye tracker to adjust to each participant's eyes, improving the accuracy of the data. Afterwards, the experiment started in-app with a brief overview of the interaction method and the in-app questionnaires. The users would then complete a free-viewing and a visual search task, followed by the IPQ. After repeating this three-step process for the second time in different lighting conditions, participants would remove the headset and fill the last questionnaires in the laptop. A diagram summarizing the experimental flow is shown in 4.2. All of the questionnaires used in this study can be verified on Appendix C.

Figure 4.1: The Vive XR Elite headset with the Full Face Tracker accessory.

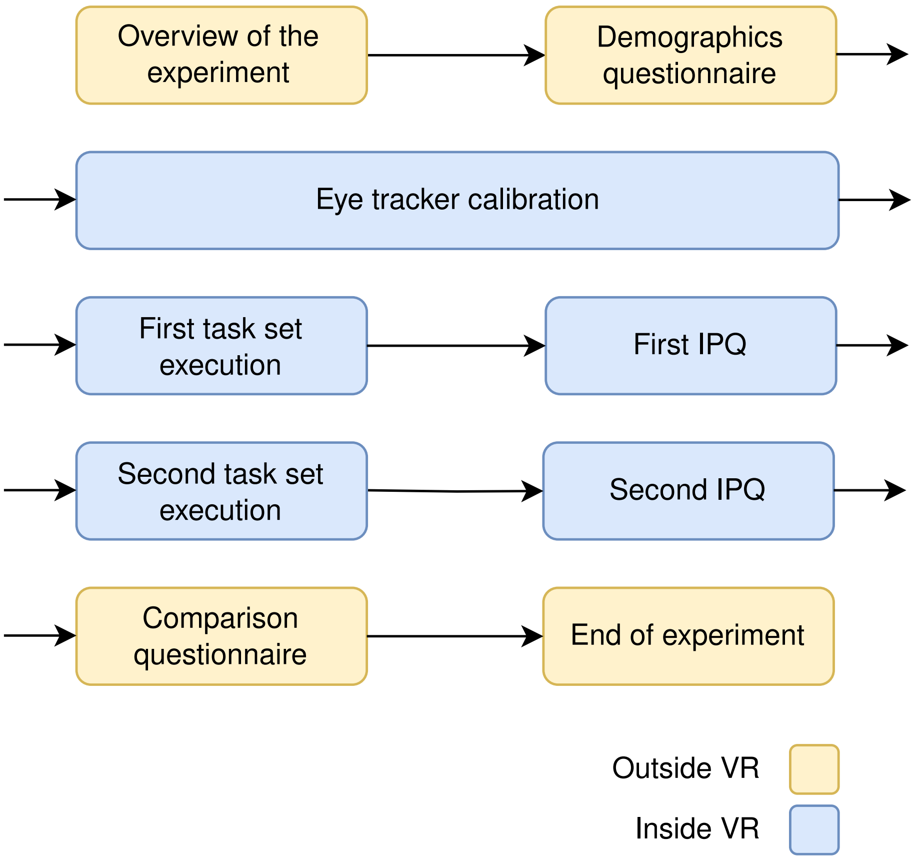


Figure 4.2: Diagram of the experimental flow.

4.1.2 Design

Initially, the study was designed to follow a between-subjects design. However, as previously mentioned, it was switched to a within-subjects design, due to requiring a smaller sample size and preventing individual factors from affecting the conditions unevenly. This way, not as many participants are required to achieve the same statistical power, although it introduces other concerns such as carryover effects. This change naturally prompted other modifications in the methodology which will be mentioned later. The primary independent variable in this study is the presence or absence of lighting in certain spots, including the target object area during the visual search tasks. Additionally, the target object positioning, as well as the order of the scenes across tasks, are also independent variables, counterbalanced to avoid false correlations between the primary independent variable and the dependent variables. These dependent variables are objective metrics such as the gaze/fixation data and subjective measures from the IPQ and comparison questionnaire, which will be described in more detail in Section 4.2

4.1.3 Tasks

Two tasks were implemented for this study, whose order remained the same throughout every experiment. The first task, free-viewing, was basically a sandbox, the absence of a concrete task: users merely had to look around the scene on their own for one minute. This was useful to answer the first research question, since the goal is to evaluate the impact of the additional lights on their own, without the influence of a search task which can condition the way users look at a scene [12], for example. Figure 4.3 shows the description of the free-viewing task, displayed prior to its start.

Figure 4.3: In-game description of the free-viewing task

The second one, visual search, prompted users to look for an object in the least amount of time possible, shown previously in section 3.4. Before starting the timer, in order to practice the controller movement and acknowledge the object's appearance, each user had to emulate the interaction with the target object, by pointing at an identical copy in front of them and pressing the trigger button, as seen in figure 4.4. Once that step was completed successfully, the "Start" button would become available and users were able to begin at any time, but only upon pressing the button would the timer start to count and spawn the target object.

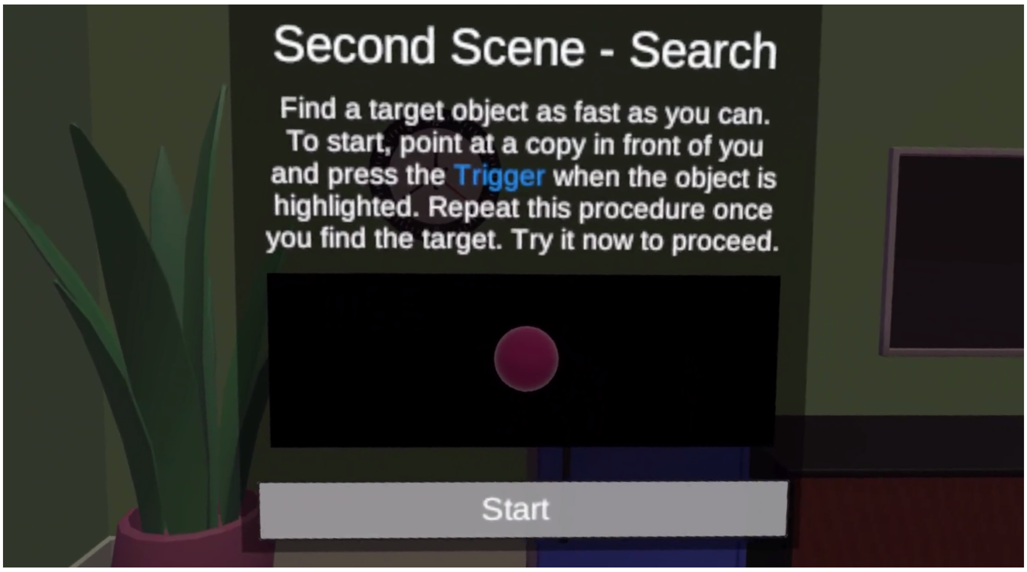


Figure 4.4: In-game description of the visual search task

4.1.4 Order

In within-subjects designs, avoiding carryover effects from the first to the second target variable is paramount, as well as any other undesirable impact from variables such as target object positioning or scene order. As such, using a partial counterbalancing strategy, a set of eight versions of the experiment were defined, ensuring a balanced group of results for every condition, according to table 4.1.

