TÉCNICO LISBOA

EyeDepth: Hands-free Selection in VR

David Martins Correia

Thesis to obtain the Master of Science Degree in

Telecommunications and Informatics Engineering

Supervisors: Prof. Joaquim Jorge
Prof. Anderson Maciel

October 2025

Declaration
I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

I would like to thank my parents for their friendship, encouragement and caring over all these years, for always being there for me through thick and thin and without whom this project would not be possible. I would also like to thank my grandparents, aunts, uncles, and cousins for their understanding and support throughout all these years.

I would also like to acknowledge my dissertation supervisors Prof. Joaquim Jorge and Prof. Anderson Maciel. I would also like to thank Professor Pedro Belchior for his insight, support and sharing of knowledge that has made this Thesis possible.

Last but not least, to all my friends and colleagues who helped me grow as a person and who were always there for me during the good and bad times in my life. Thank you.

To each and every one of you – Thank you.

Abstract

With the rapid evolution of virtual reality technology, eye tracking has become a key enabler of natural and intuitive interaction, offering the potential to replace traditional input devices and enhance user immersion. Among its many applications, gaze depth estimation through vergence presents a promising path toward more precise and responsive virtual interactions. Accurate interpretation of user intent based on gaze depth can significantly improve object selection, focus control, and depth-based rendering, enabling more seamless and engaging experiences.

However, vergence estimation remains challenged by hardware noise, calibration inconsistencies, and human factors such as fixation instability and eye jitter. This thesis addresses these challenges by developing and evaluating vergence-based gaze interaction techniques that emphasize robustness and usability in cluttered 3D environments. The work was conducted in three stages: assessing vergence reliability, testing gaze confirmation methods, and designing depth-aware selection techniques. Each stage built upon empirical insights from the previous one, progressively forming a complete gaze interaction framework.

The proposed techniques were tested in controlled but visually complex virtual environments, highlighting how depth-aware strategies can mitigate gaze inaccuracies while maintaining comfort and efficiency. Empirical evaluations demonstrated robust depth estimation, higher selection accuracy compared to baseline gaze input, and improved user comfort through reduced gaze instability and task-induced fatigue. The findings contribute to the design of stable and precise gaze-based interfaces that operate effectively even under depth ambiguity, laying the groundwork for future adaptive and dynamic

gaze interaction systems.

Keywords

Vergence; Eye Tracking; Gaze Depth; Virtual Reality; Fixations; Head-Mounted Displays;

Resumo

Com a rápida evolução da tecnologia de realidade virtual, o rastreamento ocular tornou-se um elemento-chave para promover interações naturais e intuitivas, possibilitando a substituição de dispositivos de entrada tradicionais e melhorando a imersão do utilizador. Entre as suas aplicações, a estimativa da profundidade através da vergência revela-se um caminho promissor para alcançar interações mais precisas e responsivas. Uma interpretação exata da intenção do utilizador pode melhorar significativamente a seleção de objetos e apoiar a renderização baseada em profundidade, permitindo experiências mais contínuas e envolventes.

Apesar deste potencial, persistem desafios relevantes, nomeadamente ruído durante rastreamento ocular, inconsistências de calibração e fatores humanos como instabilidade do olhar e micro-movimentos oculares. Este trabalho procura enfrentar essas limitações ao investigar técnicas de interação baseadas na vergência, equilibrando precisão, conforto e robustez em ambientes virtuais visualmente complexos. Foram conduzidos três estudos que analisaram separadamente a fiabilidade da vergência, métodos de confirmação e a eficácia de técnicas de seleção em cenários tridimensionais. Esta abordagem iterativa resultou em soluções capazes de manter precisão e usabilidade mesmo sob elevada ambiguidade visual.

As avaliações empíricas demonstraram uma estimativa de profundidade mais robusta, maior precisão na seleção em comparação com métodos de base e melhorias no conforto do utilizador graças à redução da instabilidade do olhar e da fadiga induzida pela tarefa. Os resultados contribuem para o avanço das interfaces de interação baseadas no olhar, mostrando como modelos sensíveis à profundidade podem mitigar imprecisões, reduzir o esforço visual e promover uma experiência de realidade

virtual mais natural e imersiva.

Palavras Chave

Vergência; Rastreamento Ocular; Profundidade do Olhar; Realidade Virtual; Fixações; Dispositivos Montados na Cabeça

Contents

List of Figures

List of Tables

Acronyms

AR	Augmented Reality
DL	Deep Learning
DoF	Depth of Field
EVA	Eye Vergence Angle
FC	Floor and Ceiling
XR	Extended Reality
HCI	Human–Computer Interaction
HMD	Head-Mounted display
IR	Infrared
IPD	Interpupillary Distance
MAE	Mean Absolute Error
NFC	No Floor and Ceiling
RMSE	Root Mean Squared Error
RQ	Research Question
VR	Virtual Reality

1

Introduction

Contents

In recent years, the increasing accessibility of affordable Virtual Reality (VR) headsets equipped with integrated eye-tracking technology has begun to redefine the landscape of Extended Reality (XR) interfaces. Devices such as Apple Vision Pro [9] have demonstrated that gaze-based interaction is not only practical but also a central input modality for next-generation immersive systems, offering unprecedented levels of precision and presence. Eye tracking enables natural and intuitive gaze-based control, reducing dependence on traditional input devices such as controllers and hand tracking [10]. Nevertheless, despite these advances, current implementations continue to face notable limitations, particularly regarding reliable depth perception, stable object targeting, and sustained interaction comfort [11–13].

1.1 Motivation

Accurate gaze-based object detection in VR remains a challenging problem, particularly in complex or cluttered environments where objects at multiple depths overlap or occupy similar visual space. Under such conditions, the ambiguity between the depth layers can substantially degrade the interaction performance. These issues arise not only from hardware constraints, such as tracking noise or limited sampling frequency, but also from human factors, including natural micro-saccades, fixation instability [14], and individual differences in oculomotor control [15], producing inconsistent targeting results. Although head or body movement can sometimes resolve these ambiguities by altering the user's perspective, this solution is not universally practical, especially for seated users or users with physical impairments, and can also be limited by the physical space available around the user during virtual interaction.

Various approaches have been proposed to mitigate these challenges. Some rely on external tracking systems or multi-camera setups to enhance spatial accuracy in natural environments [16], although these often increase complexity and cost. More accessible approaches focus on improving gaze-based interaction directly within standard Head-Mounted displays (HMDs), leveraging vergence, the inward rotation of the eyes during fixation, as a natural indicator of gaze depth. Studies inspired by Wibirama and Hamamoto [17] and Bourke [18] have demonstrated that vergence can serve as a key mechanism for supporting 3D selection tasks, although its practical reliability remains influenced by distance, calibration stability, and individual visual characteristics.

This thesis builds upon these foundations by investigating vergence-based gaze interaction within visually cluttered virtual environments, emphasizing the balance between precision, comfort, and robustness under varying visual and interaction conditions. Specifically, it explores how vergence data can be integrated into a complete gaze-based interface by developing and evaluating multiple techniques for both selection and confirmation, each addressing the limitations identified in previous studies.

1.2 Research Questions and Aims

Motivated by these goals, the investigation aims to answer the following research questions: **Research Question (RQ) 1.** How can vergence techniques be refined to improve gaze depth estimation in cluttered and immersive VR environments?; **RQ 2.** How can interaction techniques mitigate inaccurate eye-tracking data and improve the reliability and precision of gaze-based systems in VR?; **RQ 3).** How do tracking inaccuracies and task complexity contribute to visual fatigue, and what design strategies can mitigate these effects in VR systems?.

To achieve this, the research was organized into three complementary stages: vergence evaluation, confirmation testing, and selection testing. Each stage isolated a specific component of the interac-

tion process, allowing for a focused quantitative analysis without cross-interference between variables. This modular approach ensured that design decisions were empirically grounded, progressing from evaluating raw vergence reliability to creating comprehensive selection techniques capable of operating effectively in overlapping, densely populated 3D scenes.

The goal of this research is to develop a robust vergence-based interaction interface that maintains accuracy and usability even under visual clutter and depth ambiguity. The resulting interface demonstrates that vergence can support consistently accurate depth estimation and reliable object selection, offering resilience to tracking noise and visual complexity while maintaining user comfort across varied tasks and viewing conditions. This offers a solid foundation for future adaptive extensions.

Ultimately, this thesis contributes to the advancement of gaze-based interaction design by (1) refining the understanding of vergence limitations and error tolerance in complex VR contexts, (2) identifying efficient and comfortable confirmation methods for gaze-based selection, and (3) demonstrating how depth-aware interaction models can enhance usability in cluttered 3D environments. Collectively, these findings move toward the broader vision of seamless, controller-free interaction in immersive virtual systems.

1.3 Structure of the Document

The remainder of this work is organized as follows. Chapter 2 introduces the fundamental concepts necessary for understanding this thesis. Chapter 3 reviews key related works and themes, providing a comprehensive overview of the current state of the art. Chapter 4 presents the study conducted to assess the quality and reliability of the vergence algorithm. Chapter 5 describes the development of the gaze-based interaction techniques and the methods used to evaluate them. Chapter 6 discusses the findings of all experiments and their implications in a unified context. Finally, Chapter 7 concludes the thesis by summarizing the main contributions, acknowledging its limitations, and suggesting directions for future research.

2

Key Concepts

Contents

This chapter presents several recurring concepts that underpin this thesis. Beginning with eye-tracking, I discuss how these concepts relate to visual perception, depth in VR, and related ocular measurements.

2.1 Eye-Tracking

Eye tracking involves measuring and interpreting where and how a person looks at something, making it instrumental for understanding visual attention and behavior. Eye trackers typically fall into two categories: (a) those measuring the eye's position relative to the head and (b) those determining the "point of regard," or where the eye is directed in space. The second approach is most commonly used in human-computer interaction and visual attention studies.

Modern systems primarily rely on video-based corneal-reflection technology, which uses a directed light source and cameras to detect features like the pupil and the iris-sclera boundary for precise, non-invasive measurements [19]. As illustrated in Figure 2.1, these systems calculate gaze direction by comparing the corneal reflection and pupil positions, enabling accurate tracking despite minor head movements.

Eye movements play a central role in human vision, allowing precise focusing on points of interest and rapid shifts of attention. This underscores the importance of fixations and saccades, which will be addressed in subsequent sections.

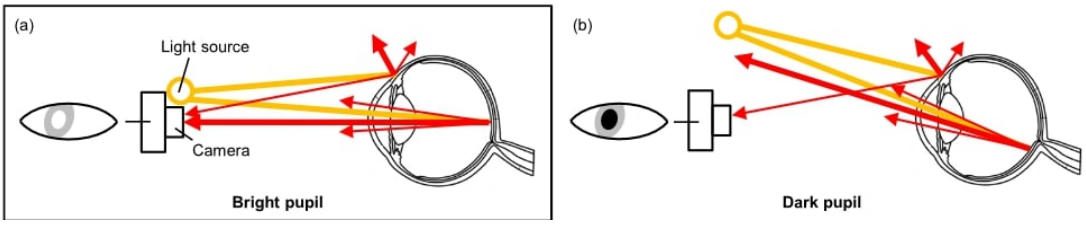


Figure 2.1: Illustration of video-based eye tracking using corneal reflection to estimate gaze direction [1]

2.2 Vergence

Vergence refers to the coordinated movement of the eyes to focus on an object at a particular depth. In VR, vergence is essential for accurate depth perception, as the technology must simulate 3D environments with objects appearing at varying distances. However, VR headsets often cannot perfectly replicate natural vergence, leading to vergence-accommodation conflicts and potential discomfort.

Accurate eye-tracking technologies, including pupil-size monitoring and corneal reflection, enable VR systems to estimate vergence and adjust stereoscopic images to align with users' focus, enhancing realism and reducing visual fatigue [5, 20]. Combining vergence data with Interpupillary Distance (IPD) measurements improves depth perception, making VR experiences more comfortable and immersive [2, 21]. Figure 2.2 from the 2022 study demonstrates how vergence measurements, including the Eye Vergence Angle (EVA), enhance 3D gaze estimation and depth perception in virtual environments.

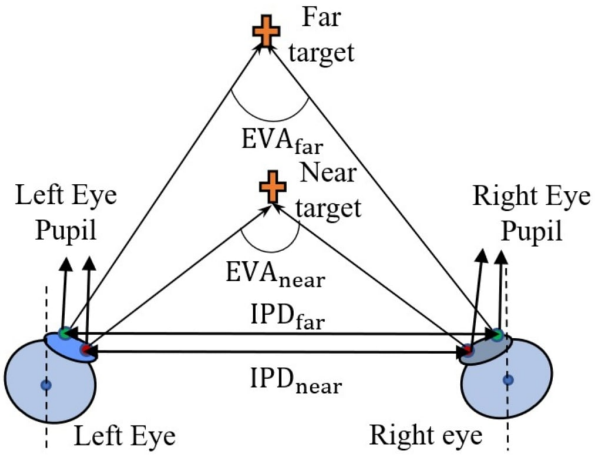


Figure 2.2: Vergence measurement in VR using IPD and EVA [2]

2.3 Interpupillary Distance

IPD is the separation between the centers of a person's pupils, typically measured in millimeters. Properly aligning the headset's lenses with the user's IPD ensures optimal image clarity and reduces discomfort. The variability of human IPD, ranging from roughly 50 mm to 75 mm, poses challenges for designing VR systems that accommodate a wide range of users. Some studies [22, 23] highlight its importance in achieving visual comfort and an immersive VR experience.

2.4 Foveal Vision and Parafoveal Vision

Foveal vision is the acute central vision mediated by the fovea, a small retinal area with the highest density of cone cells. Crucial for reading and object recognition, the fovea spans less than 1% of the visual field yet accounts for a significant portion of visual acuity [24]. Surrounding the fovea is the parafoveal region, providing broader peripheral support to detect large objects and movement [25].

Together, these regions enable the integration of fine detail and peripheral awareness, vital for navigation and maintaining situational awareness in tasks such as reading or scene searching (see Figure 2.3).

2.5 Depth of field

Depth of Field (DoF) refers to the range of distances within a visual scene that appear in sharp focus, influenced by factors such as pupil size and lens accommodation. Smaller pupils (in bright light) provide a greater DoF, enabling a clearer focus across a wider range, whereas larger pupils (in dim light) result in a shallower one. This change in depth of field can be seen in photography (see Figure 2.4), although

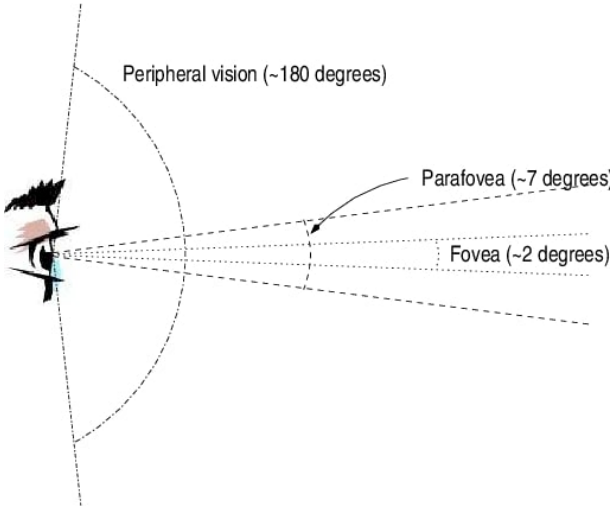


Figure 2.3: The three regions of visual perception: fovea, parafovea and peripheral vision [3]

for the human eye, the difference is not as pronounced. Simulating DoF in VR enhances realism by mimicking how objects naturally transition between sharpness and blur as the user's focus shifts [24].

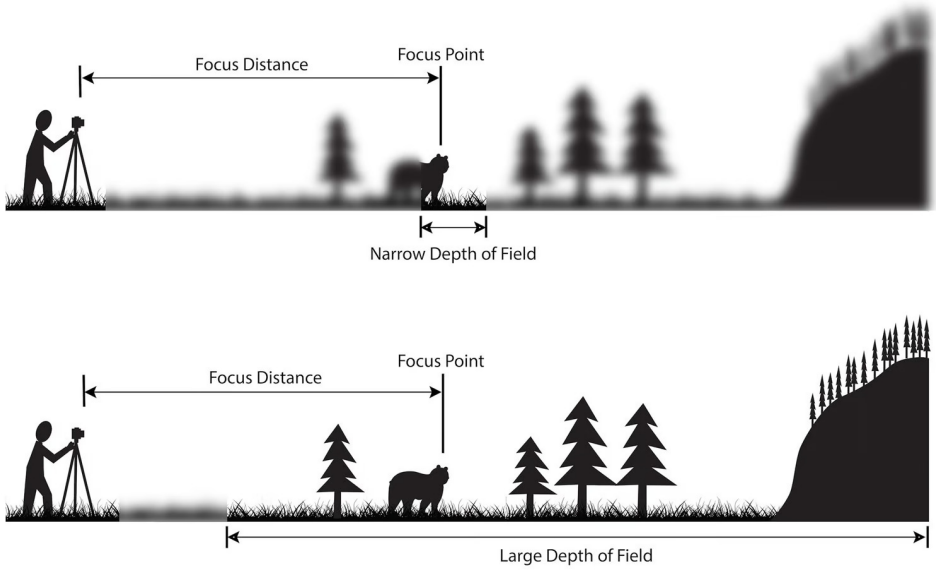


Figure 2.4: Narrow and Large Depth of Field [4]

2.6 Fixations

Fixations are the brief periods during which the eyes pause to gather visual information, typically lasting 200 to 400 milliseconds. During these pauses, the fovea focuses on specific points of interest, enabling

detailed perception, such as reading or object recognition [26]. In VR, eye-tracking systems analyze these fixation data to improve user interactions and enhance immersion [19].

2.7 Saccades

Saccades are rapid eye movements that shift the eyes between points of interest, enabling users to explore a visual scene efficiently. These movements typically last only a few milliseconds but are vital for reading, scanning the environment, and identifying objects. During saccades, the brain briefly suppresses visual input, instead focusing on the new fixation point after movement [27].

In VR, understanding saccadic movements is essential for optimizing user interfaces and interactions. VR systems can use collected eye-tracking data to evaluate how users perform saccadic shifts, helping developers refine the design and flow of virtual environments [28].

2.8 Eye Limitations/Vision Problems

Refractive errors, such as myopia, hyperopia, and astigmatism, are increasingly common and can alter how the eye appears to video-based tracking devices (pupil shape, corneal reflections), which, in turn, can reduce calibration quality and gaze accuracy in VR [29].

Users wearing glasses or contact lenses often cause additional reflections and occlusions that degrade the performance of eye-tracking devices. Many studies recommend design/testing protocols to handle these cases [30, 31].