Scene Pair Order

- **1** – Park → Room
- **2** – Room → Park

Target Object Positions

- **A** – Bench at coordinate Z− (Park) / Big Table (Room)
- **B** – Bench at coordinate Z+ (Park) / Small Table (Room)

Lighting Condition Order

- **X** – No Lighting → Lighting

- **Y** – Lighting → No Lighting

Assuming that the two pairs of tasks are executed in the same order for each environment condition:

Table 4.1: Experimental versions based on conditions

Version	**Conditions**	
I	1AX	2AX
II	1AY	2AY
III	1BX	2BX
IV	1BY	2BY
V	2AX	1AX
VI	2AY	1AY
VII	2BX	1BX
VIII	2BY	1BY

The different versions were assigned sequentially for each user, resulting in an equal amount of results for every version.

4.2 Gathered Data

During the testing sessions, several data was provided by each participant to be later analyzed, using demographics questionnaires before the session, as well as after each pair of tasks using the IPQ. A set of metrics was also automatically collected via the eye-tracker and using scripts inside the Unity project.

4.2.1 Demographics

The first questionnaire obtains general information about the user, as well as some domain-specific information relevant to the study:

- Age
- Gender
- Experience with type of VR device
- Type of VR experience (games, simulation, research. . .)
- VR usage frequency ("never", "once every few months", "once a month", "once a week", "every day")

4.2.2 Metrics

Throughout the experiment, several visualization and performance metrics were also collected, some of them provided by the Cognitive3D integration. These include:

- Average gaze count
- Average time to first gaze
- Average fixation count
- Average fixation length
- Average time to first fixation
- Visual search task completion time

It is important to note that the visualization metrics will only be analyzed on free-viewing scenarios, since Cognitive3D had trouble registering fixations during the visual search task, potentially due to the placement of the target object (relatively distant to the user and on the edge of tracked objects such as the benches). The visualization metrics were tracked on the potential target object locations of the visual search task, as well as the two additional benches present in the park scene and a floor lamp in the room scene (replacing the mostly hidden small table, the second target location in the room scene).

4.2.3 Follow-up Questionnaires

Inside the HMD, the IPQ measures the user's presence in 4 separate categories: general presence (PRES), spatial presence (SP), involvement (INV) and experienced realism (REAL).

Finally, a comparison questionnaire at the end of the experiment was added to better understand the perception of each user regarding the different lighting conditions. It is composed of two binary choice questions (1 and 4) and two 5-point Likert scale questions (2 and 3):

1. There was a noticeable difference in the lighting conditions between the two pairs of tasks.
2. The added lighting affected my exploration of the scenes.
3. The added lighting helped me find the target object more easily.
4. Overall, I preferred the experience provided in the (1st/2nd) pair.

4.3 Results

A total of 24 users participated in the experiment, one of them shown in figure 4.5. These participants ranged from 18 to 41 years old, with 12 men, 9 women, 3 non-binary and 1 unspecified (one person selected both "Man" and "Non-binary"). 15 users had prior VR experience, and within a 12-month span prior to the experiment, most of them had only used it once every few months, although 2 used their headsets once a month and 1 used it every day. 7 participants had already taken part in other user studies related to VR. Outliers were evaluated using a boxplot visualization method, and no extreme outliers were removed, as only one existed and it represented a legitimate attempt at finding the target object, not stemming from confusion or errors in the system.

When reporting the results, we sometimes use abbreviations, which are all listed below:

- M - Mean
- MDN - Median

- SD - Standard Deviation
- IQR - Interquartile Range

Figure 4.5: Participant performing the experiment

4.3.1 Gaze / Fixation Metrics and Task Completion Times

To assess the impact of lighting on user attention and task performance, a set of quantitative metrics was collected and analyzed. The data was segmented by lighting condition to evaluate how the presence or absence of lighting influenced user behavior. Table 4.2 summarizes the collected visualization metrics per condition, while table 4.3 displays the average time to first fixation for each tracked object. Further statistical analysis is presented in tables 4.4 and 4.5, which compare time to first fixation values across conditions within each scene. Finally, table 4.6 presents statistics for task completion times under both lighting conditions. The last three tables showcase the results of the corresponding statistical significance test performed as well.

Table 4.2: Visualization and performance metrics by lighting condition

Metric	With Lighting	Without Lighting
Gaze Count ∣ M (SD)	11.39 (10.85)	11.72 (13.23)
Fixation Count ∣ M (SD)	4.06 (4.00)	4.33 (4.87)
Fixation Length (s) ∣ M (SD)	0.61 (0.42)	0.64 (0.64)
Time to First Gaze (s) ∣ M (SD)	14.05 (6.38)	14.67 (4.69)
Time to First Fixation (s) ∣ M (SD)	17.34 (6.28)	19.18 (5.83)
Task Completion Time (s) ∣ M (SD)	8.33 (5.703)	9.90 (10.41)

Table 4.3: Average time to first fixation per object under both lighting conditions

Object	Avg. Time to First Fixation - With Lighting (s)	Avg. Time to First Fixation - Without Lighting (s)
Big Table ∣ M (SD)	13.11 (14.20)	10.74 (7.14)
Floor Lamp ∣ M (SD)	9.02 (7.79)	15.99 (13.42)
Bench Section (Behind Right) ∣ M (SD)	18.40 (16.44)	25.03 (13.03)
Bench Section (Behind Left) ∣ M (SD)	26.98 (12.40)	21.50 (6.64)
Bench Section (Ahead Right) ∣ M (SD)	20.96 (16.32)	16.28 (18.26)
Bench Section (Ahead Left) ∣ M (SD)	15.56 (18.30)	25.52 (21.46)

Table 4.4: Time-to-first fixation under both lighting conditions in the park scene

Time to First Fixation - Park Scene		
Statistics	**No Lighting**	**Lighting**
M (SD)	22.20 (15.70)	21.46 (15.41)
Mann-Whitney U		
U (p)	477 (0.98)	

Table 4.5: Time-to-first fixation under both lighting conditions in the room scene

Time to First Fixation - Room Scene		
Statistics	**No Lighting**	**Lighting**
M (SD)	13.12 (10.54)	11.15 (11.52)
Mann-Whitney U		
U (p)	213 (0.36)	

Table 4.6: Task completion time under both lighting conditions

Task Completion Time		
Statistics	**No Lighting**	**Lighting**
M (SD)	9.50 (7.77)	7.98 (5.84)
Wilcoxon Signed-Rank		
Z (p)	-0.29 (0.76)	

A Shapiro-Wilk normality test was performed for each set of data from both tasks. In every scenario, we concluded that the data was not normally distributed for either condition ($p < 0.05$). For the task completion times, a Wilcoxon signed-rank test was performed, since every user experienced the visual search task in both conditions and the objective remained the same. However, in the free-viewing task, there were a different number of objects being tracked in each scene (four in the park scene and two in the room scene), so we decided to perform a separate analysis for each of the two scenes. Since each user only executed the free-viewing task in one of the lighting conditions per scene, we have to consider both groups independent in each scene: a Mann-Whitney U test is more adequate in this situation. The Mann-Whitney U test indicates that no statistical difference was found in time to first fixation for each scene ($p > 0.05$). Similarly, the Wilcoxon signed-rank test showed that additional lighting in the target object area did not cause a statistically significant change in the completion times of visual search tasks ($p > 0.05$), as shown in 4.6.

4.3.2 IPQ

For the IPQ, we retrieved and analyzed the scores from each category and for the questionnaire's total score, to evaluate the effect of added lighting in user presence as a whole and in each of its components. Tables 4.7, 4.8, 4.9, 4.10 and 4.11 statistically describe the scores for each category and for the sum of every category in both lighting conditions.