3

Related Work

Contents

This chapter presents a review of the existing literature that provides the conceptual and technical foundation for the present research. Over the past decade, eye-tracking and gaze-based interaction in VR have undergone significant development, leading to substantial advances in gaze estimation, user modeling, and interaction design. These contributions have deepened our understanding of how users perceive, attend to, and manipulate virtual content through eye movements, establishing gaze as a viable input modality within immersive systems.

Despite these advances, several aspects of gaze-based interaction remain underexplored. Many existing studies address specific challenges, such as the accuracy of gaze estimation, fixation stability,

or eye–head coordination, without fully examining how these factors jointly affect interaction performance when dealing with dynamic 3D contexts. Furthermore, questions persist about the impact of visual fatigue, workload, and depth perception on the consistency and usability of gaze-based selection techniques. These open issues underscore the need for continued investigation into how gaze can be leveraged effectively and comfortably across varied interaction conditions.

For clarity and focus, the chapter is organized thematically. It begins with a discussion of the work by Duchowski [5], which lays important groundwork in 3D gaze estimation methods and real-time event detection. The subsequent sections then expand on related themes, including gaze tracking of moving targets, the integration of gaze data in deep learning systems, eye–head coordination, gaze depth estimation, fatigue effects in virtual environments, and interaction techniques. Together, these sections outline the current state of research in gaze-based VR interaction, emphasizing both the achievements that have shaped the field and the unresolved challenges that motivate the present work.

3.1 3D Gaze Estimation in VR: Methods and Applications

Accurate gaze estimation is fundamental to enabling immersive and intuitive interactions in virtual environments. Among the notable contributions in this field, the work by Andrew T. Duchowski [5] stands out for its integrated exploration of vergence estimation, continuous calibration, and real-time 3D event detection, three interrelated components that collectively advanced the precision and usability of XR eye-tracking systems.

Among these, the technique of vergence-based gaze depth estimation was of particular interest for this work, as it provides a means of determining where a user is focusing in 3D space by computing the intersection of gaze vectors from both eyes, rather than relying solely on gaze ray–object intersections (Figure 3.1). By leveraging the geometric relationship between both eyes, this approach enhances the perception of spatial relationships and improves accuracy in object selection and manipulation, directly addressing one of the most persistent challenges in virtual environments and enabling intuitive, precise interaction with objects at varying depths.

In parallimproved spatial consistency and al-time 3D event detection contributed to improving spatial consistency and the temporal understanding of user gaze behavior. Continuous calibration reduces spatial drift and error accumulation during prolonged sessions, while event detection extends gaze analysis beyond static fixations by incorporating eye–head coordination and motion dynamics.

Together, these three methods offer a strong foundation for the development of more natural and adaptive gaze-based interfaces. Although each addresses a specific problem, the combined insights provide valuable guidance for enhancing depth perception, calibration accuracy, and event recognition in immersive systems. The vergence-based depth estimation technique, in particular, serves as the

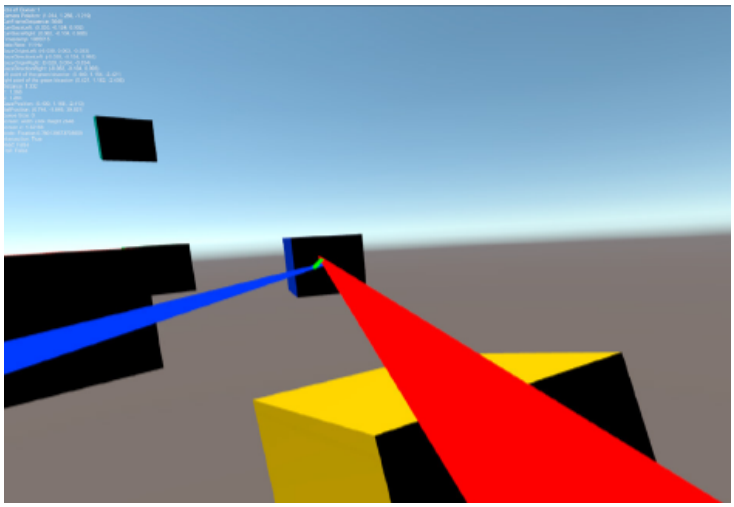


Figure 3.1: Vergence estimation in Unity 3D [5]

conceptual basis for this work, informing both the methodological and experimental design choices.

Building on these foundational methods, several studies have sought to refine or extend gaze depth estimation in VR. Traditional binocular approaches combine vergence and interpupillary distance to estimate focal depth, improving realism and interaction precision in virtual environments [2]. However, to reduce dependency on dual-eye tracking, alternative monocular techniques have emerged. For instance, Zhang et al. [32] introduced a vestibulo-ocular reflex-based method that integrates head movement with single-eye tracking, simplifying hardware requirements while maintaining acceptable accuracy.

Other contributions include Gaze3DFix [33], which models gaze convergence using an ellipsoidal bounding volume, enabling reliable depth estimation even with irregular or partially occluded targets. Similarly, GazeStereo3D [34], although developed for AR contexts, demonstrates that manipulating stereo disparity can enhance depth perception and consistency in 3D interactions, offering insights that are transferable to VR. More recently, McAnally et al. [20] revisited vergence dynamics, emphasizing the need for adaptive vergence models to handle rapidly changing focal points, an increasingly relevant challenge in interactive and dynamic 3D scenes.

Recent work continues to explore and refine the role of vergence in 3D gaze estimation. For instance, Monier et al. [35], conducted a pilot study on ocular vergence measurement in virtual reality, demonstrating that controlled VR environments can accurately reproduce the near- and far-focus conditions observed in real-world vision. Their results reaffirm the feasibility of using vergence as a reliable indicator of depth awareness, particularly under well-calibrated optical conditions. Complementarily, Wang et al. [36] proposed a geometrical model to mitigate the vergence–accommodation conflict (VAC), showing that compensating for binocular distortion can significantly improve spatial accuracy during gaze-based interaction.

Overall, these studies collectively illustrate the evolution of gaze estimation from discrete, planar

approaches toward dynamic, depth-aware systems that better reflect human visual behavior. Despite these advances, challenges remain in balancing real-time performance, accuracy, and comfort, particularly when integrating eye and head motion in complex virtual environments. Addressing these challenges remains central to improving the precision and intuitiveness of gaze-based interaction in future XR systems.

3.2 Non-static Targets Focusing Accuracy and How to Effectively Recognize Them

Accurately focusing on non-static targets in VR is essential for creating immersive and responsive user experiences. These tasks can be effectively addressed using eye-tracking technology, which allows systems to monitor and interpret user gaze behavior in real time. A key aspect of this process is the measurement of vergence, which is crucial for depth perception and focus adjustment. As users engage with dynamic objects in 3D spaces, effective recognition of these objects becomes just as critical. The ability to track a user's gaze directly influences interaction, especially in environments with varying depths and occlusions.

Recent studies have highlighted the importance of precise vergence estimation in VR environments, particularly when users interact with dynamic objects. Research on vergence matching has introduced interaction techniques that leverage motion correlation to select small targets in 3D spaces, thereby enhancing user interaction with moving elements [37]. Accurate gaze tracking is essential for identifying which objects are being focused on, a crucial step for tasks like object selection and manipulation.

Additionally, advancements in eye-tracking calibration methods have been shown to improve spatial accuracy and precision across various visual angles, further supporting effective focus on non-static targets [38]. These developments, combined with enhanced object recognition capabilities, underscore the critical role of accurate vergence measurement and calibration in enabling more seamless, natural interactions with dynamic objects in VR settings.

3.3 Applications of Eye Gaze in Deep Learning Systems

The role of gaze in Deep Learning (DL)-based models for virtual reality remains a topic of debate. While some studies argue for the inclusion of gaze signals, highlighting their potential to enhance model accuracy, others caution against their use due to dependence on contextual background features or the need for extensive task-specific labels. Research on gaze, such as the work by Suneeta [39], demonstrates that gaze-driven attention maps can guide DL models, improving performance by focusing on regions that align with human visual attention.

However, these methods often face challenges in VR environments, where the 3D structure of the scene is predefined, and the user's viewpoint is central to object targeting. In such cases, simpler geometric techniques, including ray casting or gaze intersection, are typically more efficient and interpretable, as they directly utilize the known spatial layout of virtual objects without requiring computationally expensive DL inference. While DL approaches such as SAMURAI [40] and YOLOX [41] excel in unstructured environments like Augmented Reality (AR), where depth and object positions must be inferred, they often fall short in VR due to their dependence on large amounts of labeled data and pixel-level depth classification [42, 43].

Recent studies have attempted to bridge this gap by combining geometric and data-driven principles. For example, work by Von et al. [44] introduced a convolutional neural network that integrates vergence cues with depth maps to estimate gaze distance in VR. Although such hybrid models show promise for improving gaze estimation accuracy in complex scenes, they remain dependent on large training datasets and high computational overhead, limiting their practicality for real-time interaction tasks.

Therefore, this thesis adopts a deterministic vergence-based approach rather than a DL framework, prioritizing precision, transparency, and responsiveness over data-driven generalization. This choice ensures that interaction performance remains interpretable and reproducible across diverse user conditions without reliance on pre-trained models or extensive calibration. This choice also ensures that the results remain directly comparable to prior gaze-based selection studies that employ similar deterministic methods, avoiding the ambiguity introduced by fundamentally different model architectures. Nevertheless, once the experimental findings of this thesis are established, future work could incorporate deep learning or other machine-learning techniques as an alternative pipeline to benchmark against the presented method.

3.4 Eye-Head Coupling in VR Systems

Eye-head coupling, the coordination between eye and head movements, is crucial to achieving instinctual interactions in VR. This interplay enables users to engage with targets across their field of view, improving the realism and usability of virtual environments.

In virtual environments, gaze and head movements are tightly interlinked. For example, when users focus on a distant object, they often use a combination of eye movements for precise targeting and head movements for broader angular shifts. This coupling has been shown to influence gaze accuracy and interaction speed, particularly in scenarios involving dynamic or peripheral targets. Studies have highlighted that head movement compensates for the limitations of eye movement range, enabling users to maintain engagement with objects outside their central visual field [45].

The design of VR systems can benefit significantly from integrating eye-head coupling models. For

example, predictive algorithms that anticipate head movements based on gaze direction can improve the performance of gaze-based object selection and tracking. This is useful to achieve reduced latency, a persistent challenge in VR systems. Additionally, understanding the nuances of eye-head coupling can inform more effective calibration techniques, ensuring that gaze-tracking systems remain accurate during rapid, concurrent eye-head movements. [46, 47] offer examples of predictive modeling that emphasize these benefits.

As VR systems continue to evolve, exploring and optimizing eye-head coupling dynamics represents a promising direction for enhancing both usability and accessibility across a wide range of applications.

3.5 The Role of Fatigue in Virtual Environments Systems

Fatigue, either physical or cognitive, is a common challenge in VR that can hinder prolonged use and immersion. Mitigating fatigue is essential for creating more user-friendly and accessible VR systems.

Eye-tracking technology reduces fatigue by enabling natural hands-free interactions. By allowing users to navigate and interact with virtual environments using gaze, it minimizes reliance on controllers and repetitive gestures, thereby reducing physical strain over time. Additionally, eye-tracking has been used to assess fatigue directly on HMDs, helping to identify strain-inducing interactions and achieve ergonomic improvements [48].

An alternative approach is the RayHand navigation method [49], which combines gaze direction with relative hand positioning to enable seamless navigation. This multimodal system balances the workload between hand gestures and gaze input, distributing the strain and providing a versatile approach to reducing fatigue.

Research has also explored the foundational factors that influence visual fatigue, such as color schemes and DoF simulation. For example, studies highlight how optimized color schemes can reduce strain during visually demanding tasks [50], while DoF effects may either alleviate or aggravate fatigue depending on individual sensitivity [51]. These insights emphasize the importance of designing VR environments with perceptual ergonomics in mind to minimize strain during extended use.

A study on eye movement features in VR gaming environments [52] proposed a visual fatigue detection algorithm that dynamically monitors users' eye behavior throughout the game. By analyzing the relationship between head and eye displacements, the system can estimate fatigue levels in real time, enabling adaptive feedback or interaction adjustments. This approach marks a shift from passive fatigue mitigation to active fatigue management, allowing the system to intelligently respond to users' physiological states to preserve comfort and engagement.

3.6 Interaction Techniques

Interaction techniques are critical in VR because they can both mitigate the inherent inaccuracies of tracking devices and enhance the overall responsiveness and usability of 3D environments. When dealing with eye-tracking data, these techniques are often divided into two categories: *detection techniques*, which focus on identifying the correct target among multiple targets, and *confirmation techniques*, which ensure the system accurately interprets the user's intentions.

For detection techniques, the objective is to select one or multiple targets. This can be achieved with a simple gaze-ray, but such methods are not always reliable in complex scenarios, so alternative detection strategies have been explored. For example, some approaches duplicate the selected targets or introduce proxy shapes (such as indicators connected by a line to the originals), enabling the user to visually verify whether the intended object has been correctly identified [53].

For confirmation techniques, the goal is to verify the intended target. This can be achieved by leveraging the user's motor behavior, such as winks or blinks [8], fixations or dwell times, and saccades [54]. Some approaches combine multiple techniques to improve reliability. For instance, the "Gaze+Hold" method [55] leverages the eyes as separate input channels, with one eye modulating the interaction state (open/closed). At the same time, the other provides continuous input, and a similar work [6] developed a mini game aimed at firing by closing one eye while aiming with the other, as shown in fig. 3.2.

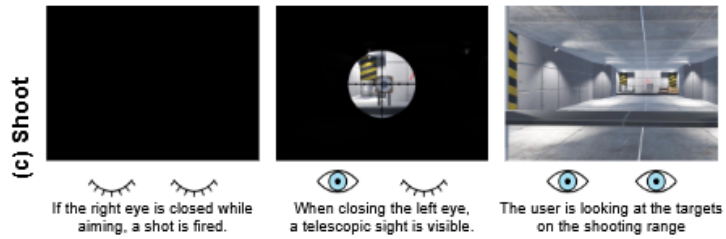


Figure 3.2: Illustration of how both eyes are independently used to complete the task of shooting, in a mini game developed by Rebsdorf et al. [6].

These techniques vary in naturalness and reliability. For example, dwell-based selection has been shown to produce fewer errors than other methods, but at the cost of significantly slower interaction and throughput [56]. In contrast, blink-based or wink/blink confirmations can be fast but are prone to false activations and user fatigue, especially when distinguishing between natural and intentional blinks [31].

Since no single technique is universally optimal, selecting an approach requires balancing speed, accuracy, and user comfort for the specific task and interaction context.

3.7 Research Gaps

Based on a review of the current literature, several research gaps were identified that this work aims to address.

Research Gap 1 (RG1). Current methods for gaze depth estimation in VR often rely heavily on traditional vergence techniques, which struggle to adapt to dynamic and complex virtual environments. This limits their effectiveness in scenarios requiring rapid focal changes or intricate object interactions.

Research Gap 2 (RG2). Inaccuracies in eye-tracking data, often linked to physiological limitations of the eyes or hardware constraints, continue to skew results in gaze-based systems. While several studies propose interaction techniques to mitigate these inaccuracies, many of these approaches remain untested in complex VR scenarios or rely on additional manual input to compensate for a lack of precision.

Research Gap 3 (RG3). Visual fatigue caused by prolonged focus tasks or imprecise tracking in VR environments is poorly understood. Existing studies often fail to examine how tracking inaccuracies or task complexity contribute to user fatigue, particularly during interactions that require extended precision or rapid focus shifts.

Based on these Gaps, the three Research Questions in Section 1.2 serve as guidelines for developing the experiment and evaluating it. They were not answered equally by all experiments, but they are all addressed in this work.

4

Vergence Study

Contents

This chapter presents the first stage of research, focusing on the evaluation of vergence precision and reliability in VR. It describes the experimental design, methodology, and results that address the initial research questions defined in chapter 1, establishing the technical foundation for the subsequent confirmation and selection studies (Chapter 5).

4.1 Explanation and Motivations

The current state of the art in gaze depth estimation does not identify any vergence-based method as consistently superior or more reliable. This section outlines the motivation for evaluating the reliability of vergence when working in VR and describes the methodological approach adopted for this study. The goal of this study is to assess whether vergence can provide stable and precise depth information suitable for real-time interactive use in virtual environments, directly addressing RQ 1.

To ensure accurate data collection, a high-performance HMD capable of providing binocular eye-tracking data was required, allowing access to both eye positions relative to the head and their corresponding gaze ray directions.

The first stage of this research, therefore, focused on evaluating the reliability of the selected HMD and verifying the effectiveness of the vergence estimation method proposed by Duchowski [5]. This validation step served as a technical foundation for subsequent experiments, ensuring that later interaction techniques were developed on a verified, consistent depth-estimation process.

The following formulas explain the algorithm that makes it possible to obtain the fixation point using vergence:

$$P_a = P_1 + t_1(P_2 - P_1) = P_1 + t_1 \cdot R_2 \tag{4.1}$$

$$P_b = P_3 + t_2(P_4 - P_3) = P_3 + t_2 \cdot R_4 \tag{4.2}$$

$$P_m = \frac{P_a + P_b}{2} \tag{4.3}$$

$$t_2 = \frac{(P_3 - P_1) \cdot R_2(R_4 \cdot R_4) - (P_3 - P_1) \cdot R_4(R_4 \cdot R_2)}{(R_2 \cdot R_4)^2 - (R_2 \cdot R_2)(R_2 \cdot R_4)} \tag{4.4}$$

$$t_1 = \frac{(P_3 - P_1) \cdot R_2 + t_2(R_2 \cdot R_4)}{R_4 \cdot R_2} \tag{4.5}$$

The input variables used in this algorithm have the following meaning:

- P_1 and P_3 are the positions of the left and right eyes, respectively.
- $(P_3 - P_1)$ is the IPD.
- R_2 and R_4 are the left and right gaze rays, respectively.
- P_2 and P_4 were used in the original iteration of this technique, by Wibirama and Hamamoto [17], where the gaze rays intersect a 2D screen providing two natural gaze ray intersections. This method was enhanced by Duchowski, to avoid the need for this step.

This method computes the point of closest approach (P_m) between the gaze rays of both eyes, assuming that their intersection approximates the user's fixation point.

Equation (4.1) and Equation (4.2) calculate the closest points on the gaze rays of both eyes (P_a and P_b). These calculations use the origin points of the gaze rays, along with the scalar constants t_1 and t_2 from Equations 4.4 and 4.5, which represent the parametric distances along each gaze ray obtained. Once the coordinates of these closest points are identified, Equation 4.3 determines the midpoint of the line segment connecting them. This midpoint, derived from vergence, provides an estimate of gaze depth. The vergence method is illustrated in Figure 4.1.[1]

[1] In the equations, R_2 and R_4 correspond to R_1 and R_2, respectively in Figure 4.1.