Table 4.7: Presence score under both lighting conditions

Presence Score		
Statistic	**No Lighting**	**Lighting**
MDN (IQR)	5.00 (1.75)	4.00 (1.00)
Wilcoxon Signed-Rank		
Z (p)	-0.97 (0.34)	

Table 4.8: Spatial presence score under both lighting conditions

Spatial Presence Score		
Statistic	**No Lighting**	**Lighting**
M (SD)	4.43 (0.89)	4.31 (1.12)
Dependent T-Test		
t(23) (p)	0.75 (0.46)	

Table 4.9: Involvement score under both lighting conditions

Involvement Score		
Statistic	**No Lighting**	**Lighting**
M (SD)	3.48 (1.22)	3.38 (1.29)
Dependent T-Test		
t(23) (p)	0.63 (0.54)	

Table 4.10: Experienced realism score under both lighting conditions

Experienced Realism Score		
Statistic	**No Lighting**	**Lighting**
M (SD)	2.28 (0.90)	2.48 (1.03)
Dependent T-Test		
t(23) (p)	-2.15 (0.04)	

Table 4.11: Total score under both lighting conditions

Total Score		
Statistic	**No Lighting**	**Lighting**
M (SD)	14.77 (3.24)	14.63 (3.52)
Dependent T-Test		
t(23) (p)	0.33 (0.74)	

The Shapiro-Wilk normality test was once again employed for each variable, concluding that only the presence score had a non-normal distribution out of the five. As such, for the presence score, a Wilcoxon signed-rank test was performed, while a Dependent T-Test was chosen for the remaining variables. Every variable had a non-significant difference between the two lighting conditions ($p > 0.05$), except for experienced realism ($p < 0.05$), in which the condition with added lighting was perceived as more realistic.

4.3.3 Comparison Questionnaire

In the comparison questionnaire, 14 users declared a noticeable difference in the lighting conditions between the two pairs of tasks (Q1). Out of these 14, only 3 people preferred the first pair of tasks, which were all pairs with no additional lighting (Q4). Tables 4.12 and 4.13 outline the scores for questions 2 and 3 of the questionnaire.

Table 4.12: Results from question 2 of the comparison questionnaire

Q2 - Lighting helps exploring	
Statistic	**Value**
MDN	4.00
IQR	1.00

Table 4.13: Results from question 3 of the comparison questionnaire

Q3 - Lighting helps finding targets	
Statistic	**Value**
MDN	4.00
IQR	1.00

4.4 Discussion

To address the research questions proposed in this study, whether lighting effectively draws user attention in free-viewing conditions (RQ1), and whether it improves visual search performance without compromising user presence (RQ2), we now examine and interpret the results.

The time to first fixation results tell us that the difference between the presence or absence of additional lighting is not statistically significant, in other words, the objects are not fixated significantly faster with the additional lighting, even if it shows a slight improvement. However, we can compare the results from tables 4.4 and 4.5 and see that the difference between both lighting conditions is steeper in the room scene than in the park scene. This can be explained by the park scene being an open area with much more detail, thus causing the lighting to be less effective, while the closed environment of the room scene with fewer objects to look at made the added lighting matter slightly more. Alternatively, we can hypothesize that this change loses its impact as more objects are involved in it: four objects were manipulated in the park scene compared to just two in the room scene. This greater difference seems to be aligned with related work by Sitzmann et al. [38] discussed in section 2.2.4: fewer salient regions lead to faster attention allocation. Based on these results, we can say that the answer to RQ1 is "no", since there is no statistical difference between both lighting conditions regarding time to first fixation.

When analyzing the visual search completion time results, although being slightly lower with additional lighting, we concluded that there is no statistically significant difference between the two lighting conditions, meaning that the addition of lighting did not directly help users in finding the target object. Along with the relative simplicity of the scenes used, the target object positions were all visible from the user's starting position, making the visual search task easier, which can explain the results obtained.

In regards to the IPQ answers, the tests reported a lack of significant difference in all but one category: experienced realism. In general, users believed the scenes with added lighting spots were more realistic than the ones without it, an expected behavior if we consider that we usually turn on lights in our rooms or light up the lamp posts on the streets whenever the sun is set.

Finally, the comparison questionnaire shows that almost half of the users did not consciously notice a difference in the lighting conditions between the two sets of tasks. The majority of those who did feel a distinction reported that the lighting affected their exploration of the scenes and helped them find the target object. However, all but three of these users preferred the second pair of tasks, which were not always the tasks with added lighting. This can be justified by a recency bias, where users had a tendency of feeling like they performed better in the tasks after experiencing them before in the first pair, thus having a preference for the last set. Observing the results from visual search task completion time, IPQ scores and comparison questionnaire answers, the answer to RQ2 is that lighting does not affect visual search task performance, since there is no statistical difference between the task completion times across both conditions. However, the inclusion of lighting can positively affect user presence, specifically in the experienced realism category, as we saw that the difference in its scores were statistically significant. Additionally, users that noticed the change in lighting conditions reported a positive influence of the additional lighting in both types of tasks.

In conclusion, we were able to answer both research questions based on the results obtained. Despite not being able to verify a significant improvement in user attention or task performance with the presence of lighting, there are avenues of future work that can be explored, such as the question of balancing the amount of tracked objects per scene. This, along with limitations and an overview of the study, will be discussed in chapter 5.

Chapter 5

Conclusions

Virtual Reality, and IVEs in particular, present many fundamental differences from traditional 2D or 3D environments and, as such, require a distinct body of research to be carried out. The impact of the virtual environment on user attention is an underdeveloped topic in the virtual reality area and was part of the motivation behind this research, with a focus on lighting.

In the related work, we established fundamental understanding of VR and visual attention concepts: immersion and presence, as well as behavior like fixations, saccades and blinking, along with object characteristics such as saliency, within the context of IVEs. While the influence of saliency on attention has been studied in previous works, we identified gaps regarding its role in task-oriented scenarios, and the lack of exploration of lighting as a medium to attract attention. This research addresses the described gaps by implementing a solution which focuses on exploring the impact of lighting on both free-viewing and visual search task scenarios.

After evaluating the results of a user study with 24 participants, conducted in laboratory environment, we concluded that the effect of added lighting, despite causing a noticeable difference in time to first fixation and visual search task execution times, had no statistically significant impact. Nonetheless, users considered that the scenes with added lighting were more realistic and those who consciously noticed a difference in the lighting conditions generally felt like it affected their exploration of the scenes and helped them find the target object during the visual search tasks. We found that, while lighting does not meaningfully contribute to better task performance and more effective user attention capturing in free-viewing settings, it provides a positive influence on user presence, particularly in experienced realism. These findings are crucial for developing more stimulating VEs, and open up the path for related future work discussed in section 5.1.

Overall, we consider that this research presents important information regarding the improvement of IVEs with respect to the compromise between presence and user experience. While this field's research body of knowledge has a large room for development, studies like these can contribute to expanding it, providing developers with important insights to perform informed decisions when constructing their virtual experiences, such as when and where to include light sources in the environment, in this case. The developed testing platform could also be adapted to other experiments with similar objectives and extended to encompass more types of tasks, benefiting VR research with heavy user testing components.