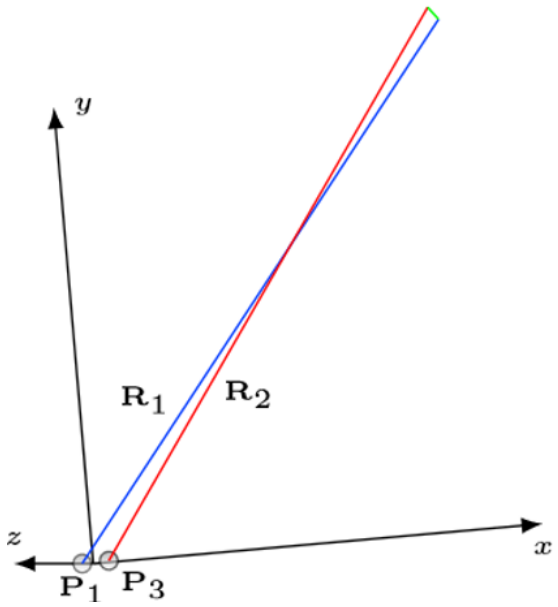


Figure 4.1: Visual representation of the vergence algorithm ([5]). The lines R_1 and R_2 represent the gaze rays originating from the left (P_1) and right (P_3) eyes. The shortest connecting segment (green line) defines the vergence distance, and its midpoint corresponds to the estimated 3D fixation point.

4.1.1 Experiment Setup

The first step was to determine which HMD to employ for the experiment. The Pico Neo 3 Pro was initially considered due to its integrated eye tracking capabilities, but was discarded because it did not provide the necessary variables in the eye-tracking data, or such variables were inaccessible through its SDK. Specifically, this HMD lacks both individual eye-position data relative to the head and gaze direction per eye, making it incompatible with the selected vergence estimation algorithm. Although the Pico Neo 3 Pro offers competent eye tracking for general applications, its limited developer access renders it unsuitable for research that requires full binocular tracking and geometric vergence computation.

The second option, which was finally selected, is the Varjo Aero, which has none of the above problems. The Varjo can capture comprehensive eye data, including pupil size, which may affect both DoF and IPD, and its high frame rate further ensures precise and reliable measurements, making it an ideal tool for vergence-based experimentation. The main drawback is that the Varjo Aero is a PC-tethered device, requiring an external computer for computation and at least two HTC Vive Base Stations [57] for spatial tracking. While this increases cost and setup complexity, the resulting precision and stability justify its use in this study.

With the hardware selected, a test environment was created to evaluate the data collection quality of the Varjo Aero. The environment was implemented in Unity, chosen for its seamless integration with commercial HMDs, its suitability for gaze-tracking applications, and the extensive support offered through its available plugins and research extensions [58]. The virtual scene featured a single moving

sphere that maintained a fixed position relative to the user's head. The sphere was selected as the target stimulus because spherical objects are widely used in gaze-tracking studies, offering consistent visibility and minimizing orientation-related biases. The sphere cycled through five depth levels and various horizontal and vertical locations. These positions were defined using spherical coordinates, azimuthal (Φ) and polar (Θ) angles, to maintain a constant distance while changing orientation. The full path was generated randomly before each trial, and the sphere paused briefly at each position before transitioning. The participants were instructed to fixate on the sphere at every position while the system continuously recorded gaze data (Figure 4.2).

To verify whether the user's gaze intersected with the sphere, an invisible gaze ray was cast from the headset's eye-tracking data. When the ray intersected the sphere, the target turned green; otherwise, it remained red. This visual feedback confirmed fixation detection and helped identify which positions were more challenging for the participants.

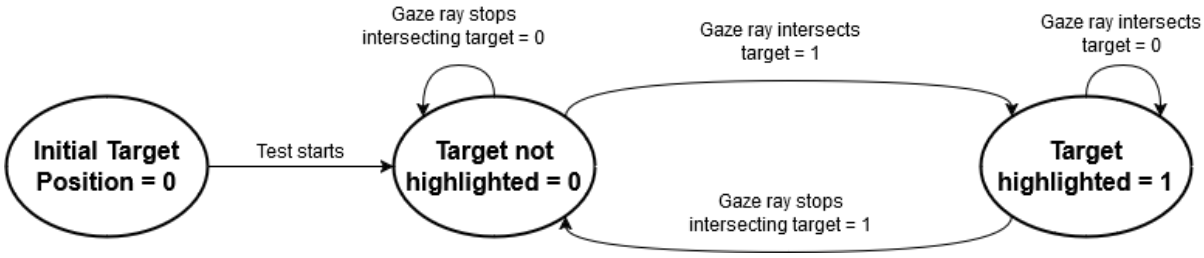


Figure 4.2: Vergence test workflow: the target starts in a default position and then, when the experiment starts, moves sequentially through a randomized set of positions, pausing briefly at each before transitioning.

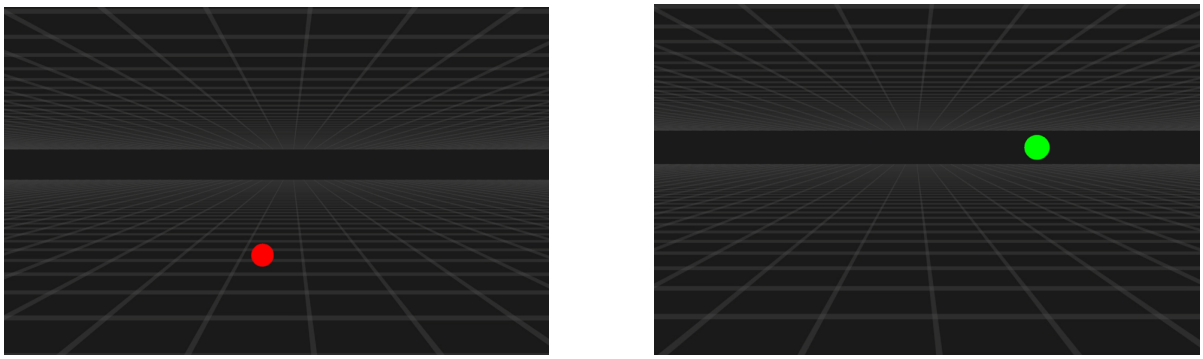
We initially explored several experimental variables in pilot tests to refine the conditions, as for example:

- **Floor and Ceiling (FC) vs. No Floor and Ceiling (NFC)** – The presence of a floor and ceiling was compared with a void condition (NFC) to determine whether visual reference surfaces improved user comfort and gaze stability. The FC setup positioned the surfaces near the upper and lower bounds of the sphere's possible locations without occluding it.
- **2 Seconds vs. 3 Seconds** – Two dwell durations were tested to determine whether the shorter 2-second interval allowed sufficient fixation time for accurate data collection or if a longer 3-second interval was necessary. This duration accounts for the time it takes the object to move from its current position to the next.
- **Target Object Size** – Initially, the sphere maintained a constant angular size of 20 pixels at all distances. Later, this was increased to 30 pixels to improve visibility and measurement consistency.

After analysis and feedback, the configuration using the FC environment, a 3-second dwell time, and a 30-pixel sphere diameter was adopted as the standard setup.

Initially, through different iterations, it was decided that the range used would be $\pm 21^{\circ}$ horizontally and vertically. Still, the range was later shortened to $\pm 10^{\circ}$ in both the horizontal and vertical directions, consistent with comfortable and accurate gaze movement limits for similar experimental tasks [59, 60]. Five depth levels were tested: 0.3 m, 0.6 m, 0.9 m, 1.2 m, and 1.5 m, each containing 25 positions (5×5 grid), resulting in 125 total fixation points per participant.

Additional refinements were implemented to enhance user comfort and reduce measurement noise. The background was changed to black to minimize visual strain, and the sphere colors were adjusted for strong contrast: red when the sphere is not fixated and green when the sphere is fixated, as shown in Figure 4.3.



(a) not focused **(b)** focused

Figure 4.3: Vergence test captures: 4.3(a), sphere is red when the gaze ray does not intersect the sphere; 4.3(b), sphere is green when fixation is detected.

Data corresponding to transitions between positions were excluded from the analysis, as rapid sphere movement during these intervals precluded stable fixations and thus smooth pursuit eye data. The main performance objectives were to maintain position error below 1%, achieve fixation times under 400 ms, and approximate the natural human focusing speed observed in real-world conditions.

4.1.2 Analysis Methodology

This experiment evaluated the focus speed and target-detection accuracy, as well as the capabilities and limitations of the selected HMD. The goal was to verify whether the chosen vergence method performed as expected and provided stable fixation data suitable for later stages of the research.

During the test, only one process was dedicated to data handling and to the calculation of the fixation point using the vergence algorithm described in Section 4.1. After concluding the experiment, the collected data were analyzed in two stages: first through a direct observation of the error distribution for each trial, as illustrated in Figure 4.4, and later through quantitative accuracy metrics.

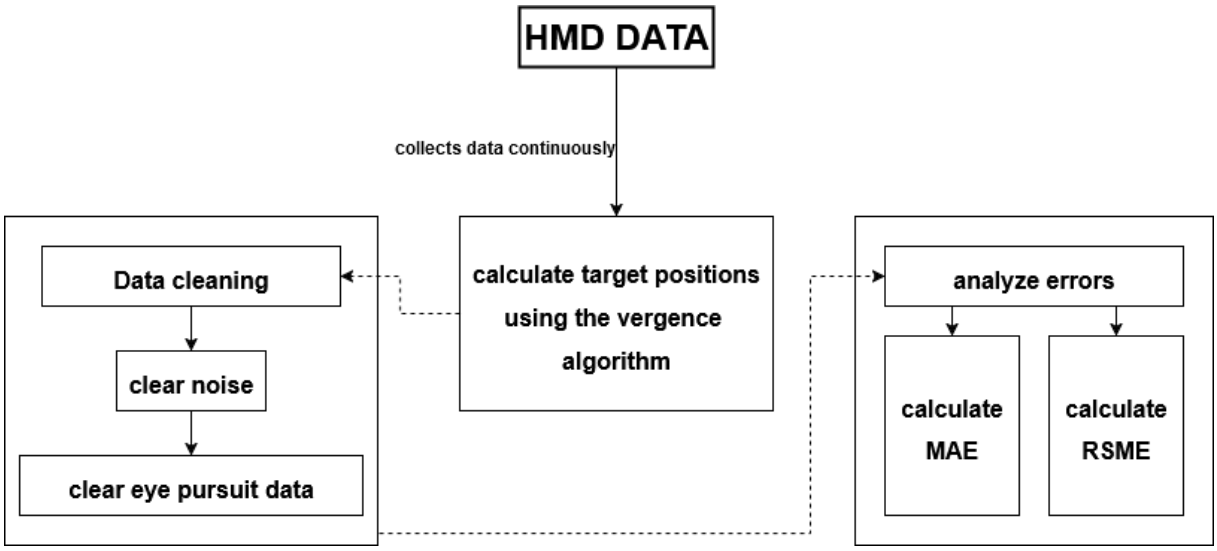


Figure 4.4: Data Treatment: All the data collected in the experiment is analyzed post-experiment, clearing all the non-relevant data when the object is moving or is not being focused. Performance was then assessed using both Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

To evaluate accuracy and outlier influence, both the MAE and the RMSE were computed for each successfully detected position. The following formulas show how they will be used to answer our questions and what the variables mean in the context of this test:

- **MAE**, used as a simple way to determine horizontal, vertical and depth accuracy and the formula is as follows, $\frac{\sum_{i=1}^{n} |th - ch|}{n}$, where **th** represents the true horizontal location of the object and **ch** is the horizontal location calculated using the vergence. The same procedure was used for the vertical angle and depth.

- **MAE%**, calculated using, $\frac{MAE}{range} \times 100$, where $range$ is equivalent to $(MaxAngle - MinAngle)$ for both horizontal and vertical angles, and $(MaxDistance - MinDistance)$ for depth.

- **RMSE**, used as a way to determine the existence of big outliers in the results data and determine how influential they are. The formula is as follows, $\sqrt{\frac{\sum_{i=1}^{n} (th - ch)^2}{n}}$.

- **RMSE%**, calculated using $\frac{RMSE}{range} \times 100$, where $range$ is equivalent to $(MaxAngle - MinAngle)$ for both horizontal and vertical angles, and $(MaxDistance - MinDistance)$ for depth.

The results of the vergence algorithm [5] were compared with the HMD 's native estimations to determine any changes in tracking accuracy. The analysis included both focused and unfocused states, excluding the brief transition periods when the sphere moved to a new position to allow natural saccades to complete.

As this is a small-scale preliminary study, no formal pre- or post-experiment questionnaires were used to assess fatigue. Instead, qualitative feedback from participants was collected informally to confirm that the task did not induce noticeable visual strain.

4.2 Results

This section presents the findings from the vergence reliability study, focusing on both gaze direction and depth estimation accuracy. The analysis aims to determine whether the vergence algorithm proposed by Duchowski [5] provides consistent, precise results across different spatial conditions. In particular, the section examines horizontal and vertical gaze errors, overall vergence depth estimation performance, and the variability observed among the participants. Quantitative results are supported by statistical indicators such as the MAE and the RMSE, followed by an interpretation of the outcomes and their implications for subsequent stages of the research.

4.2.1 Participants

A total of eleven individuals (6 males, 5 females) participated. Five of them wore glasses, which had to be removed during testing because the Varjo Aero requiring the absence of reflective surfaces for accurate eye calibration. Contact lenses did not appear to interfere during calibration, and their influence likely went unnoticed at this stage. This factor may have affected depth estimation, but had minimal impact on fixation direction. All eleven participants were included in the analysis and most had no prior experience with eye-tracking or VR systems.

4.2.2 Data Analysis

The analysis showed that participants consistently and smoothly detected all target positions without reporting noticeable fatigue or eye strain.

As shown in Table 4.1, the horizontal angle errors remained below 1%, with slightly higher values observed at shorter distances. When comparing the MAE obtained from the vergence algorithm with that provided directly by the HMD, the algorithm yielded more accurate results (0.29% vs Varjo's 0.85%).

Similarly, Table 4.2 shows that the vertical angle errors followed the same trend, smaller at closer distances and slightly increasing with depth. The vergence algorithm achieved an average MAE of 0.32%, compared to Varjo's 0.39%, confirming a modest but consistent improvement.

Overall, the RMSE remained low for both horizontal and vertical directions, slightly above the MAE but never exceeding 3%, indicating that outliers had little influence on the accuracy of the estimation of gaze direction.

Table 4.1: Data measurements analysis: Horizontal Angle using Vergence Algorithm

Depth	MAE	MAE%
0.3 m	0.151°	0.76%
0.6 m	0.061°	0.31%
0.9 m	0.033°	0.16%
1.2 m	0.029°	0.15%
1.5 m	0.028°	0.14%

Table 4.2: Data measurements analysis: Vertical Angle using Vergence Algorithm

Depth	MAE	MAE%
0.3 m	0.096°	0.48%
0.6 m	0.066°	0.33%
0.9 m	0.063°	0.32%
1.2 m	0.048°	0.24%
1.5 m	0.057°	0.28%

The results in Table 4.3 indicate that the vergence algorithm performed better in closer ranges, while the built-in Varjo estimation showed higher accuracy beyond 0.9 m. However, the overall variability between the participants was considerable, with MAE differences reaching up to eightfold at certain distances (see fig. 4.5 for percentage comparison). This variability was more pronounced among participants who removed their glasses, although even within the group without vision correction, significant fluctuations remained. This could explain the high results.

On average, RMSE values were relatively high, aligning with the variability trends, but this was true only when looking at all participants together. Individual-level analysis revealed that some participants consistently achieved low errors values, suggesting that user-specific visual and physiological factors strongly affect vergence-based estimation. The outliers increased as target distance grew, reinforcing the depth sensitivity of the method.

Table 4.3: Comparison of data measurements: Depth provided by Varjo vs. estimated using the Vergence Algorithm.

Depth (m)	Varjo Provided Depth		Vergence Algorithm Depth	
	MAE (m)	MAE%	MAE (m)	MAE%
0.3	0.335	27.94	0.213	17.72
0.6	0.376	31.30	0.316	26.37
0.9	0.496	41.36	0.553	46.12
1.2	0.560	46.68	0.657	54.79
1.5	0.558	46.48	0.715	59.57

4.2.3 Discussion

The vergence algorithm produced highly reliable gaze direction estimates, maintaining errors below 1% and ensuring responsive system behavior comparable to real-world accuracy. Depth estimation, however, proved less consistent than initially expected, with accuracy decreasing with increasing distance, diminishing its reliability. Variability between participants was substantial, likely due to individual visual differences (such as myopia severity or lens usage) and limited strong environmental depth cues. Even under simple conditions, these inconsistencies suggested that relying solely on vergence depth for selection tasks could introduce significant errors, particularly in complex environments.

To mitigate this issue, my approach focuses on building upon the reliable elements of vergence,

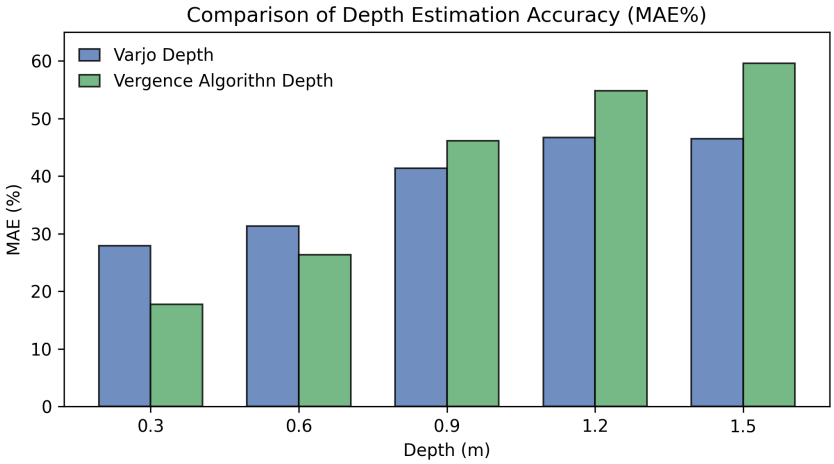


Figure 4.5: Comparison of depth estimation accuracy between the vergence algorithm and the Varjo-provided depth values across multiple distances. The bars represent the mean absolute error percentage (MAE%) for each depth.

namely gaze direction and the vergence point, while compensating for its depth imprecision. The proposed approach integrates these metrics into novel gaze-based selection techniques that incorporate a tolerance margin based on the observed MAE, enhancing robustness without sacrificing intuitiveness. This strategy ensures that vergence remains useful as a core input while addressing its limitations, preparing the groundwork for application in more challenging scenarios.Chapter 5 details the development and evaluation of these interaction techniques.

5

Eye-Tracking-Based Object Selection in 3D

Contents

This chapter explains how we developed 3D interaction techniques using eye tracking with depth estimation.

5.1 Techniques Design

As discussed at the end of chapter 4, the vergence algorithm demonstrated strong precision in terms of gaze direction, but its depth estimation proved less reliable. Although these limitations were manageable

in the controlled test environment, they highlighted clear challenges for scaling to more demanding scenarios where multiple targets may be closely spaced or partially occluded. In such contexts, even small depth inaccuracies could significantly compromise selection accuracy, making raw vergence data insufficient on its own.

At the same time, gaze-based interaction in VR still lacks a universally accepted "ideal" selection method. Existing approaches often trade off between precision, speed, and usability, and none have emerged as a clear standard across applications. This gap creates both a challenge and an opportunity: the challenge of overcoming inherent limitations in gaze tracking and the opportunity to design techniques that make gaze a more practical, reliable, and comfortable input modality.

To address this, our work takes advantage of the reliable aspects of vergence, such as gaze direction and the vergence point, while compensating for its shortcomings in depth. Selection techniques were designed to incorporate these inputs alongside a margin of error derived from the MAE obtained in the vergence study, ensuring that targets within natural tolerances are detected correctly. This approach transforms vergence from a limited raw signal into a more robust and adaptable interaction method, laying the foundation for effective use in cluttered and visually complex VR environments. In doing so, it contributes to the broader search for a gaze-based interface that can balance precision, efficiency, and user comfort. The techniques developed in this thesis include improved versions of classic methods enhanced through gaze-depth disambiguation, entirely new methods introduced for the first time, and adaptations of existing approaches redesigned to operate solely with eye-tracking input.

Data processing for all techniques follows the procedure outlined in chapter 4, where outliers that could skew results are removed (see fig. 4.4). In addition, a running mean of the last five valid gaze samples is maintained and updated at each instant. This averaged gaze data is then used as input for all selection and confirmation techniques, providing a smoother signal that improves stability and ensures a more consistent user experience.

The following sections introduce the selection and confirmation techniques developed in this work, along with the experimental setup designed to evaluate their performance in cluttered and visually demanding scenarios, the very conditions where traditional gaze-only approaches tend to break down.

5.1.1 Target Selection

Our first objective was to develop techniques for accurately detecting the user's intended target using **Varjo** gaze data and the algorithm's vergence point. Since there is no consensus on a single superior selection method for eye-tracking interactions, I implemented multiple techniques inspired by previous work, leveraging existing Unity features, and accounting for the errors observed in the preceding vergence experiment.

To achieve reliable selection in practice, it is crucial to consider both the raw gaze data and the

display update rate. The Varjo Aero headset samples eye gaze data at up to 200 Hz, capturing fine-grained eye movements. However, the application's effective update frequency is constrained by the headset's display refresh, which varied between 70 and 90 Hz during testing. Unity's Update() method executes once per rendered frame, meaning selection techniques operate at this refresh-dependent rate. Consequently, the responsiveness of gaze-based interactions is inherently tied to the headset refresh rate, ensuring that the application progresses at a speed acceptable for real-time interaction in dynamic 3D environments [61]. This downsampling is typical in VR systems and provides sufficient temporal resolution for accurate target selection, even if some micro-movements occur between frames [5, 62].

5.1.1.A OverlapSphere

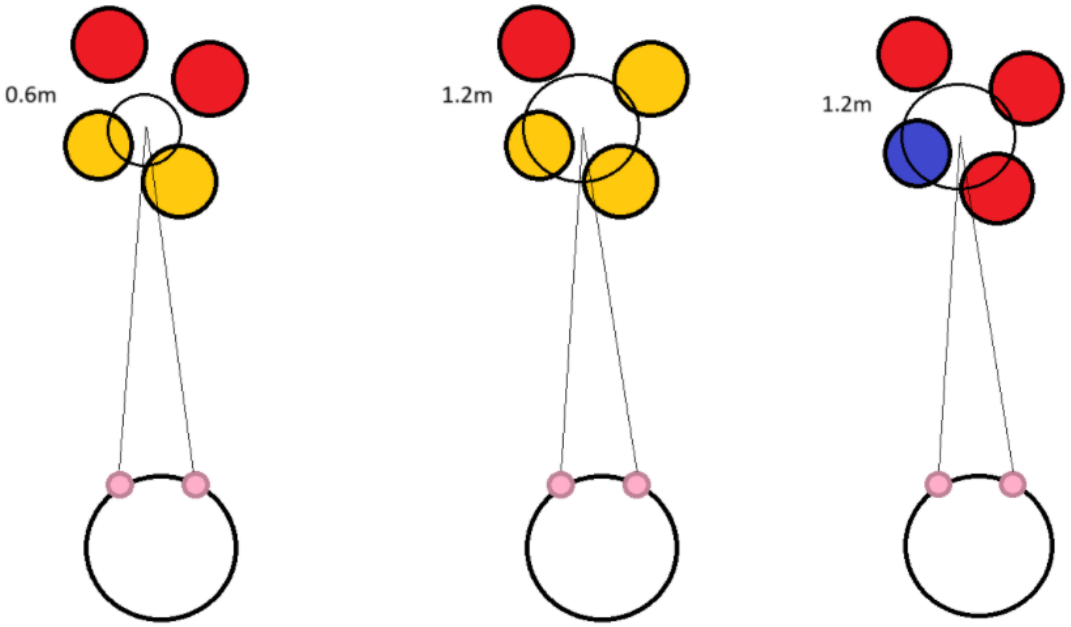
The first selection technique was designed with a specific goal: to investigate the impact of depth errors when using vergence. Specifically, this method ignores the knowledge that depth errors exist, allowing for determining whether accounting for depth improves target selection accuracy or whether similar results are obtained without it. By not factoring in depth error, this technique serves as a baseline for comparison with subsequent selection methods.

This technique consists of checking for collisions inside a generated sphere in the 3D space where the user is fixating. Its implementation leverages the built-in Unity Physics function `OverlapSphere()` [63]. This function takes a center point (a 3D vector) and a radius in meters (Unity's standard distance unit) and returns a list of all objects detected within the generated sphere. The center defines the sphere's position, while the radius determines its size, allowing for straightforward detection of nearby targets in a 3D space.

This method is not perceived by the user, considering that the sphere used to detect collisions is invisible. The first implementation consisted of a simple process: all objects caught within the black circle (detection sphere) are highlighted in orange, as shown in Figure 5.1(a). This iteration allowed me to test the detection's stability. After seeing the results, it was clear that multiple objects would be highlighted frequently, indicating that a method to trim down the available targets would be required. To achieve this, I compare the distance between every detected object's center and the center of the generated sphere created by the function; the closest object would be the selected target and would, in turn, be highlighted, as seen in Figure 5.1(b), where the target turned blue.

The radius (r) of the sphere will be decided based on two factors, as shown in Equation (5.1), where the 2° represents the central degrees of the visual field that the fovea can process, fig. 2.3, and the error is the MAE found in 4.2.2, also in degrees. This MAE is used to account for the error in the Varjo eye data measurements. Considering that I measured the MAE only for five specific depths, I needed to consider how to handle different depths in between them, therefore, to solve this problem, I used interpolation[1].

[1] Interpolation is a mathematical and statistical method for estimating unknown values that fall within the established range of



(a) concept **(b)** final iteration

Figure 5.1: Top-down view of an individual using the selection technique OverlapSphere: 5.1(a) first idea where there was no distinction between possible targets, so all of them are selected; 5.1(b) final concept where the possible targets are no longer visible, only highlighting the selected target.

Using this method, I assume that, between two adjacent depths, my error follows a pattern. As an example, if the vergence algorithm indicates that 0.4 meters is the current depth, my method will recognize that the error is between 0.3 meters and 0.6 meters, having 0.151° and 0.061° as their MAE values, respectively, according to table 4.1. In this scenario, as the MAE decreases, the pattern is to lower the error; however, in table 4.2, between 1.2 meters and 1.5 meters, the pattern is to increase the error. In eq. (5.1), $distance$ is equivalent to the current depth.

$$r = \tan(2^\circ \ + error) * distance \tag{5.1}$$

Since I am determining the radius of a sphere, another rule was added to my method, the error value I choose to use will be the biggest one between the horizontal MAE, table 4.1, and vertical MAE, table 4.2. Following the prior example where the current depth is 0.4 meters, my method will calculate the error through interpolation and then decide which one to use. The radius of the sphere has a minimum size equivalent to the largest possible target in the 3D environment for the experiments, which will be explained in greater detail in Section 5.3, which means that if the radius calculated using eq. (5.1) is below that value, it will be overwritten.

known data points.

The center point used in the `OverlapSphere()` function is determined using the vergence point estimated by the algorithm, with an added error margin along the vector connecting the midpoint between the eyes to the vergence point. This error margin corresponds to the size of the largest possible target in the 3D environment. This adjustment is motivated by a known limitation in VR gaze estimation, namely that vergence-based depth often diverges from the true fixation depth [64]. When humans fixate on an object, vergence tends to align with its front surface rather than its geometric center [20], which means that half of the detection sphere would be ineffective depending on whether it extends in front of or behind the intended target. By incorporating this additional depth, the method approximates the value that would be obtained if the fixation were directed at the center of the object.

The strength of this selection technique lies in its simplicity to implement and in the way it avoids selecting more than two or three targets most of the time when it decides which target is going to be highlighted, which helps avoiding possible wrong targets, but it also has a problem. Specifically because it is more precise when verifying which targets are colliding (detection sphere is small), if the error is larger, like it happens with the depth as seen in **??**. For example, at larger distances, the target may not be properly detected because this technique does not benefit from the depth errors I analyzed in chapter 4.

5.1.1.B SphereCast

My second selection technique builds on the first by explicitly incorporating depth error into the process. The aim is to improve both accuracy and comfort by addressing the limitations of the previous approach, which ignored vergence error. By accounting for depth error, this technique is designed to deliver superior selection performance in more demanding scenarios.

The implementation relies on Unity's built-in Physics function `SphereCastAll()` [65]. As the name suggests, this function casts a sphere from a starting point in a specified direction over a defined distance, sweeping through the 3D environment and detecting every object it intersects. The function then returns these objects as a list. Its input parameters include:

$origin$ Is the 3D vector that will decide where the sphere will be cast in the 3D environment.

$radius$ The radius of the sphere that will be cast, measured using meters.

$direction$ Is the 3D vector that represents the direction in which to cast the sphere.

$maxDistance$ Maximum length for the sweep, measured using meters.

This method is not perceived by the user since the sphere being cast is invisible at all times. As seen in fig. 5.2, the area created by the SphereCast selects all the possible targets. Then it reduces the options to only one, which is decided by comparing the distance between every detected object's center, to the vergence point, and selecting the closest one. The shape of the sweeping area is reminiscing of a cylinder, but it is more accurate to call it a capsule considering that both ends of the "cylinder" have half

a sphere protruding out.

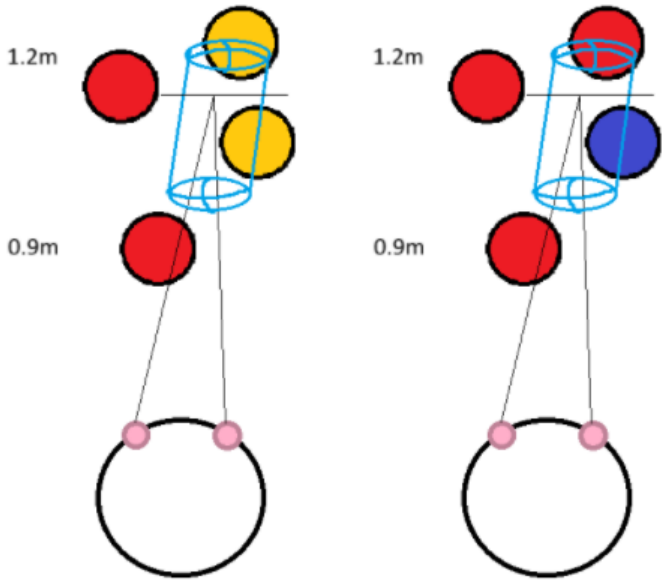


Figure 5.2: Top-down view of an individual using the selection technique SphereCast: on the left side, in blue can be seen the representation of the area where collisions are detected, the selected object candidates are highlighted in orange; on the right side, the same scenario is showed but only the target that will be selected is highlighted.

The $origin$ is determined once again using the vergence point estimated by the algorithm, with an added error margin along the vector that connects the midpoint between the eyes to the vergence point, which would be $direction$. This error margin corresponds to the size of the largest possible target in the 3D environment, thus yielding an approximate value that would be obtained if the fixation were directed to the center of the object. After receiving the vergence point, it is subtracted from the depth error along the $direction$ used for the sweep because the error can be closer to or farther from the intended target.

This error will be calculated using interpolation. It will have a similar implementation, but the values used come from table 4.3 instead and will not be compared to any other values, since only depth is relevant for distance. Seeing how the MAE values increase consistently the farther away the target is, the pattern of interpolation will always be to increase the value. The $maxDistance$ is twice the error, so it accounts for the error before and after the vergence point.

The $radius$ used is 1 cm, since this technique's main purpose is to leverage depth error and because the directional error is small enough that 1 cm is already sufficient to select the intended target, assuming no external factors degrade the gaze ray direction, such as poor calibration. The strength of this selection technique is that it leverages the known depth error, which helps compensate for its imprecision relative to direction accuracy, reducing the likelihood that the intended target is not between the colliding objects.

The only problem is that in more complex scenarios there might be cases where the wrong target is highlighted because too many targets are being caught by the collider.

5.1.1.C Raycast

My third selection technique takes a different approach by discarding both horizontal/vertical gaze errors and depth error. Instead, it relies on a direct-gaze ray cast through the vergence point, intersecting potential targets along its path. This method is designed to test whether a simplified geometry-based approach can achieve reliable selection without compensating for vergence inaccuracies.

The implementation combines the vergence point obtained from the algorithm with a gaze ray cast in that direction. The ray itself is invisible to the user, but in practice it extends from the midpoint between the eyes through the vergence point, as illustrated in fig. 5.3. To ensure that out-of-range objects are not excluded, the ray was given a fixed length of 3 meters. All objects intersected by the ray are flagged as potential targets and added to a candidate list. The final selection is then determined by comparing the depth of the vergence point with the distance between the center of each candidate object and the midpoint between the user's eyes.

In fig. 5.3, the purple cross represents the midpoint between the eyes, and the highlighted orange spheres represent the objects intersected by the gaze ray. The method compares the distance from this midpoint to the center of each intersected sphere and selects the one whose depth is closest to that of the vergence point.

The strength of this selection technique is its simplicity and the ease with which it can increase the likelihood of hitting the target, because simply tilting the head or looking at a different side of the target can eliminate incorrect targets from consideration. It also takes advantage of the direction's precision by not relying on a large intersection area, thereby avoiding more potential mis-targets. However, there is one disadvantage: since it does not account for depth error when selecting a target, this technique might choose the wrong targets in a highly complex scenario where objects are positioned behind one another and close together.

5.1.1.D ConeCast

My fourth and final selection technique, ConeCast, was designed as a counterpoint to the precision-based methods presented earlier. Instead of relying on highly accurate gaze measurements and precise fixations, this technique adopts a two-step process that prioritizes simplicity and tolerance to inaccuracy. The goal is to reduce the burden of precision on the user and make target selection in complex, occluded environments more manageable.

This technique was not initially planned, but after reviewing Xu's work EyeExpand [7], I implemented a solution inspired by their two-step method to address imprecision in 3D object selection, which is

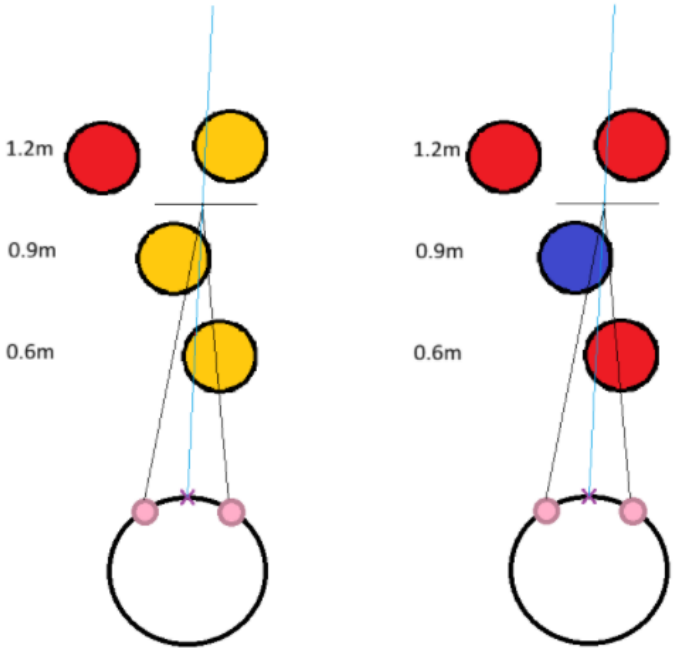
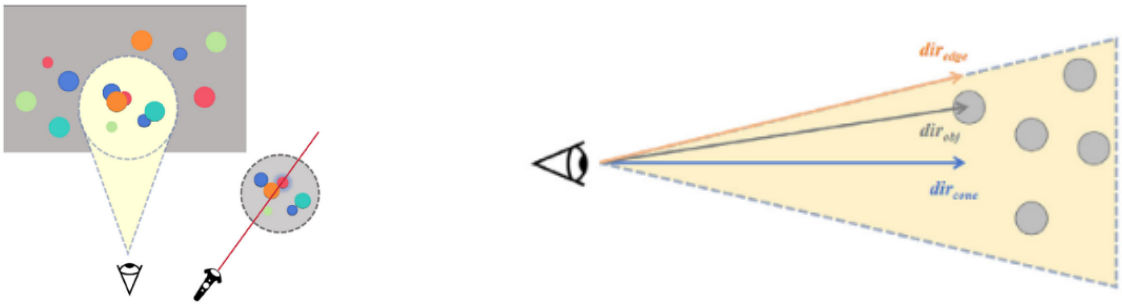


Figure 5.3: Top-down view of an individual using the selection technique RayCast: on the left side, in blue can be seen the representation of the gaze ray, the selected object candidates are highlighted in orange; on the right side, the same scenario is shown but only the target that will be selected is highlighted.

similar in goal to my own work. The first step is ConeCast, which starts when the user holds the controller trigger. The VR headset tracks eye movement and activates an invisible cone aligned with their gaze, fig. 5.4(b). Objects inside the cone are outlined to show what the user is focusing on. Once the trigger is released, those objects are rearranged in a flat circular area in front of the user. They are arranged on the circular plane to reflect their relative positions in the cone, scaled down, and spaced to avoid overlaps, creating a clear miniature of the scene where the user can easily select targets with the controller fig. 5.4(a).