5.1 Limitations and Future Work

This study had some limitations: we opted to focus on one characteristic of the virtual environment - lighting, at risk of being unable to effectively control confounding variables and being forced to perform a more surface-level analysis given the time constraints. A greater variety of tasks, such as navigation, could have also been explored to evaluate the effect of lighting in a broader sense, potentially discovering a task in which it has greater influence. Moreover, several changes had to be made throughout development, namely the transition to a standalone build instead of PCVR, a version downgrade of the Unity Editor due to VIVE Wave SDK compatibility issues and switching from a between-subjects to a within-subjects design. The latter was a quick process due to the ease of experimental configuration, where every setting was centralized in a file (in the case of independent variables) or automatically set (participant ID and experimental condition), but it cost development time as a whole.

One very important avenue of potential future work is exploring other object/environment properties mentioned in 2.2 (shape, movement...), as well as other tasks. These properties may yield better results than simply manipulating lighting, and might work best if employed together. Another aspect worth noting is the fact that all locations of the target objects were already visible from the user's starting position, only requiring head movements: we decided to keep the experiments as simple as possible due to aforementioned constraints, but an approach where some positions of the target object require the user to move around the scene could result in a bigger difference between experimental conditions. In fact, regarding movement inside the scene, the implications might differ from a stationary exploration approach, potentially yielding distinct and interesting results. As most studies focus on the latter, an in-depth study on how saliency affects attention while moving can provide valuable insights. Measuring the effect of lighting on different amounts of objects at once can also be an interesting avenue of work: as we mentioned before, the scene with less objects affected by the added lighting benefited from a higher improvement in average time to first fixation.

References

[1] Bhojan Anand. Do Guidance Affect Immersion in Educational VR Simulations? In *2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, pages 1–8, December 2024.

[2] Sahand Azarby and Arthur Rice. Understanding the Effects of Virtual Reality System Usage on Spatial Perception: The Potential Impacts of Immersive Virtual Reality on Spatial Design Decisions. *Sustainability*, 14:10326, August 2022.

[3] Sara Baldoni, Mahmoud Z. A. Wahba, Marco Carli, and Federica Battisti. A Study on the Impact of Virtual Reality on User Attention. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, June 2023.

[4] Fatma Beji, William de Paula Ferreira, Isabelle Pivotto Dabat, and Vitor Matias. CAVE Automatic Virtual Environment Technology: A Patent Analysis. *Engineering Proceedings*, 89(1):9, 2025.

[5] ALAN Chalmers and KIRSTEN Cater. 41 - Exploiting Human Visual Perception in Visualization. In Charles D. Hansen and Chris R. Johnson, editors, *Visualization Handbook*, pages 807–816. Butterworth-Heinemann, Burlington, January 2005.

[6] Murat Coban, Yusuf Islam Bolat, and Idris Goksu. The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review*, 36:100452, June 2022.

[7] H. Creagh. Cave Automatic Virtual Environment. In *Proceedings: Electrical Insulation Conference and Electrical Manufacturing and Coil Winding Technology Conference (Cat. No.03CH37480)*, pages 499–504, September 2003.

[8] Rudolph Darken and Barry Peterson. Spatial Orientation, Wayfinding, and Representation. *Handbk Virtual Environ*, 2002, November 2001.

[9] Krista A. Ehinger, Barbara Hidalgo-Sotelo, Antonio Torralba, and Aude Oliva. Modeling Search for People in 900 Scenes: A combined source model of eye guidance. *Visual Cognition*, 17(6-7):945–978, August 2009.

[10] Martin Feick, Niko Kleer, Anthony Tang, and Krueger Antonio. The Virtual Reality Questionnaire Toolkit. In *The 33st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, UIST '20 Adjunct, New York, NY, USA, 2020. Association for Computing Machinery. event-place: Minneapolis, Minnesota, USA.

[11] Carlos Flavián, Sergio Ibáñez-Sánchez, and Carlos Orús. The impact of virtual, augmented and mixed reality technologies on the customer experience. *Journal of Business Research*, 100:547–560, July 2019.

[12] J. Hadnett-Hunter, G. Nicolaou, E. O'Neill, and M.J. Proulx. The effect of task on visual attention in interactive virtual environments. *ACM Transactions on Applied Perception*, 16(3), 2019. Publisher: Association for Computing Machinery.

[13] Honglei Han, Aidong Lu, and Unique Wells. Under the Movement of Head: Evaluating Visual Attention in Immersive Virtual Reality Environment. In *2017 International Conference on Virtual Reality and Visualization (ICVRV)*, pages 294–295, October 2017.

[14] Zhiming Hu. Gaze Analysis and Prediction in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 543–544, March 2020.

[15] Zhiming Hu. [DC] Eye Fixation Forecasting in Task-Oriented Virtual Reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 707–708, March 2021.

[16] Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. FixationNet: Forecasting Eye Fixations in Task-Oriented Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2681–2690, May 2021. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[17] Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. EHTask: Recognizing User Tasks From Eye and Head Movements in Immersive Virtual Reality. *IEEE transactions on visualization and computer graphics*, 29(4):1992–2004, April 2023.

[18] Zhiming Hu, Sheng Li, and Meng Gai. Temporal continuity of visual attention for future gaze prediction in immersive virtual reality. *Special issue on Visual interaction and its application*, 2(2):142–152, April 2020.

[19] Zhiming Hu, Congyi Zhang, Sheng Li, Guoping Wang, and Dinesh Manocha. SGaze: A Data-Driven Eye-Head Coordination Model for Realtime Gaze Prediction. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2002–2010, May 2019. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[20] Md Shafiqul Islam and Sol Lim. Vibrotactile Feedback in Virtual Motor Learning: A Systematic Review. *Applied Ergonomics*, 101, May 2022.

[21] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, November 1998.

[22] Aelee Kim, Minha Chang, Yeseul Choi, Sohyeon Jeon, and Kyoungmin Lee. The Effect of Immersion on Emotional Responses to Film Viewing in a Virtual Environment. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 601–602, March 2018.

[23] Bruno Alexandre Krinski, Daniel Vitor Ruiz, Guilherme Zasyeki Machado, and Eduardo Todt. Masking Salient Object Detection, a Mask Region-Based Convolutional Neural Network Analysis for Segmentation of Salient Objects. In *2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE)*, pages 55–60, October 2019.

[24] Zhuoran Li, Shiqi Zhang, Muhammad Shahid Anwar, and Jing Wang. Applicability Analysis on Three Interaction Paradigms in Immersive VR Environment. In *2018 International Conference on Virtual Reality and Visualization (ICVRV)*, pages 82–85, October 2018.

[25] Tie Liu, Jian Sun, Nan-Ning Zheng, Xiaoou Tang, and Heung-Yeung Shum. Learning to Detect A Salient Object. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[26] Omar López Chávez, Luis-Felipe Rodríguez, and J. Octavio Gutierrez-Garcia. A comparative case study of 2D, 3D and immersive-virtual-reality applications for healthcare education. *International Journal of Medical Informatics*, 141, September 2020.

[27] F. Orabona, G. Metta, and G. Sandini. Object-based Visual Attention: A Model for a Behaving Robot. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, pages 89–89, September 2005.

[28] Robert J. Peters and Laurent Itti. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[29] Jaziar Radianti, Tim A. Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147:103778, April 2020.