My implementation will adopt the same general concept and steps as EyeExpand, but with a different interaction approach. The key distinction is the removal of controllers, as this work focuses on developing eye–tracking–based selection techniques without additional input devices. The first step will be to adapt Conecast. Rather than relying on a controller trigger to confirm selection, I introduce an eye-based triggering mechanism, which will be described in section 5.1.2. In contrast to EyeExpand's static cone, my approach uses a more flexible cone whose depth adapts to the vergence point, while maintaining the same alignment logic by orienting the cone along the gaze direction, with its vertex at the eyes and its axis extending outward.

As illustrated in fig. 5.6(a), the cone originates at the midpoint between the user's eyes and extends to the vergence point with an added error margin. This margin corresponds to the size of the largest possible target in the 3D environment. The angle of the cone varies with depth, determined by the mean



(a) Front View **(b)** Side view

Figure 5.4: Figure 5.4(a), an illustration of EyeExpand technique. EyeExpand allows users to select an area with gaze, where objects are rearranged on a circular plane in front of them, enabling them to select occluded objects quickly. Figure 5.4(b), An illustration of cone-casting. [7]

absolute error (MAE) values reported in table 4.1 and table 4.2 by interpolation. The larger of the two MAE values is applied, consistent with the procedure described in section 5.1.1.A.

Once selection is confirmed, a reconstruction of the selected targets is presented on a circular plane positioned 0.35 meters in front of the user and tilted 10° to the right relative to the gaze direction. Clones of the targets are placed sequentially from nearest to farthest relative to the cone's center. To ensure clarity, the method prevents overlaps between clones and resizes them based on their relative depth. Additionally, slight depth adjustments are applied to enhance the spatial impression of the recreated scene. When targets are located at similar depths, their scaling is identical; conversely, targets at greater depth differences will appear with proportionally larger size differences, with closer targets rendered noticeably larger than those farther away.

For ease of use, the original objects are highlighted in orange once selected, and their corresponding clones replicate this color, as shown in fig. 5.6(a). The only exception is the intended target: its clone retains the original color, ensuring it remains easily recognizable within the circular plane. To further support clarity, when a clone is selected, a line is drawn between the clone and its corresponding original object, providing a clear visual link that reinforces their association, as seen in Figure 5.5.

The final step is selecting the target, which is done using the third technique described in section 5.1.1.C. This method was chosen for its simplicity and reliability under conditions where the clones do not overlap significantly, as illustrated in fig. 5.6(b). Once the user selects the clone corresponding to the intended target via the eye-based triggering mechanism, the original object is highlighted, and the selection is confirmed, as shown in fig. 5.6(c). Suppose the user wishes to cancel the current selection, for example, after missing the intended target or because too many targets were selected. In that case, this can be done by diverting their gaze outside the circular plane in any direction and activating the eye-based triggering mechanism. After confirmation or cancellation, the circular plane and its clones

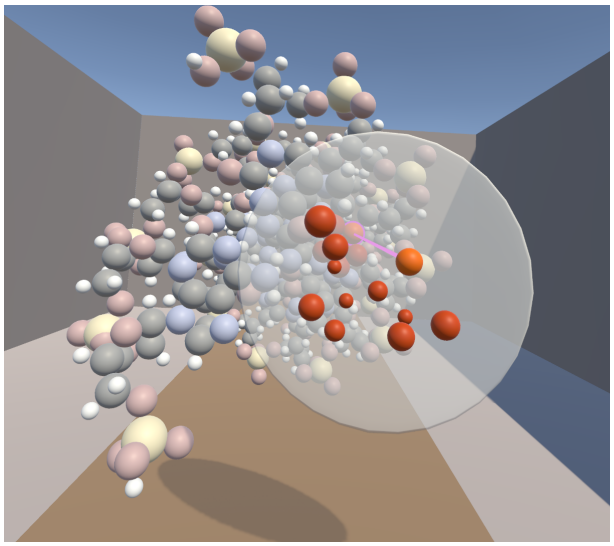
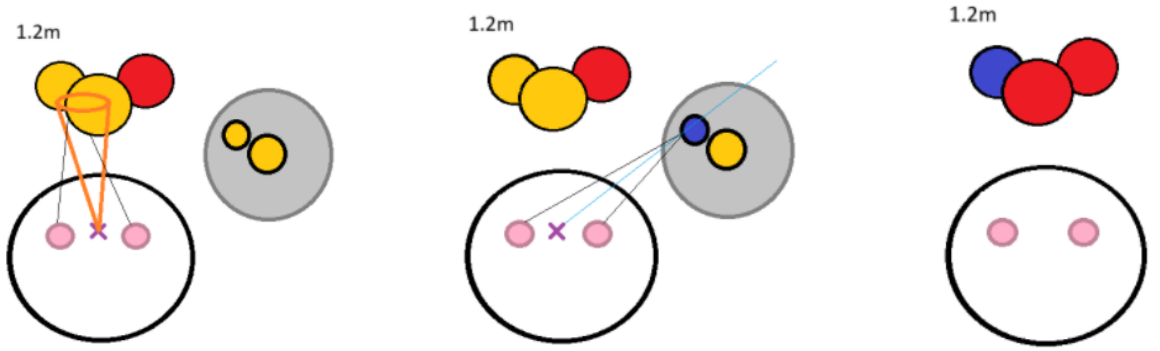


Figure 5.5: Circular Plane example where the selected clone is highlighted and connected to its original form, which is also highlighted.

are removed from the scene, leaving only the selected object highlighted in the case of confirmation.



(a) selecting targets **(b)** selecting clone **(c)** target confirmed

Figure 5.6: An illustration of the ConeCast technique. Figure 5.6(a) represents the moment the Conecast captures a group of objects. Figure 5.6(b) shows the RayCast selection technique being used to select one of the clone. Figure 5.6(c) shows that the selected clone's counterpart, the original, is confirmed as the selected target.

An advantage of the cone-based approach is that it reduces incorrect collisions at closer depths, a limitation not addressed in SphereCast section 5.1.1.B, whose shorter and more precise depth range made this less of a concern. A key strength of ConeCast is its high success rate: rather than requiring direct fixation on the exact target, it only requires selecting the area where the target lies. This property reduces visual strain in complex scenarios with frequent occlusions, consequently reducing task burden.

Conversely, the method is less efficient when targets are close or only partially occluded, as it converts a simple selection into an unnecessary two-step process. Moreover, since the confirmation mech-

anism is eye-based, each selection theoretically requires double activation, which could increase user fatigue during longer sessions. However, this drawback may be negligible in practice, as the technique's simplification of the selection process significantly reduces the likelihood of missed targets.

5.1.2 Target Confirmation

Having addressed selection, the next objective was to design a method that allows users to confirm the choice of one or multiple objects **without relying on controllers** and **without disrupting the current selection during confirmation**. As with selection techniques, no single gaze-based confirmation method is universally recognized as superior. Therefore, this work explores several alternatives inspired by existing research.

Varjo headsets collect gaze data at a 200 Hz refresh rate (200 readings per second) and provide detailed ocular metrics, including eye openness, which ranges from 0 (fully closed) to 1 (fully open). Eye openness is derived using high-speed Infrared (IR) cameras and illuminators that create corneal reflections ("glints"). Computer vision algorithms analyze the relative position of the pupil and glints over time to infer eye state, including whether the eyes are open or closed [66]. This metric played a key role in the development of two of the three confirmation techniques described in the following sections.

5.1.2.A Dwell

My first confirmation technique is Dwell, which relies on sustained fixation to confirm a target and serves as a widely studied baseline for gaze-based confirmation. It is one of the most common approaches in gaze-based interfaces and does not rely on eye openness measurements. Controlled studies on 2D displays have identified effective fixed thresholds; for example, Paulus and Remijn [67] reported that thresholds around 600 ms were both effective and preferred. The same basic dwell paradigm has been widely applied in 3D HMDs, such as the DEEP system by Yi et al. [68], which examined dwell-based behaviors and complementary eyelid cues to improve target selection in dense and occluded virtual scenes.

For my implementation, I adopted a simple dwell mechanism that confirms a target after 1500 ms of sustained gaze. Although 600 ms was effective in simpler 2D settings, shorter thresholds are insufficient for my more complex 3D environment, where occlusions are frequent and accidental selections are more likely. Preliminary testing found that 1500 ms reliably prevents unintended selections while remaining comfortable for the user.

The implementation also accounts for natural gaze instability. When a new target is focused but the threshold has not yet been reached, the confirmation timer is reset. If the user looks away without selecting another target, a secondary timer of 700 ms begins. If the user refocuses within this interval,

the confirmation timer resumes, and the secondary timer resets; otherwise, the target is deselected and no object remains under confirmation.

This dwell mechanism is compatible with all the implemented selection techniques, except ConeCast. Due to ConeCast's design, targets are almost always located within the cone. Applying dwell confirmation in this context would cause the circular plane to appear repeatedly, leading to a less comfortable, less natural user experience.

This dwell mechanism is compatible with all selection techniques except ConeCast. Due to the nature of this technique, targets are almost always within the cone's bounds, and applying dwell would cause the circular plane to appear repeatedly, resulting in a less comfortable, less natural user experience.

The main advantage of dwell selection is its minimal cognitive and motor demand: confirmation requires only sustained fixation on the intended target. However, if the underlying selection technique is unstable, dwell time may increase frustration and lead to inaccurate confirmations. Additionally, because dwell is always active, scenarios in which the user does not wish to select any target require intentional gaze aversion, which can cause discomfort or unnecessary eye strain.

5.1.2.B Wink

For the second confirmation technique, Wink, the objective is to provide a faster method that leverages the user's motor eye skills to confirm targets. Unlike dwell, which relies on sustained fixation, wink-based confirmation allows users to close one eye to register their selection deliberately.

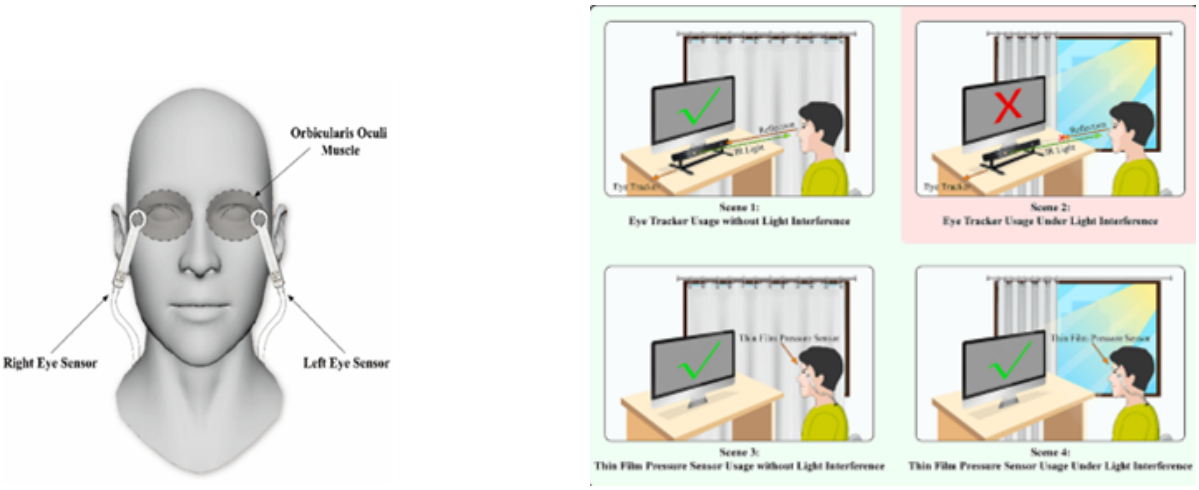
Although less prevalent than dwell, wink-based confirmation has been explored in several studies. For example, Fan et al. [8] investigated the characteristics and efficiency of voluntary blink actions, offering insights for optimizing blink-based Human–Computer Interaction (HCI) systems. In their work, eye openness was detected through a pressure sensor placed near the orbicularis oculi muscle[2], as shown in fig. 5.7(a). Unlike camera-based implementations, the sensor-based approach in [8] is unaffected by ambient lighting conditions, as illustrated in fig. 5.7(b). This aspect aligns with my own implementation within an HMD, which similarly operates independently of external light. This robustness is especially valuable for ensuring reliable performance in varied or complex environments.

My implementation of Wink is straightforward. First, the system detects when one eye begins to close: if the eye openness drops below 0.7, the currently selected target is locked, and no new targets can be chosen. This threshold was determined through testing, where 0.7 proved to be the lowest value at which gaze data remained stable; below this point, partial eye closure caused noticeable shifts in gaze tracking. Thus, 0.7 was chosen as the final reliable instant to record new data.

If ConeCast is the active selection technique, the gaze direction and head position are also locked at this stage to ensure that the clones are correctly positioned on the circular plane, disregarding minor

[2]The sphincter-like muscle surrounding the eye responsible for closing the eyelids, blinking, and aiding in tear drainage.

movements until confirmation is complete. The system then checks if the eye openness falls below 0.35 to register a deliberate wink. This lower threshold accounts for partial closures and ensures the system distinguishes intentional winks from normal eyelid variations. If both eyes close simultaneously, the action is classified as a blink rather than a wink. When the closed eye reopens above 0.35, the wink is registered and the selected target is highlighted. Finally, once the eye opens above 0.7, the system resumes and allows new target selections.



(a) Pressure Sensor **(b)** Experimental environment

Figure 5.7: 5.7(a): Pressure sensors used by Fan et al. [8] to detect eye activity. 5.7(b): Illustration of how different lighting conditions affect the experimental setup.

The main strengths of this technique are its speed and flexibility, as users can choose which eye feels more natural to wink. However, its limitations are user-dependent. Some individuals may find it difficult to wink on command, may unintentionally close both eyes, or may only partially close the intended eye. Task demands and individual differences in wink control can influence fatigue, particularly in complex, occluded environments, which is the primary focus of this work.

5.1.2.C Double Blink

My third and final confirmation technique, Double Blink, also relies on eye openness like Wink but does not require the potentially demanding motor skill of winking with a single eye, which can vary between individuals. Double Blink is similar to Wink but involves both eyes simultaneously. The rationale for this technique is based on the findings of Fan et al. [8], who compared various combinations of blink and wink with one, two, and three repetitions, as shown in fig. 5.8. They concluded that blinking twice achieved the best performance, which informed the decision to use double blinking in this implementation.

The implementation of Double Blink follows a logic similar to Wink, but with both eyes involved. First,

the system detects when both eyes begin to close: if eye openness in both eyes drops below 0.65, the currently selected target is locked, and no new targets can be chosen. This threshold was determined through testing, where 0.65 proved to be the lowest value at which gaze data remained stable; below this point, partial closures caused noticeable shifts in gaze tracking. Thus, 0.65 was chosen as the reliable cutoff point for recording gaze data.

If ConeCast is the active selection technique, the gaze direction and head position are also locked at this stage to ensure clones are correctly positioned on the circular plane, disregarding minor movements until confirmation is complete. If the openness of both eyes drops below 0.35, it is considered closed. This lower threshold accounts for partial closures and ensures that the system distinguishes deliberate blinks from normal eyelid variations. When both eyes reopen above 0.35, the first blink is recorded. Once eye openness returns above 0.7, a 700 ms timer starts. The second blink is then executed in the same manner; since gaze data is already locked, the user does not need to maintain fixation on the target. If the second blink occurs within the timer threshold, the double blink is registered, and the selected target is highlighted. Regardless of whether the second blink occurs, the system resumes and allows new target selections.

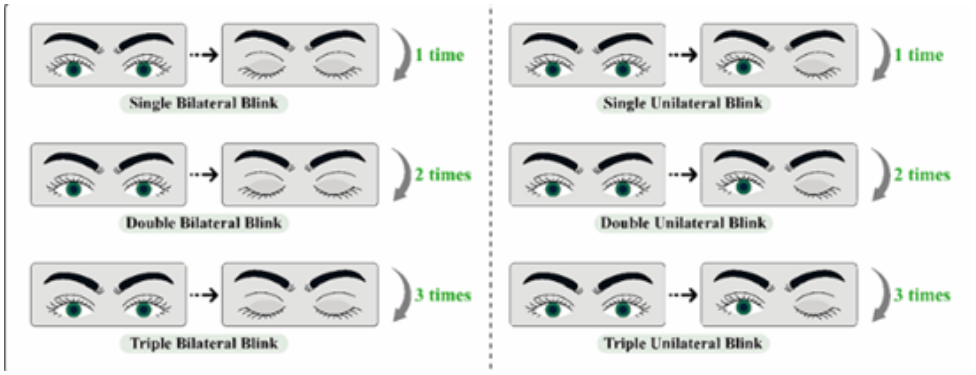


Figure 5.8: Winking variations trialed by Fan et al. [8].

The main strengths of Double Blink are its speed (slightly slower than Wink) and its lower demand for fine-motor eye control compared to winking. However, some users may find it challenging to synchronize both eyes or to fully reopen their eyes above 0.65 after the first blink, potentially requiring a third blink or restarting the process. These limitations may affect user comfort and fatigue, especially in complex, cluttered environments.

5.2 Metrics

To rigorously assess the performance of the developed techniques and address the research questions outlined in this thesis, a clear set of evaluation metrics were established. These metrics serve as the

link between the experimental results and the underlying research objectives, providing quantitative and qualitative evidence for refining and applying vergence-based interaction effectively in VR.

Specifically, the selected metrics were chosen to evaluate accuracy, response efficiency, and user comfort, aligning with the two central aims of this work. Metrics related to gaze and selection performance address RQ 2, determining whether the proposed vergence techniques enable precise and rapid selection of dynamic and partially occluded targets. Complementary measures, such as error rates, subjective workload, and indicators of visual fatigue, support RQ 3, revealing how interaction design can mitigate the effects of tracking inaccuracies and task complexity on user comfort.

To address RQ 2, each target will be associated with a limited number of selection attempts. Whether the participant successfully selects the target or not, the final attempt will be recorded for analysis. Each atom object is assigned a unique identifier, which will be logged along with the corresponding trial data. Additionally, a timer will record the duration taken per target, starting when the target appears and stopping when it is selected or when all attempts are exhausted. These measures allow evaluation of both speed and accuracy, offering insight into how efficiently and precisely participants perform target selections. For example, a participant may exhibit fast responses but low success rates, reflecting a trade-off between speed and precision. After each trial, participants will also provide a subjective rating of the technique's usability on a five-point scale (see fig. 5.9).

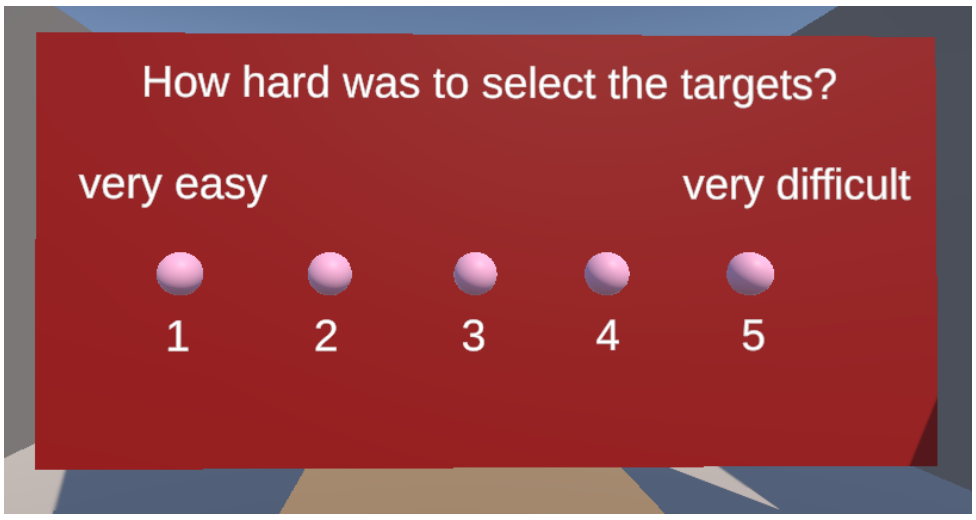


Figure 5.9: Rating Menu after a technique is tested to completion.

To address RQ 3, post-experiment questionnaires will be administered to assess visual and physical comfort. Participants will be asked about experiences of eye fatigue, general discomfort, dizziness, or nausea, providing direct indicators of how experimental conditions affected them. These subjective responses will be cross-referenced with both performance data and information gathered in the pre-experiment profile questionnaire, which records previous VR experience and any known vision impair-

ments. This comparison may reveal associations, such as inexperienced participants showing higher fatigue or longer selection times. Participants will be informed that they can interrupt the experiment at any time if they experience excessive discomfort.

To mitigate order effects such as learning, fatigue, or context bias, the sequence in which techniques are presented will be counterbalanced between participants according to their assigned user ID. This ensures that observed differences in performance are attributable to the methods themselves rather than the order in which they were experienced.

The following section outlines the experimental setup in which these evaluations were performed, detailing the hardware configuration, software environment, and data flow architecture that support the experiments.

5.3 Experimental Setup

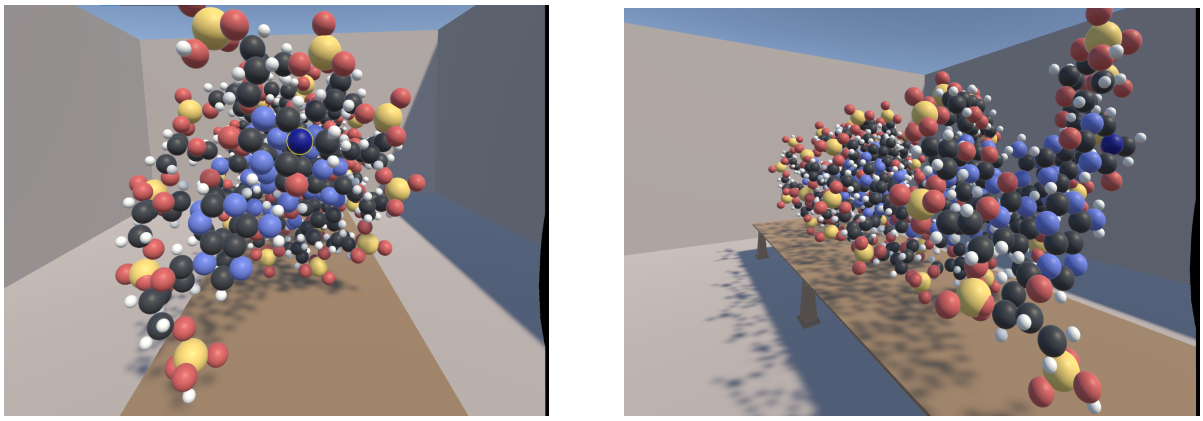
With four selection techniques and three confirmation techniques developed, the next step was to design a 3D environment that would enable the execution of demanding tasks to test their reliability. After considering several scenarios, a DNA molecule was chosen as a suitable test case. A detailed model was acquired from the Superhive marketplace[3] [69]. The model includes atoms of the chemical elements but not their connections, since only the atoms themselves are relevant as gaze targets. It was imported into the Unity project and colored using five distinct colors, each representing one of the elements found in DNA: carbon, hydrogen, oxygen, nitrogen, and phosphorus, as shown in fig. 5.10.

The experimental environment is illustrated in fig. 5.10. It consists of a simple cubic gray room without a roof, with the DNA model placed in the center. To reduce complexity, the model was cut in half, leaving two helices, which are sufficient for the tasks. A table was added beneath the model to enhance the sense of spatial presence, as VR environments often struggle to reproduce spatial perception at real-world scales. The virtual table is positioned to match the real table that will be present during laboratory experiments, thus strengthening the user's spatial recognition.

The DNA molecule is composed of multiple spheres, maintaining consistency with the standard use of spherical targets in gaze-tracking research. Besides aligning with established practice, this choice facilitates direct comparison with previous work, even though such studies typically use far less cluttered scenes. The spherical structure also ensures uniform visibility and minimizes orientation-related bias, while allowing the evaluation of gaze-based interaction in a more complex, cluttered visual environment.

In the final iteration of the experimental environment, the visual design was optimized to maximize the clarity of the target. The colors of the surrounding atoms were adjusted to reduce opacity, as shown in fig. 5.11, to ensure that the focus of the study remained on the quality of selection rather than the search

[3] Formerly known as Blender Market, Superhive is an independent online marketplace and community for Blender users to buy and sell 3D assets, add-ons, and training courses.



(a) DNA Molecule Model Front Review

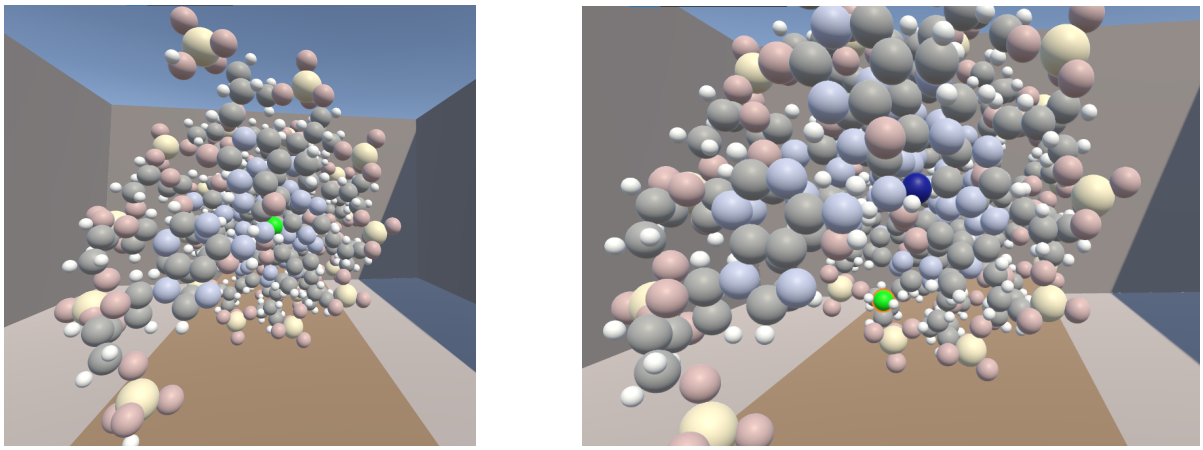
(b) DNA Molecule Model Side Review

Figure 5.10: DNA molecule inside the application from two angles: fig. 5.10(a) (front view) and fig. 5.10(b) (side view). Small white spheres represent hydrogen, medium black spheres represent carbon, medium blue spheres represent nitrogen, medium red spheres represent oxygen, and large yellow/orange spheres represent phosphorus.

task itself. To further enhance visibility, the current target is assigned a bright green-like hue, allowing it to remain recognizable even when partially occluded. In addition, shadows were removed to eliminate unnecessary visual cues that could otherwise distract participants from the task. Atom diameters in Unity were scaled to approximate relative sizes, hydrogen (0.0248 m), oxygen (0.0487 m), nitrogen (0.0499 m), carbon (0.0512 m) and phosphorus (0.0697 m), balancing two key goals: preserving the structural realism of the DNA molecule and creating a selection scenario that remained challenging enough to test the robustness of the techniques developed without being unreasonably difficult for participants.

Another improvement implemented in this final iteration concerns the visual highlighting of selected objects. As shown in fig. 5.11(b), the target highlight now appears as an orange outline scaled to 1.15 times the size of the original object. This configuration was chosen to maximize visibility while avoiding the occlusion of neighboring atoms. When using the Conecast technique to select groups of objects, all selected clones are displayed in orange. To preserve clarity and consistency, the outline color in this case was changed to the same bright purple used for the connecting line that links the selected clone to the previously focused object (see fig. 5.5).

The experimental setup, illustrated in fig. 5.12, is designed to ensure precise eye tracking and accurate rendering of the virtual environment. The experimental system is built around a standard PC equipped with 16 GB of RAM, an Intel Core i5-10400 CPU 2.90 GHz, and an NVIDIA GeForce RTX 2060 GPU , which is directly connected to the Varjo headset. This setup ensures adequate computational performance for real-time rendering of the virtual environment and reliable processing of gaze data. The Varjo Base software, installed on the PC along with SteamVR, facilitates setting up the environment, user calibration, and ongoing experiment monitoring. The PC sends the rendered frames



(a) DNA molecule with reduced-opacity colors

(b) Example of target atom with orange outline highlight

Figure 5.11: Final iteration of the DNA molecule model within the application. (a) shows the reduced-opacity environment designed to emphasize target visibility, and (b) illustrates the orange outline applied to selected targets.

of the virtual environment to the Varjo Aero HMD, which presents the scene to the participant on its integrated displays while simultaneously capturing eye-gaze data at up to 200 Hz. The headset functions effectively as both a display and a data-acquisition device, returning gaze information to the PC for processing.

Each participant in all experiments completed the Varjo Aero's built-in calibration procedure, which involves fixating on a point that gradually shrinks and disappears. Because the headset's internal calibration algorithms are not exposed to developers, it is not possible to determine precisely why some participants may have achieved poorer calibration than others. The procedure also guides users to position their eyes optimally for tracking, ensuring that both eyes remain clearly visible to the HMD's cameras.

To determine the precise location of the headset in physical space, the system uses the HTC Vive base stations, which track the HMD and provide positional feedback to the PC. This allows the virtual environment to maintain alignment with participants' movements, ensuring spatial consistency and accurate interaction. Data flows bidirectionally: the PC sends rendered frames to the Varjo HMD while receiving eye-gaze and positional data, as shown in fig. 5.12. This architecture represents the setup used for all experiments, supporting synchronized rendering, gaze tracking, and spatial alignment throughout testing.

Since testing all possible combinations in a single session would be too demanding (4 selection techniques × 3 confirmation techniques = 12 conditions), the evaluation was divided into two experiments. The first, referred to as the Confirmation Test, focuses on comparing the confirmation techniques to

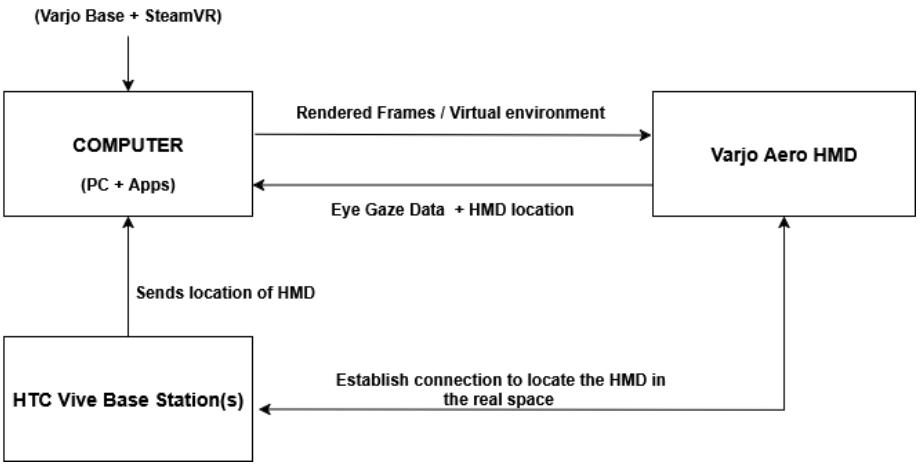


Figure 5.12: Architecture of the experimental setup using the Varjo Aero headset. The diagram shows the data flow between the computer, Varjo HMD, and HTC Vive base stations. The PC renders the virtual environment and manages gaze data through Varjo Base and SteamVR, while the HMD displays the scene and streams gaze information. The base stations provide spatial tracking for accurate headset positioning during the experiments.

identify the most effective and user-friendly option. The second, referred to as the Target Selection Test, uses the confirmation technique chosen in the Confirmation Test to evaluate all selection techniques. This stage aims to determine which technique is the most accurate, user-friendly, and efficient in terms of speed. Details of these experiments are presented in section 5.4 and section 5.5.

5.4 Confirmation Test

As introduced earlier, the purpose of this first experiment is to evaluate and compare the efficiency and comfort of the three developed confirmation techniques. The pilot test consisted of three blocks, each dedicated to one of these techniques. For all blocks, the GazeRay selection method was employed because of its simplicity and stability at the time of testing, as other selection techniques were still undergoing refinement.

Each block contained ten targets that differed in size, occlusion, and distance (around 0.3 to 1.5 m), introducing a range of difficulty levels. To maintain experimental consistency, the same target set was used across all techniques. However, their presentation order varied between blocks to prevent learning effects; this way participants could not rely on memorized target positions. The sequence for each block was randomly generated and adjusted until it exhibited no repeating spatial or pattern-based similarities.

The first five targets in each block were designated as practice trials, allowing participants to familiarize themselves with the system before data collection began. This phase helped participants adjust to both the VR environment and gaze-based interaction, which can initially feel unfamiliar or counterintuitive for those without prior experience. Between blocks, participants were given a short break accompanied

by on-screen instructions for the next technique. As shown in fig. 5.13, a pink sphere was placed below the instruction panel. Selecting and confirming this sphere using the current technique allowed participants to rehearse the interaction before starting the next block.

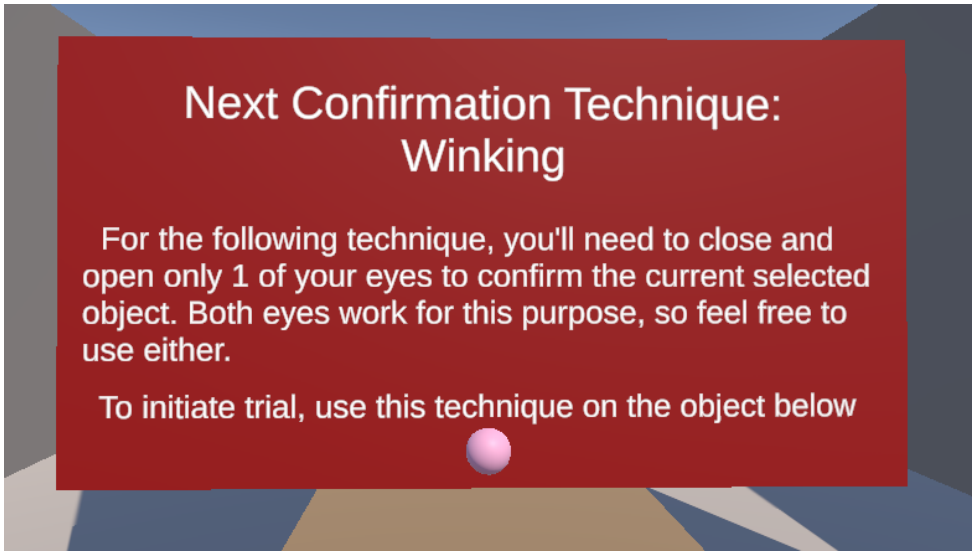


Figure 5.13: Example of the on-screen instructions presented between blocks, including the pink confirmation sphere used to rehearse the current technique (shown here with the Wink method).

If a participant was unable to perform a given technique after repeated attempts, the corresponding data were excluded from the analysis, following the criteria described in section 5.4.1. In cases where participants were unable to perform any of the techniques due to visual or motor limitations, such as uncorrected eye impairments, the session was terminated and discarded to ensure data integrity.

5.4.1 Participants

Before the pilot experiment, participants were asked to complete a brief profile questionnaire designed to assess their visual health and previous experience with virtual environments (see Appendix A). The primary screening question concerned the presence and severity of any eye impairment, ensuring that participants could effectively perceive objects in 3D space without significant visual limitations. Individuals reporting severe impairments, such as an inability to perceive spatial depth without glasses or lenses that would interfere with calibration (like rigid contact lens), were excluded from participation.

A total of 12 individuals (8 males and 4 females) took part in the study. Among them, six reported binocular visual conditions: four had both myopia and astigmatism, and the other two only astigmatism. The remaining six participants reported no known visual issues. Because Varjo Aero requires the absence of reflective surfaces for accurate eye calibration, participants wearing glasses were asked to remove them during the test. Contact lens users were allowed to keep them on unless the HMD detected calibration issues.

Regarding educational background, 10 participants held master's degrees, while the remaining 2 held bachelor's and doctorate degrees, respectively. Participants were also asked to identify their dominant eye, eight were dominant in the right eye and four in the left eye, allowing the analysis to explore whether dominance influenced target perception in the 3D environment.

Finally, the questionnaire assessed previous exposure to VR and 3D gaming environments. Most of the participants (10 out of 12) had used VR before, four more than five times and six fewer than five times, while only two had never used it. Experience with 3D video games was more varied: three participants had never played, four played less than once a year, two played about once per month, two played weekly and one played daily. This information provided a useful context for interpreting participant performance and adaptation to gaze-based interaction, as familiarity with virtual environments can influence both comfort and accuracy.

5.4.2 Data analysis

Data were analyzed using two complementary sources: the post-experiment questionnaire completed by participants (see appendix B) and the CSV file generated by the application, which recorded quantitative performance metrics during the test.