[30] Eric D. Ragan, Doug A. Bowman, Regis Kopper, Cheryl Stinson, Siroberto Scerbo, and Ryan P. McMahan. Effects of Field of View and Visual Complexity on Virtual Reality Training Effectiveness for a Visual Scanning Task. *IEEE Transactions on Visualization and Computer Graphics*, 21(7):794–807, July 2015.

[31] Yashas Rai and Patrick Le Callet. Chapter 3 - Visual attention, visual salience, and perceived interest in multimedia applications. In Rama Chellappa and Sergios Theodoridis, editors, *Academic Press Library in Signal Processing, Volume 6*, pages 113–161. Academic Press, January 2018.

[32] Nicolas Riche, Matthieu Duvinage, Matei Mancas, Bernard Gosselin, and Thierry Dutoit. Saliency and Human Fixations: State-of-the-Art and Study of Comparison Metrics. In *2013 IEEE International Conference on Computer Vision*, pages 1153–1160, December 2013.

[33] Radiah Rivu, Ville Mäkelä, Mariam Hassib, Yomna Abdelrahman, and Florian Alt. Exploring How Saliency Affects Attention in Virtual Reality. In *Human-Computer Interaction – INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30 – September 3, 2021, Proceedings, Part V*, pages 147–155, Berlin, Heidelberg, August 2021. Springer-Verlag.

[34] Ruth Rosenholtz, Jie Huang, and Krista Ehinger. Rethinking the Role of Top-Down Attention in Vision: Effects Attributable to a Lossy Representation in Peripheral Vision. *Frontiers in Psychology*, 3:13, February 2012.

[35] Dario D. Salvucci and Joseph H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, pages 71–78, New York, NY, USA, November 2000. Association for Computing Machinery.

[36] Frederick Shic, Katarzyna Chawarska, and Brian Scassellati. The amorphous fixation measure revisited: With applications to autism. In *30th Annual Meeting of the Cognitive Science Society*, pages 1–6, 2008.

[37] Valerie Shute, Seyedahmad Rahimi, and Benjamin Emihovich. Assessment for Learning in Immersive Environments. In Dejian Liu, Chris Dede, Ronghuai Huang, and John Richards, editors, *Virtual, Augmented, and Mixed Realities in Education*. Springer, Singapore, 2017.

[38] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. Saliency in VR: How Do People Explore Virtual Environments? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1633–1642, April 2018. Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[39] M. Speicher, C. Rosenberg, D. Degraen, F. Daiber, and A. Krüger. Exploring Visual Guidance in 360-degree Videos. In *TVX - Proc. ACM Int. Conf. Interact. Exp. TV Online Video*, pages 1–12. Association for Computing Machinery, Inc, 2019. Journal Abbreviation: TVX - Proc. ACM Int. Conf. Interact. Exp. TV Online Video.

[40] Emma E. M. Stewart, Matteo Valsecchi, and Alexander C. Schütz. A review of interactions between peripheral and foveal vision. *Journal of Vision*, 20(12):2, November 2020.

[41] Bob G. Witmer and Michael J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence*, 7(3):225–240, June 1998.

[42] André Zenner, Kristin Ullmann, Oscar Ariza, Frank Steinicke, and Antonio Krüger. Induce a Blink of the Eye: Evaluating Techniques for Triggering Eye Blinks in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pages 1–12, New York, NY, USA, April 2023. Association for Computing Machinery.

Appendix A

Literature Review

This chapter will describe, in detail, the methodology used to collect and select papers to be referenced and analyzed in later stages of this dissertation, supplemented by the results of each phase. Section A.1 introduces the method being used and provides the reasoning behind its choice over other options. Section A.2 focuses on the logic behind the creation of the search queries, as well as the results obtained by querying each database. Section A.3 sheds light into the process of choosing which papers are most relevant for the task at hand and why, providing a set of factors that contribute to that decision. Finally, section A.4 refers to the search and collection of more relevant papers through the investigation of the bibliography of each previously selected paper, complementing the group of papers obtained in the aforementioned steps.

A.1 Methodology

Due to the interdisciplinary and open-ended nature of the problem, given that it is related to the rapidly growing field of Virtual Reality, a mixed studies review was employed for this task. Following a systematic review type would be a mistake, since its restricted scope would hinder the possibility of including important papers, while adopting a mixed studies review allows for a more flexible approach that can gather information from multiple different sources. As an example, using this type of review, it is possible to combine qualitative studies around user attention and quantitative studies regarding task completion rates, which can provide a broader perspective and drive more complete and informed conclusions.

A.2 Search Strings & Collection

Before collecting papers, it is paramount that appropriate search strings are defined to limit the scope of the results to potentially relevant papers, with appropriate thematic keywords. Four search strings were initially created:

- **SS1**: attention AND virtual reality AND salienc?
- **SS2**: performance AND virtual environment AND salienc?

- **SS3**: attention AND virtual reality AND salienc? AND object AND NOT object detection
- **SS4**: salienc? AND virtual environment AND task performance AND immersive AND measur*

After defining the search strings, these were converted into search queries for three different databases: IEEExplore, Scopus and ACM, to ensure a thorough and complete collection of the maximum amount of relevant papers. The full search queries for every database are available in Appendix B.1.

However, after querying the first database (IEEExplore), these were the number of papers obtained:

- **SS1**: 1266 results
- **SS2**: 1908 results
- **SS3**: 105 results
- **SS4**: 95 results

It is very clear that exclusively searching in full-text content yields too many results to properly analyze, so a different approach for the first two queries was adopted: targeting less fields per keyword. The results for each query and each database can be found in the table below.

Table A.1: Search Query Results

Search String	**IEEExplore**	**Scopus**	**ACM**
SS1	484	132	277
SS2	327	317	102
SS3	105	27	15
SS4	95	43	41

It is important to note that the results of these refined queries already exclude any content that could not be saved to the utilized reference manager due to unauthorized access, so the initial numbers when running each query can be slightly different from those specified above.

The theory behind the resulting search strings was based on an analysis of relevant papers cited in Chapter 1 and identifying common keywords between them that could lead to the discovery of other appropriate content. Furthermore, they were separated into two categories: two queries geared towards answering the first research question (SS1 and SS3) and the other two toward the second research question (SS2 and SS4). The use of the question mark in "salienc?" stems from the two possible forms of writing the word: "saliency" and "salience", ensuring that both are caught by the search queries. Given that Virtual Reality is a recent and growing field, all of the results were filtered on query-time to only include papers from 2017 and beyond, in an attempt to avoid very dated technologies and methods. Using Zotero's duplicate merging capabilities, 1362 papers remained after this step.

A.3 Screening and Selection

Upon analyzing each search query, it is clear that SS3 and SS4 are more advanced versions of SS1 and SS2, respectively. This approach was adopted to simplify the selection process. First, evaluate SS3 and SS4, which result in a subset of papers obtained from SS1 and SS2, since these papers are more likely to be relevant due to the presence of a greater number of keywords and are thus analyzed with an increased focus and depth. Then, look at the results from SS1 and SS2 that were not captured by the previous search queries, which have a greater chance of being irrelevant but must be evaluated nonetheless. This two-step process is done to ensure that, even though a greater number of papers was collected from the phase described in A.2, their analysis can be effective while still preventing over-engineered search queries from not capturing important papers.

With the large set of papers collected, the next step is to select them based on how relevant they are. Each paper is filtered given a set of inclusion and exclusion criteria and added to a new subset of the original group. The criteria followed are presented below. Some may seem redundant, such as limiting the publication year, as it was already filtered by each query, but these will be useful in the step described in Section A.4.

Inclusion/Exclusion Criteria

- Date of publication since 2017
- Published in English
- Full-text access
- Not duplicated
- Conference Paper, Article, Book

In this reduced set, every paper is analyzed with greater care until it is considered relevant or irrelevant, by reading these fields in order:

- Title
- Keywords
- Abstract
- Conclusions
- Full-text content

The relevance of each paper is judged according to the following set of selection criteria:

Selection Criteria

- In the field of Virtual Reality
- Not focused on other fields of Extended Reality (Augmented Reality, Mixed Reality) or traditional 2D/3D technologies
- Studies focusing on user attention OR object saliency with user testing

- In the presence of task performance analysis, methodology and measures are clear
- Tests with at least 10 participants

After following this procedure, 23 papers were considered relevant.

A.4 Reference Chaining

"Reference Chaining" is the arbitrary name chosen for the process of scanning each selected paper's bibliography in the search of additional relevant content. This step is useful to capture a handful of papers that might have been excluded by the search queries, while still being considered important after applying the same selection procedure described in section A.3. 5 papers were added to the set after this step.

A.5 Conclusion

At the end of the literature review, a group of 33 papers, including the ones added prior to the literature review, was deemed relevant to serve as a knowledge base for the remainder of the thesis, without ruling out the possibility of small additions and removal of papers in the future.

Appendix B

Literature Review - Search Queries

B.1 Detailed Search Queries

The next four pages showcase the search queries used during the literature review process written in chapter 2. Each one was fine-tuned according to the number of results returned, avoiding too many papers. It is important to note that, even though only Scopus has the publication date filter mentioned explicitly, the remaining databases were queried using a "publication date" slider set to 2017-2024.

Table B.1: Detailed Search Queries for Each Database - SS1

ID	Source	Search String
SS1 (initial)	IEEExplore	("Full Text Only":attention) AND ("Full Text Only":virtual reality) AND ("Full Text Only":salienc?)
	Scopus	Only IEEExplore was explored at this stage
	ACM	Only IEEExplore was explored at this stage
SS1 (refined)	IEEExplore	("All Metadata":salienc?) AND ("Full Text & Metadata":virtual environment) AND ("Full Text & Metadata":task performance)
	Scopus	(TITLE-ABS-KEY (salienc?) AND ALL (virtual AND environment) AND ALL (task AND performance)) AND PUBYEAR > 2016
	ACM	(Title:(salienc?) OR Abstract:(salienc?) OR Keyword:(salienc?)) AND (Title:(virtual environment) OR Abstract:(virtual environment) OR Keyword:(virtual environment) OR Fulltext:(virtual environment)) AND (Title:(task performance) OR Abstract:(task performance) OR Keyword:(task performance) OR Fulltext:(task performance))

Table B.2: Detailed Search Queries for Each Database - SS2

ID	Source	Search String
SS2 (initial)	IEEExplore	("Full Text Only":performance) AND ("Full Text Only":virtual environment) AND ("Full Text Only":salienc?)
	Scopus	Only IEEExplore was explored at this stage
	ACM	Only IEEExplore was explored at this stage
SS2 (refined)	IEEExplore	("All Metadata":salienc?) AND ("Full Text & Metadata":virtual environment) AND ("Full Text & Metadata":task performance)
	Scopus	(TITLE-ABS-KEY (salienc?) AND ALL (virtual AND environment) AND ALL (task AND performance)) AND PUBYEAR > 2016
	ACM	(Title:(salienc?) OR Abstract:(salienc?) OR Keyword:(salienc?)) AND (Title:(virtual environment) OR Abstract:(virtual environment) OR Keyword:(virtual environment) OR Fulltext:(virtual environment)) AND (Title:(task performance) OR Abstract:(task performance) OR Keyword:(task performance) OR Fulltext:(task performance))

Table B.3: Detailed Search Queries for Each Database - SS3

ID	Source	Search String
SS3	IEEExplore	("All Metadata":attention) AND ("Full Text Only":salienc?) AND ("Full Text Only":object) AND ("Full Text Only":virtual reality) NOT ("Full Text & Metadata":object detection)
	Scopus	(TITLE-ABS-KEY (attention) AND TITLE-ABS-KEY (virtual AND reality) AND TITLE-ABS-KEY (salienc?) AND TITLE-ABS-KEY (object) AND NOT TITLE-ABS-KEY (object AND detection)) AND PUBYEAR > 2016
	ACM	(Title:(attention) OR Abstract:(attention) OR Keyword:(attention)) AND (Title:(virtual reality) OR Abstract:(virtual reality) OR Keyword:(virtual reality) OR Fulltext:(virtual reality)) AND (Title:(salienc?) OR Abstract:(salienc?) OR Keyword:(salienc?) OR Fulltext:(salienc?)) AND (Title:(object) OR Abstract:(object) OR Keyword:(object) OR Fulltext:(object)) AND NOT (Title:(object detection) OR Abstract:(object detection) OR Keyword:(object detection) OR Fulltext:(object detection))