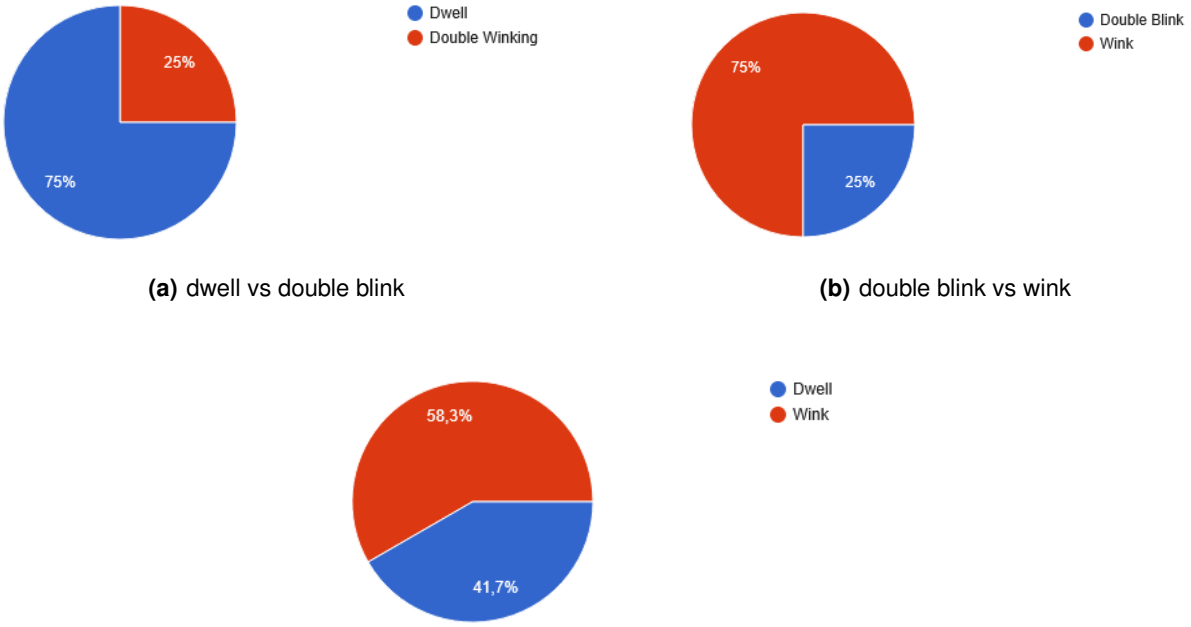
5.4.2.A Subjective data

Starting with the post-questionnaire, the results indicated that participants did not generally report eye fatigue or discomfort, aside from one who struggled with the Wink confirmation technique and another with Double Blink. This outcome was consistent with previous expectations, as Wink and Double Blink were hypothesized to be more demanding for certain users, possibly leading to frustration. Because this pilot primarily aimed to compare and identify the most effective confirmation technique, rather than to evaluate long-term comfort, questions related to fatigue or general discomfort were not included in this stage.

The first question asked participants to rate, on a scale from 1 (smaller targets) to 5 (larger targets), whether the size of the target influenced the difficulty of selection. Two participants did not see a difference (rating 3), while two rated 5, finding larger targets significantly easier to select; the rest rated 4, suggesting a moderate preference for larger targets. The following question focused on target distance, with 1 representing closer and 5 farther targets. The responses were more evenly distributed: Most of the participants rated between 1 and 3, indicating that closer targets were easier to select, while two rated 4 and 5, preferring farther targets. Three participants did not report noticeable differences. Overall, these findings suggest that the differences in target size and distance were perceptible but not substantial when GazeRay was used with the three confirmation techniques.

The third question assessed participants' perceived selection success, on a scale from 1 (never) to 7 (always). Most of the participants rated between 4 and 6 (two rated 4, six rated 5, and two rated 6), indicating a generally positive success rate. Two participants rated 2, corresponding to those who had limited VR experience or reported binocular impairments (myopia and astigmatism), both likely influenced their performance. Participants were also asked whether they noticed a consistent directional bias in their selections; half did not observe any pattern, while the remaining half reported slight drifts in the right, left, or downward directions in equal proportion.

The final three questions compared user preferences among the three confirmation techniques. As illustrated in fig. 5.14, participants consistently favored alternatives over double blink. Specifically, Dwell was preferred to Double Blink (fig. 5.14(a)), and in the comparison between Double Blink and Wink (fig. 5.14(b)), Double Blink again received the fewest votes. When directly comparing Dwell and Wink (fig. 5.14(c)), Wink emerged as the overall preferred. These preliminary results suggest that Wink may offer the most favorable balance between comfort and responsiveness, although further quantitative analysis of the CSV-recorded data is required to confirm these findings and objectively compare the techniques.



(a) dwell vs double blink

(b) double blink vs wink

(c) dwell vs wink

Figure 5.14: Comparison of user preference among the three confirmation techniques: Dwell, Double Blink, and Wink.

As mentioned in section 5.2, after completing each block, participants also provided a subjective rating of how well each confirmation technique performed. Consistent with the preferences above, par-

ticipants rated Dwell and Wink similarly (average scores of 1.92 and 2.00, respectively), while Double Blink was rated considerably worse (average score of 3.33). Analyzing these differences with Friedman's has shown significance (p = 0.0056), but after a post-hoc test with Wilcoxon and Bonferroni corrections, only Double vs Dwell is found statistically significant (p = 0.0234). Figure 5.15 illustrates the difficulty distribution from the post-task single easy question.

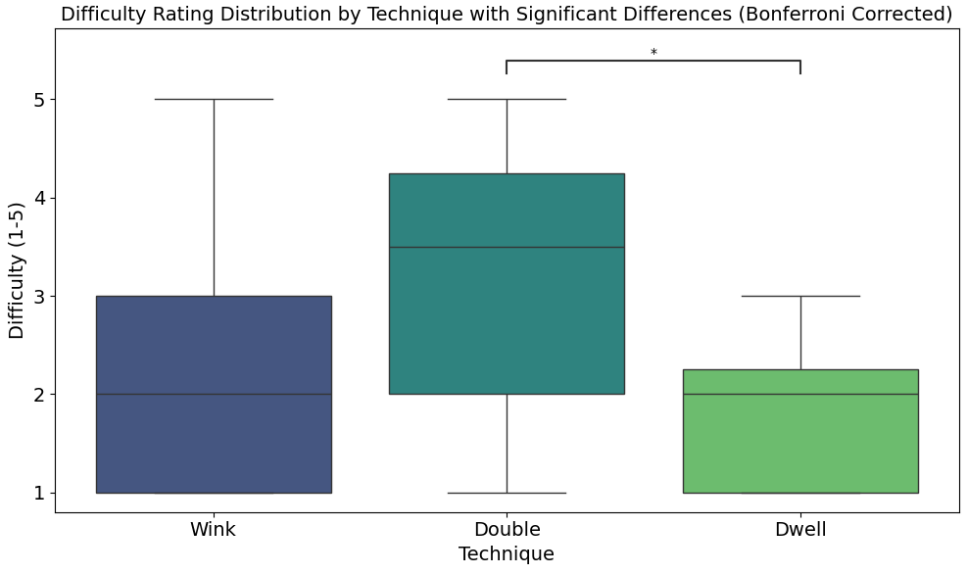


Figure 5.15: Difficulty in confirming the selection for each confirmation technique.

These results align with the preference trends observed in fig. 5.14 and reinforce the conclusion that Wink may be a preferred confirmation method for subsequent experiments.

5.4.2.B Time analysis per confirmation technique

We analyzed the log data to compare the time spent on each confirmation across the three techniques. We first removed invalid data corresponding to points where a confirmation was impossible (one participant could not wink) or took more than 30 seconds to perform. Those are degenerate cases, higher than 2 deviations above the mean, while the median is 3 seconds. Twenty-four out of 180 samples were removed. Their distribution among conditions is: Double=10, Dwell=8, Wink=6.

A Shapiro-Wilk test on the remaining 156 samples has shown that the time data cannot be assumed as normally distributed for at least one group. So, we proceeded using the Friedman test, which indicated a statistically significant overall difference in 'Time' across the three techniques (p = 0.0039). We then performed pairwise Wilcoxon signed-rank tests, which revealed potential differences.

After applying the Bonferroni correction for multiple comparisons, the pairwise comparisons revealed statistically significant (alpha = 0.05) differences for $Wink\ vs\ DoubleBlink$ and $DoubleBlink\ vs\ Dwell$ (p-value = 0.0019). We found no statistical significance for the difference between Wink and Dwell.

Figure 5.16 presents the magnitude of these differences, from which we conclude that $DoubleBlink$ takes more time to confirm than the other two techniques.

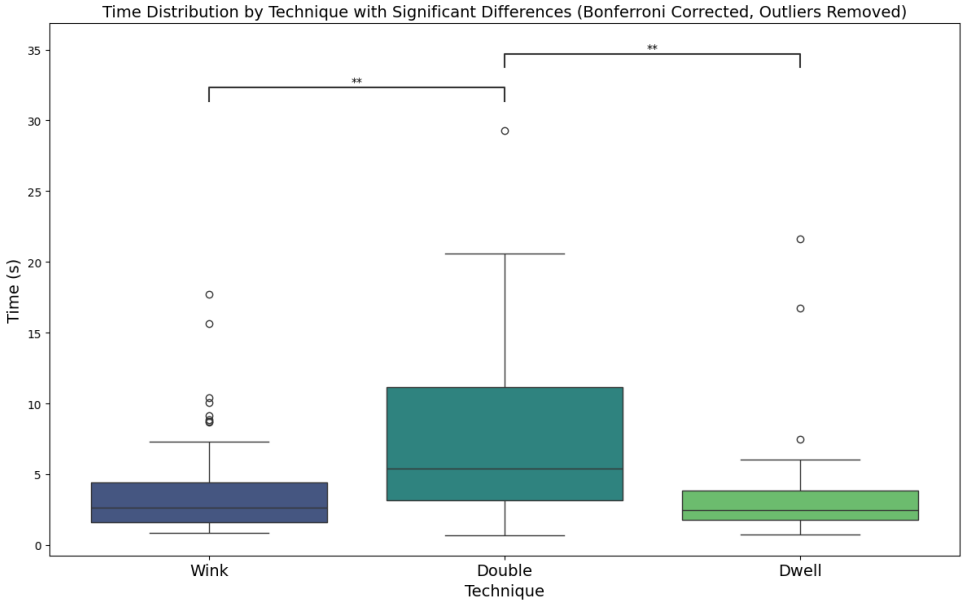


Figure 5.16: Time analysis for each confirmation technique.

5.4.2.C Selection accuracy per confirmation technique

As different confirmation techniques can affect selection accuracy due to eye motion necessary to confirm, we analyzed the confirmation techniques accuracy in target selection.

Table 5.1 show the accuracy obtained with each confirmation technique. We observed a similar accuracy for all three techniques, around 70% of correct selections. A Cochran's Q test indicated that there is no significant differences (p-value: 0.3678) in accuracy throughout confirmation techniques.

Table 5.1: Correct and Incorrect Selections per Confirmation Technique

Technique	Incorrect (False)	Correct (True)	Accuracy (%)
Double	15	40	72.73%
Dwell	18	37	67.27%
Wink	14	41	74.55%

5.4.3 Discussion

Based on the confirmation test results, Wink emerged as the most promising confirmation technique, combining responsiveness with a relatively low rate of false activations. However, the Double Blink method will be retained as a backup option for subsequent tests, providing an alternative in cases where the eye-tracking system fails to consistently detect single-eye closures.

Although Dwell demonstrated comparable subjective performance to Wink, it was ultimately excluded from further experimentation due to compatibility constraints. Specifically, Dwell confirmation cannot be reliably integrated with ConeCast without introducing implementation inconsistencies. Supporting both methods simultaneously would require participants to switch between two distinct confirmation mechanisms across four selection techniques, resulting in an unnecessarily complex interaction scheme.

From a usability standpoint, both Wink and Dwell proved generally reliable, though neither achieved perfect confirmation accuracy. Dwell offers an advantage in terms of execution, requiring minimal physical effort and no explicit gesture, but this convenience comes at the cost of precision and user control. It is more susceptible to unintentional activations and may struggle with maintaining selection stability, particularly when users are required to focus on a target for extended periods. Although this limitation was less pronounced when paired with RayCast, it could become problematic in cluttered or occlusion-heavy environments.

In terms of timing, both Dwell and Wink achieved comparable results, with similar average confirmation times and faster completion than Double Blink (see fig. 5.16). However, when analyzing accuracy, the results were slightly reversed: both Dwell and Wink showed marginally lower success rates compared to Double Blink, as shown in table 5.1. This suggests that, although Double Blink was less preferred and generally slower, likely due to failed attempts requiring more time, it nevertheless provided a reliable means of confirming selections. This finding reinforces its role as a suitable backup technique for the final test.

In summary, Wink provides a balanced trade-off between control, comfort, and robustness, making it the preferred confirmation technique for the final experimental phase of the experiment. Its combination of execution speed, minimal fatigue, and adaptability between selection methods positions it as the most practical option forward.

5.5 Selection Test

Having concluded through the data analysis in the previous section that Wink demonstrated the best overall performance among the three confirmation techniques, the selection test was conducted. The objective of this experiment was to evaluate the remaining interaction techniques, the selection methods. Similar to the pilot study, the test was divided into four blocks, each corresponding to one selection technique. For all blocks, Wink served as the default confirmation method, unless a participant was physically unable to perform a wink, in which case Double Blink was used as a substitute.

Each block followed the same general structure as described in section 5.4, containing ten targets that varied in size, occlusion, and distance (around 0.3 to 1.5 m) to introduce a range of difficulty levels. To maintain experimental consistency, the same set of targets was used across all techniques. However,

their presentation order differed between blocks to minimize learning effects and prevent participants from anticipating target positions. The order for each block was randomly generated and adjusted to avoid spatial or sequential repetition.

A tutorial section was designed to allow participants to familiarize themselves with each selection technique and the confirmation process. As shown in fig. 5.17, this tutorial was divided into two parts: in the first, participants practiced the single-step selection techniques once each; in the second, they practiced the two-step selection technique (ConeCast). This division was made because the single-step techniques shared a similar user experience, whereas ConeCast involved a distinct two-phase interaction. In total, the tutorial included six selections, giving participants time to adapt to both the VR environment and gaze-based interaction, which may initially feel unfamiliar for those without prior experience.

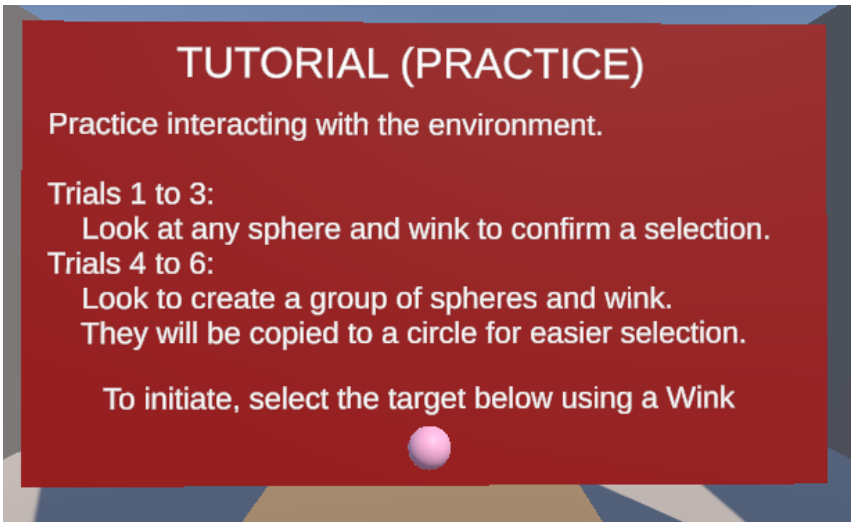


Figure 5.17: Tutorial menu at the start of the experiment, explaining the initial training section for the different selection techniques.

Between blocks, participants were given a short break accompanied by on-screen instructions for the next technique, which remained consistent across all blocks. As shown in fig. 5.18, a pink sphere was placed below the instruction panel. Selecting and confirming this sphere using the current technique allowed participants to rehearse the interaction before continuing with the next block.

At the beginning of each session, participants were briefed on the experimental procedure, including the calibration process required by the Varjo headset. During the tutorial phase, if object selection appeared unstable, the calibration was repeated until satisfactory accuracy was achieved. If a participant was unable to perform a given technique after repeated attempts, using either Wink or Double Blink as the confirmation method, the corresponding data were excluded from analysis. In cases where participants were unable to perform any of the techniques due to visual or motor limitations, the session was terminated and discarded to ensure data integrity.

Figure 5.18: Start of a block in the experiment with the instructions for the following task.

5.5.1 Participants

Before the selection experiment, participants were asked to complete a brief profile questionnaire designed to assess their visual health and prior experience with virtual environments (see Appendix C), which began by instructing participants to read and sign an informed consent form outlining the study's purpose, procedures, and their rights as participants (see Appendix E). The primary screening question addressed the presence and severity of any eye impairments to ensure that all participants could effectively perceive objects in 3D space without significant visual limitations. Individuals reporting severe conditions, such as an inability to perceive spatial depth without glasses or the use of rigid contact lenses that could interfere with calibration, were excluded from participation.

A total of 32 individuals (27 males, 5 females) took part in the study. Among them, thirteen reported binocular visual conditions: eight had only myopia; two had both myopia and astigmatism, with one of them also affected by color blindness (dyschromatopsia); two others were affected by color blindness, one of them also having myopia; and one participant had hyperopia. The remaining nineteen participants reported no known visual issues. Because the Varjo Aero requires the absence of reflective surfaces for accurate eye calibration, participants wearing glasses were asked to remove them during the test. Contact lens users were allowed to keep them on unless the HMD detected reflective interference.

The group presented a diverse educational background: thirteen participants were pursuing a bachelor's degree, eight had already completed one, eight held a master's degree, one held a doctorate, and two had completed high school (12th grade). Participants were also asked to identify their dominant eye, seventeen were right-eye dominant and fifteen left-eye dominant, allowing for a potential analysis of whether ocular dominance influenced target perception in a 3D space.

Given that the pilot experiment indicated that contact lenses and corrective procedures could affect

performance for some users, participants were specifically asked about these conditions. Two reported wearing soft contact lenses during the experiment, one used toric lenses for astigmatism, and one had undergone surgical vision correction (LASIK, PRK, or LASEK).

Finally, the questionnaire evaluated previous experience with VR and 3D gaming environments. Most of the participants (31 of 32) had used VR before, nine more than five times, and twenty fewer than five times. Although only one had never tried it, there were two who reported regular use. The experience with 3D video games was similarly varied: three had never played, ten played less than once a year, three about once per month, eight weekly, and eight daily. This information provided a useful context for interpreting participant performance and adaptation to gaze-based interaction, as prior familiarity with virtual environments can influence both comfort and accuracy.

5.5.2 Data analysis

Data were analyzed using two complementary sources: the post-experiment questionnaire completed by participants (see appendix D) and the CSV file generated by the application, which recorded quantitative performance metrics during the test.