Table B.4: Detailed Search Queries for Each Database - SS4

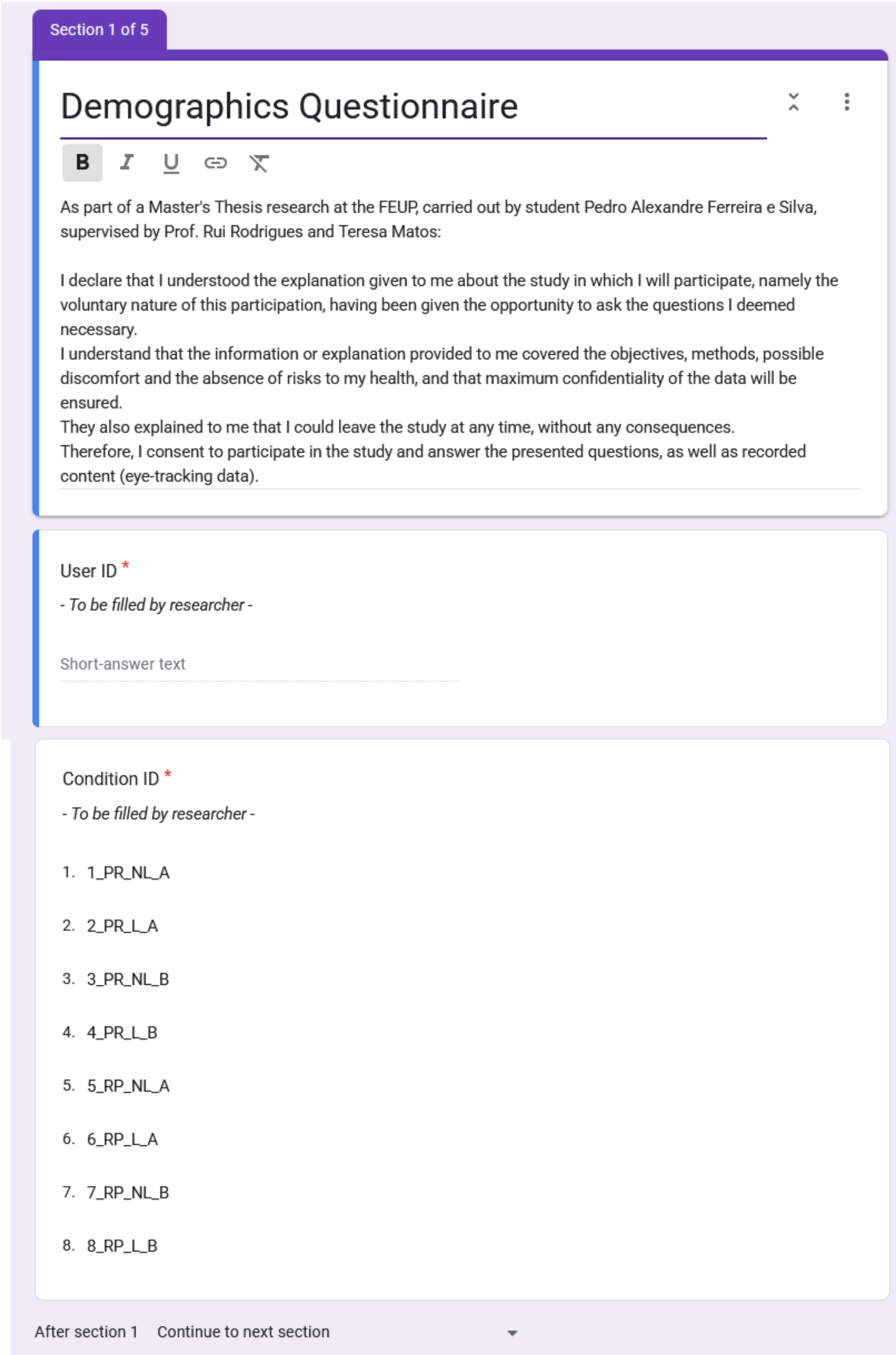
ID	**Source**	**Search String**
SS4	IEEExplore	("All Metadata":salienc?) AND ("Full Text & Metadata":virtual environment) AND ("Full Text & Metadata":task performance) AND ("Full Text & Metadata":immersive) AND ("Full Text & Metadata":measur*)
	Scopus	(TITLE-ABS-KEY (salienc?) AND ALL (virtual AND environment) AND ALL (task AND performance) ALL (immersive) AND ALL (measur*)) AND PUBYEAR > 2016
	ACM	(Title:(salienc?) OR Abstract:(salienc?) OR Keyword:(salienc?)) AND (Title:(virtual environment) OR Abstract:(virtual environment) OR Keyword:(virtual environment) OR Fulltext:(virtual environment)) AND (Title:(task performance) OR Abstract:(task performance) OR Keyword:(task performance) OR Fulltext:(task performance)) AND (Title:(immersive) OR Abstract:(immersive) OR Keyword:(immersive) OR Fulltext:(immersive)) AND (Title:(measur*) OR Abstract:(measur*) OR Keyword:(measur*) OR Fulltext:(measur*))

Appendix C

Demographics and Comparison Questionnaires

We provide the entirety of the demographics and comparison questionnaires in the next pages.

C.1 Demographics Questionnaire


Section 1 of 5

Demographics Questionnaire

As part of a Master's Thesis research at the FEUP, carried out by student Pedro Alexandre Ferreira e Silva, supervised by Prof. Rui Rodrigues and Teresa Matos:

I declare that I understood the explanation given to me about the study in which I will participate, namely the voluntary nature of this participation, having been given the opportunity to ask the questions I deemed necessary.
I understand that the information or explanation provided to me covered the objectives, methods, possible discomfort and the absence of risks to my health, and that maximum confidentiality of the data will be ensured.
They also explained to me that I could leave the study at any time, without any consequences.
Therefore, I consent to participate in the study and answer the presented questions, as well as recorded content (eye-tracking data).

User ID *

- To be filled by researcher -

Short-answer text

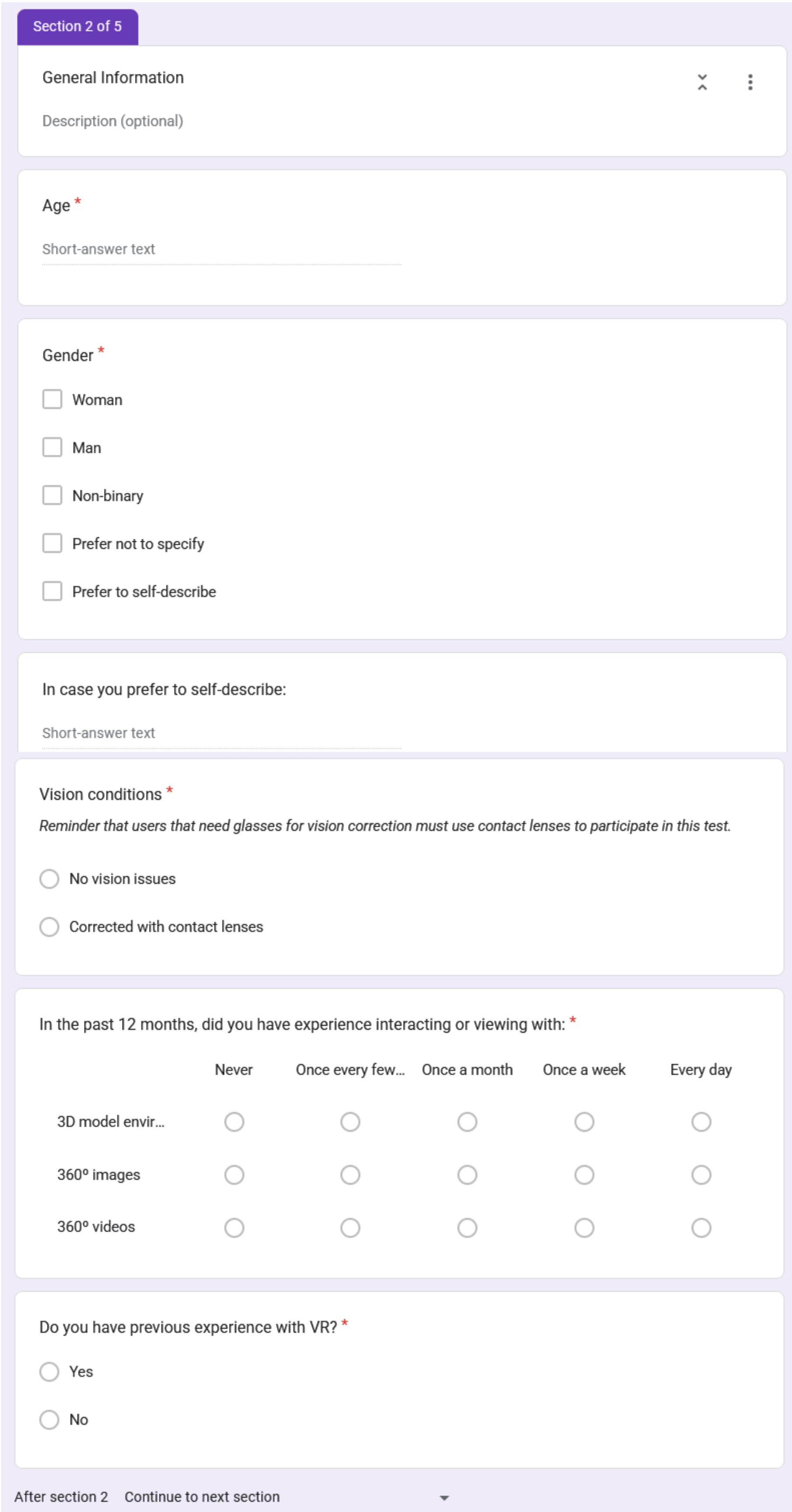
Condition ID *

- To be filled by researcher -

1. 1_PR_NL_A
2. 2_PR_L_A
3. 3_PR_NL_B
4. 4_PR_L_B
5. 5_RP_NL_A
6. 6_RP_L_A
7. 7_RP_NL_B
8. 8_RP_L_B

After section 1 Continue to next section

Figure C.1: Page 1 of the demographics questionnaire



Section 2 of 5

General Information

Description (optional)

Age *

Short-answer text

Gender *

- Woman
- Man
- Non-binary
- Prefer not to specify
- Prefer to self-describe

In case you prefer to self-describe:

Short-answer text

Vision conditions *

Reminder that users that need glasses for vision correction must use contact lenses to participate in this test.

- No vision issues
- Corrected with contact lenses

In the past 12 months, did you have experience interacting or viewing with: *

	Never	Once every few...	Once a month	Once a week	Every day
3D model envir...	○	○	○	○	○
360º images	○	○	○	○	○
360º videos	○	○	○	○	○

Do you have previous experience with VR? *

- Yes
- No

After section 2 Continue to next section

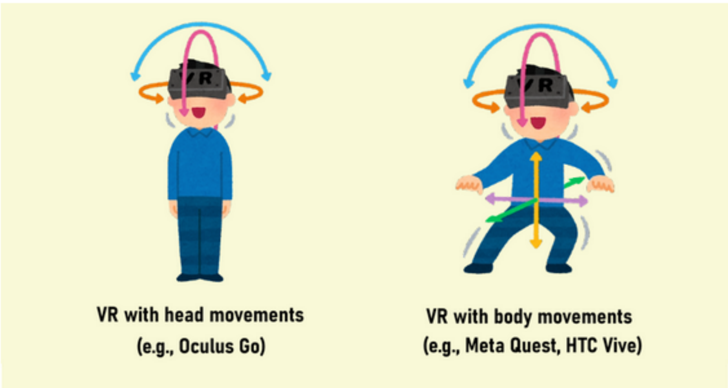
Figure C.2: Page 2 of the demographics questionnaire

Section 3 of 5

VR Experience Details

If you answered "No" to the previous question, please select "Never" in every question on this page, or leave the answer blank if it's not mandatory.

In the past 12 months, how frequently did you use the following types of VR headsets? *



	Never	Once every few...	Once a month	Once a week	Every day
Smartphone in...	○	○	○	○	○
VR headset tha...	○	○	○	○	○
VR headset tha...	○	○	○	○	○

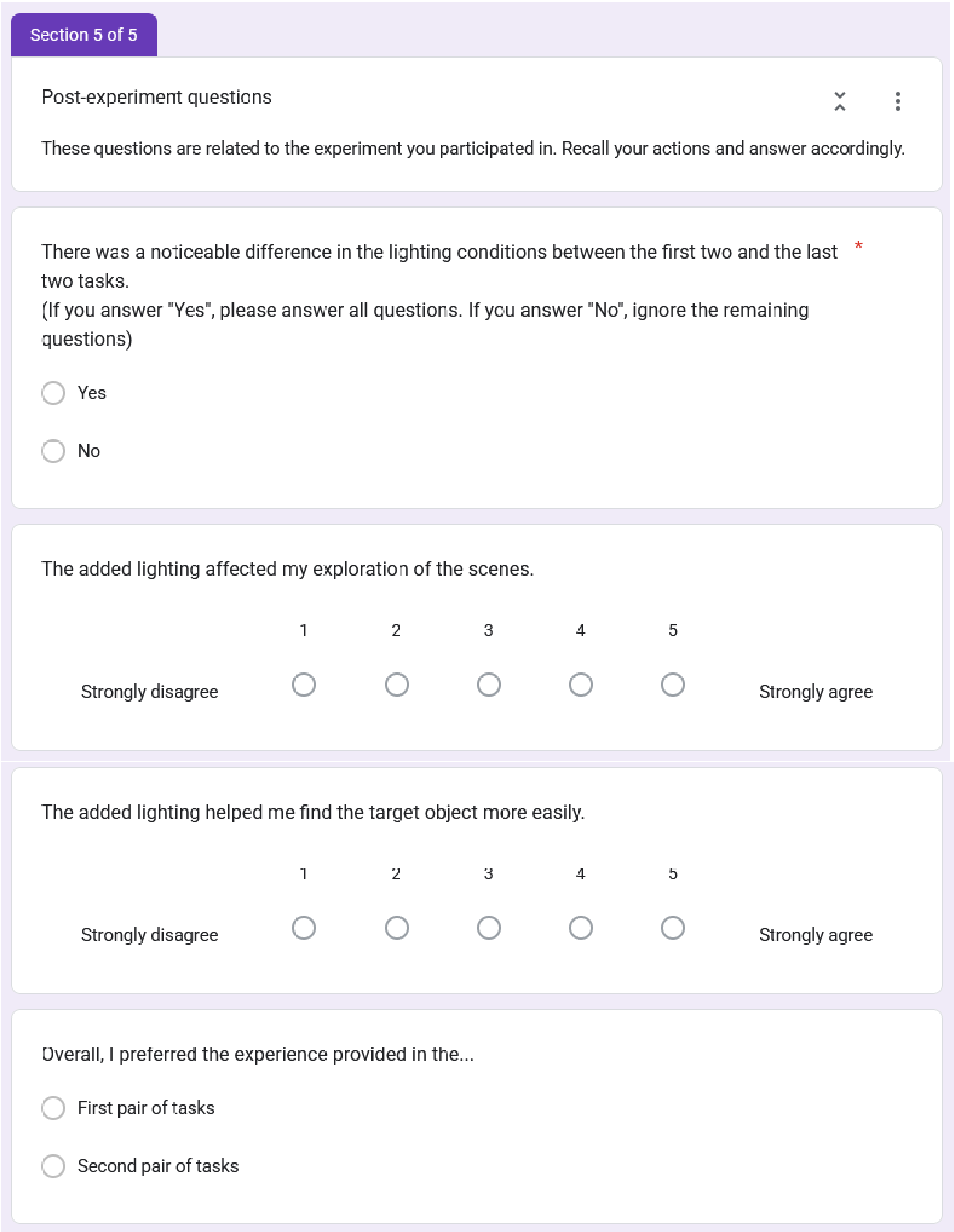
What have you used VR for?

- ☐ 360º images/videos
- ☐ Games
- ☐ Simulation
- ☐ For research and/or development
- ☐ As a participant of user studies (like this one)
- ☐ Training/Learning
- ☐ Other...

After section 3 Continue to next section

Figure C.3: Page 3 of the demographics questionnaire

C.2 Comparison Questionnaire


Section 5 of 5

Post-experiment questions

These questions are related to the experiment you participated in. Recall your actions and answer accordingly.

There was a noticeable difference in the lighting conditions between the first two and the last two tasks. *
(If you answer "Yes", please answer all questions. If you answer "No", ignore the remaining questions)

- Yes
- No

The added lighting affected my exploration of the scenes.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

The added lighting helped me find the target object more easily.

	1	2	3	4	5	
Strongly disagree	○	○	○	○	○	Strongly agree

Overall, I preferred the experience provided in the...

- First pair of tasks
- Second pair of tasks

Figure C.4: Questions from the comparison questionnaire