5.5.2.A Subjective data

The first section of the post-questionnaire addressed participants' well-being following the experiment. Participants were asked to rate general discomfort, eye fatigue, dizziness, and nausea on a scale from 1 (none) to 5 (severe). The results indicated that almost all participants experienced minimal symptoms: only two reported slight nausea and another one slight dizziness. Nineteen participants reported no discomfort at all, while eleven noted mild discomfort, and two rated moderate to high (3 and 4). Eye fatigue was more common, with six participants rating 3 and thirteen rating 2, suggesting mild strain. Twelve participants reported no eye fatigue and one experienced severe fatigue, probably related to calibration difficulties encountered during setup. Overall, these results indicate that the experiment caused little to no discomfort, and isolated fatigue cases may be attributed to suboptimal calibration or lack of familiarity with eye-tracking interfaces.

The second section of the questionnaire focused on participants' perceptions of the experimental techniques. The first two questions asked participants which block they believed performed best and worst. As expected from the limitations discussed in section 5.1.1.A, OverlapSphere was overwhelmingly identified as the least effective technique. In contrast, opinions on the best performing technique were more distributed: nine participants selected Raycast, ten selected SphereCast, and thirteen chose ConeCast, making the latter the most favored method overall. These results will be further elaborated upon through the analysis of the CSV file containing the experiment data.

Participants were also asked to rate, on a scale from 1 (smaller targets) to 5 (larger targets), whether the size of the target influenced the difficulty of selection. Half of the participants (16) did not perceive a difference (rating 3), while three rated 5 and ten rated 4, finding larger targets easier to select. Three rated 2, indicating that some found smaller targets slightly easier to select. The following question addressed the target distance, with 1 representing closer and 5 farther targets. Most participants rated 1 (9 users) or 2 (16 users), suggesting that closer targets were generally easier to select, while three slightly preferred farther ones. Four participants did not report noticeable differences. In summary, target size had a limited impact on perceived difficulty, whereas distance showed a clearer effect: closer targets were typically preferred.

The final perception-related questions addressed selection accuracy and directional bias. Participants rated their perceived success in selecting targets on a scale from 1 (never) to 7 (always). Most responses clustered between 4 and 6 (three rated 4, twelve rated 5, and ten rated 6), indicating overall satisfaction and confidence in their selections. Three participants rated 2 and another four rated 3, although no consistent pattern emerged linking these lower ratings to prior VR experience or visual conditions, suggesting the issue could be related to calibration difficulties. Finally, participants were asked whether they noticed any systematic drift or bias in their selections. The majority (20) reported none, while the remaining twelve mentioned slight directional tendencies: four toward the left, three upward, three downward, one to the right, and one who reported both rightward and upward shifts.

As mentioned in section 5.2, after completing each block, participants also provided a subjective rating of how well each selection technique performed. Consistent with the preferences described above, participants rated RayCast and SphereCast similarly, with average scores of 2.16 and 2.22, respectively. ConeCast achieved the best overall rating, with an average score of 1.91, while OverlapSphere was rated considerably lower, with an average of 3.87. Analyzing these differences with Friedman's has shown significance ($\chi^2(3) = 52.89,\ p = .0000$), and after a post-hoc test with Wilcoxon and Bonferroni corrections, there were three pairs with significance: OverlapSphere vs ConeCast (p = 0.0000); OverlapSphere vs RayCast (p = 0.0000); OverlapSphere vs SphereCast (p = 0.0000). Figure 5.19 illustrates the difficulty distribution from the post-task single easy question. These results align with the previously discussed preference trends and reinforce the conclusion that ConeCast was the most favored technique, closely followed by RayCast and SphereCast.

5.5.2.B Selection time

Before analyzing the time to select per selection technique, we observed the raw data and removed extreme outliers (52 out of 1280 samples). These are time samples more than 2 deviations above the mean. They occurred due to a calibration issue during the experiment, where the participant accidentally moved the headset and had an interruption to recalibrate. This allowed the task to continue without

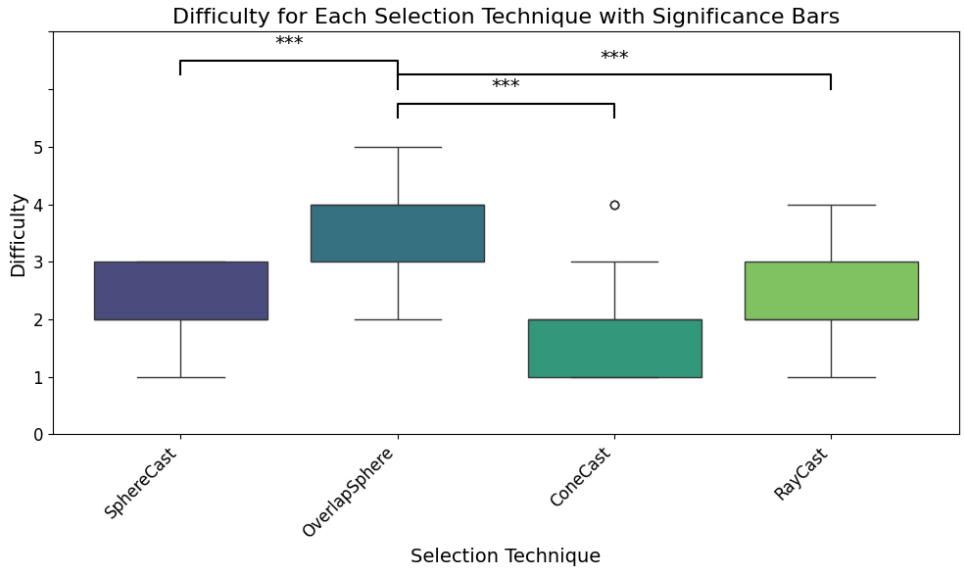


Figure 5.19: Difficulty in selecting targets for each selection technique.

Table 5.2: Accuracy per Selection Technique

Selection Technique	Accuracy
ConeCast	0.9875
OverlapSphere	0.6937
RayCast	0.9344
SphereCast	0.9187

problems for other metrics after calibration, but recorded extended times.

As illustrated in Figure 5.20, the OverlapSphere technique requires more time from the user than the other techniques, which are not significantly different among them. Since the data were not normally distributed and we had a within-subjects design, we used the Friedman test for the analysis. The Friedman test result (p = 0.0000) showed a statistically significant difference in 'Time' among the different Selection Techniques.

To determine which specific Selection Techniques were different from each other, we performed pairwise comparisons using Wilcoxon signed-rank tests with a Bonferroni correction for multiple comparisons. ConeCast vs OverlapSphere (p = 0.0033), OverlapSphere vs RayCast (p = 0.0000), OverlapSphere vs SphereCast (p = 0.0028) all displayed statistically significant differences. There were no statistically significant differences in time between ConeCast and RayCast, ConeCast and SphereCast, or RayCast and SphereCast after applying the Bonferroni correction.

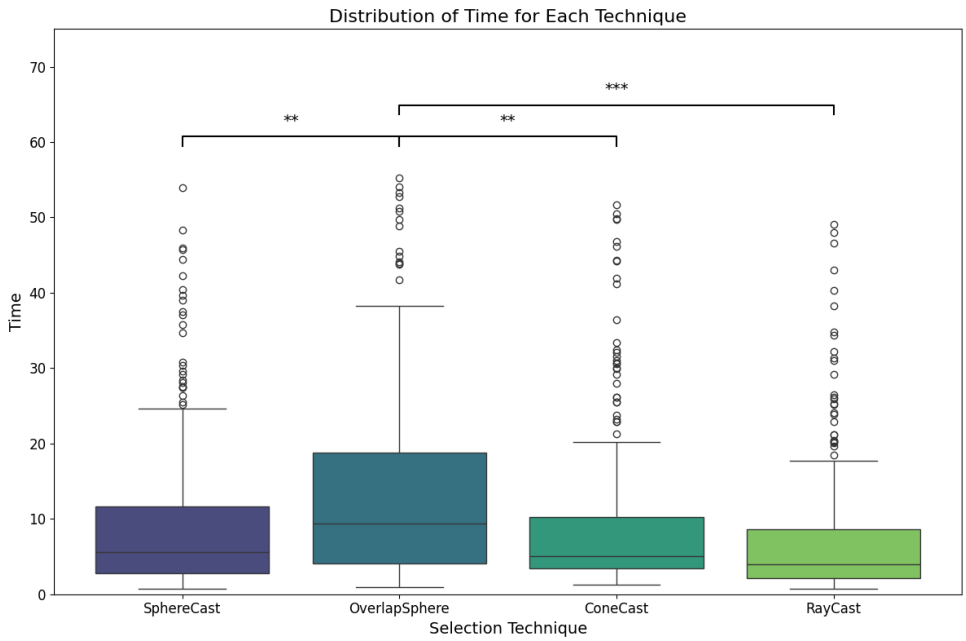


Figure 5.20: Time analysis for each selection technique. Significance bars are after Bonferroni correction.

5.5.2.C Selection accuracy per selection technique

We computed the accuracy for each technique as depicted in Table 5.2 and used a Cochran's Q test in the analysis, which indicated significance (p = 0.0000). Then, we performed pairwise McNemar's tests with Bonferroni correction for multiple comparisons. We found several significant differences with OverlapSphere, which has the poorest accuracy, and between ConeCast and SphereCast (Table 5.3). ConeCast has the highest mean accuracy, but we could not find a significant difference after Bonferroni correction to Raycast (p-value (uncorrected)=0.0147, p-value (Bonferroni)=0.0881), although for SphereCast there was a noticeable difference (p-value (uncorrected)=0.0021, p-value (Bonferroni)=0.0128). Figure 5.21 illustrates the distribution.

Table 5.3: Significant pairwise comparison p-values between selection techniques (uncorrected and Bonferroni-corrected).

Comparison	p-value (Uncorrected)	p-value (Bonferroni)
ConeCast vs OverlapSphere	0.0000	0.0000
ConeCast vs SphereCast	0.0021	0.0128
OverlapSphere vs RayCast	0.0000	0.0003
OverlapSphere vs SphereCast	0.0000	0.0000

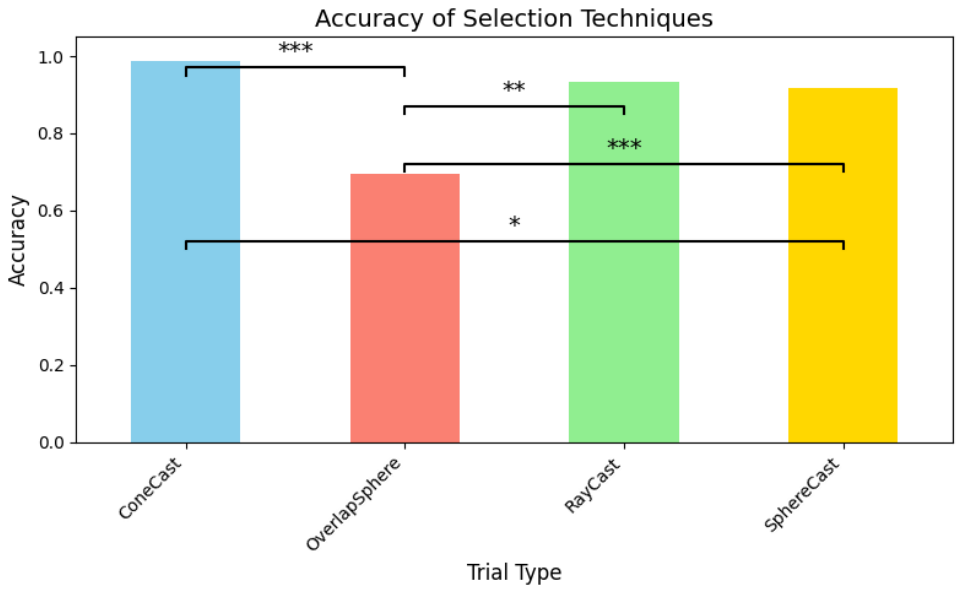


Figure 5.21: Accuracy analysis for each selection technique. Significance bars are after Bonferroni correction.

5.5.2.D Number of tries per selection technique

Participants could perform up to three tries to make a correct selection. We analyzed the number of tries required for each technique. Figure 5.22 illustrates this comparison highlighting the significantly different techniques. Significances, indicated in table Table 5.4 were computed from a Friedman test conducted across four selection techniques. The test was statistically significant, $\chi^2(3) = 60.73,\ p = .0000$, indicating that at least one condition differed from the others.

Table 5.4: Pairwise Wilcoxon Signed-Rank Test results on *Nbr. of Tries*, with Bonferroni correction.

Comparison	p-value (Uncorrected)	p-value (Bonferroni)	
ConeCast vs OverlapSphere	0.0000	0.0000	
ConeCast vs RayCast	0.0004	0.0025	
ConeCast vs SphereCast	0.0000	0.0001	
OverlapSphere vs RayCast	0.0000	0.0000	
OverlapSphere vs SphereCast	0.0000	0.0000	
RayCast vs SphereCast	0.0276	0.1657	n.s.

5.5.2.E Analysis of wrong selections per selection technique

Out of 1280 samples, 149 are wrong selections. For those, we recorded the position of the target selected by mistake and computed the distance to the correct target. Here, we study how this distance varies over techniques. Table 5.5 lists the count of wrong selections per technique and the average distances from the selected object to the target.

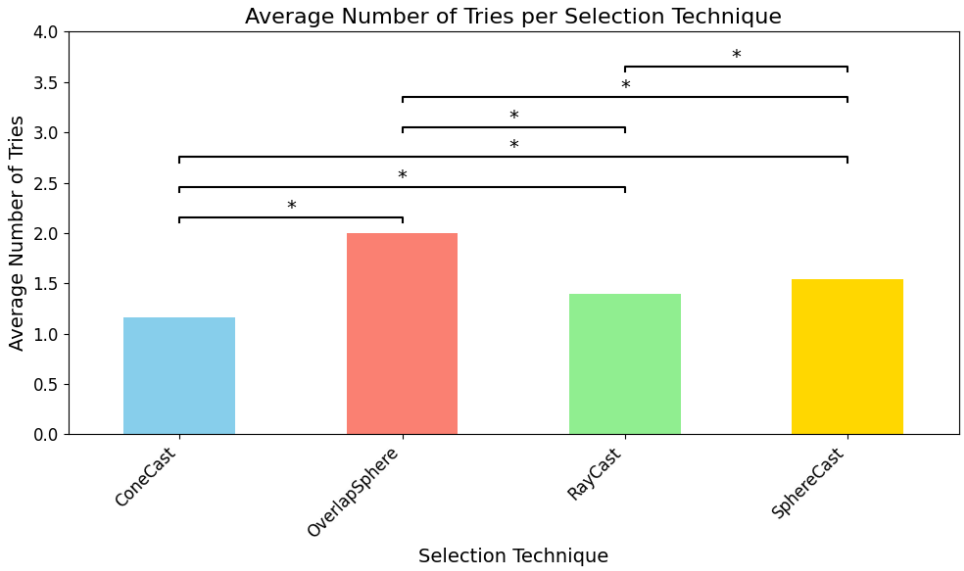


Figure 5.22: Analysis of the number of tentatives for each selection technique. Significance bars are after Bonferroni correction.

Table 5.5: Summary of wrong selections per selection technique for wrong selections.

Trial	Count	Average Distance
OverlapSphere	98	0.1831
SphereCast	26	0.2173
RayCast	21	0.2552
ConeCast	4	0.2625

We noticed that OverlapSphere is responsible for the majority of wrong selections, while ConeCast caused only 4 wrong selections. Yet, the differences in distance are not significant among techniques as a Friedman test revealed ($\chi^2(3) = 3.00,\ p = 0.3916$).

5.5.3 Discussion

The results of the selection test highlight clear distinctions in the performance and usability of the four gaze-based selection techniques. ConeCast demonstrated superior accuracy and consistency, suggesting that its two-stage confirmation process effectively mitigates common sources of error in gaze-based interaction, such as micro-saccadic drift or small calibration offsets. Although this increased reliability comes at the cost of execution speed, the trade-off appears acceptable in contexts where precision is prioritized over immediacy, such as object manipulation or interface control in dense virtual environments.

The similar performance between SphereCast and RayCast indicates that both techniques offer a practical balance between speed and accuracy. Their single-step interaction models make them inher-

ently faster and more fluid to use, particularly when the eye-tracker calibration is stable. Under well-calibrated conditions, both methods achieve near-optimal accuracy and minimal selection effort, making them ideal for rapid target acquisition tasks. However, their performance is more sensitive to calibration drift or tracking noise compared to ConeCast. The marginal advantage of SphereCast in average number of tries suggests that its volumetric interaction zone provides greater tolerance for gaze imprecision, a factor that becomes beneficial in longer sessions or for users exhibiting minor vergence instability.

The consistently poor performance of OverlapSphere further underscores the importance of accurate depth handling in vergence-based selection. Its lack of spatial flexibility likely amplified the impact of vergence estimation errors, particularly when the targets were placed in heavily cluttered spaces. These results are consistent with findings by Duchowski [5] and Arefin et al. [2], who highlighted the limitations of purely geometric or intersection-based gaze models when depth uncertainty increases. This evidence reinforces the need for depth-adaptive or error-compensated selection mechanisms that can better accommodate natural variations in vergence precision.

The general trend of higher accuracy for closer targets suggests that vergence-based estimation remains more stable at shorter fixation distances, where small eye alignment errors result in minimal positional deviation. Although accuracy tends to decrease for more distant targets, recent advances in gaze modeling and adaptive calibration have mitigated much of this issue. Techniques such as volumetric interaction zones and multistage models such as ConeCast help maintain precision by dynamically adjusting depth thresholds. Similar improvements have been reported in recent studies (see, e.g. [2,20]), demonstrating that adaptive vergence correction and gaze–head fusion can significantly reduce depth-related drift. Nevertheless, maintaining consistent precision across extended distances continues to be a subtle but relevant challenge for depth-based gaze interaction systems.

In summary, ConeCast stands out as the most accurate and reliable technique, offering the best balance between control and consistency, even though at a slight cost in execution speed. SphereCast and RayCast thrive under precise calibration conditions, providing faster and more natural interaction for well-tuned systems, while OverlapSphere's performance limitations make it unsuitable for practical deployment. These findings suggest that no single approach is universally optimal; instead, future gaze-based selection systems should dynamically adapt interaction thresholds and model parameters based on calibration quality, gaze stability, and target distance. Such hybrid adaptability could pave the way for more user-centered, context-aware gaze-based selection methods in immersive environments.

6

Discussion

This chapter summarizes the main findings of the research and discusses how each stage contributed to the development of a robust and adaptive vergence-based interface for gaze interaction in VR. The work evolved through three main stages: vergence evaluation, confirmation testing, and selection testing, each designed to isolate and refine a different aspect of the interaction process. Conducting these studies separately allowed for controlled analysis of multiple variables, such as depth precision, confirmation reliability, and selection stability, which would have been difficult to disentangle within a single, more complex experiment. Different participants were recruited for each experiment, ensuring that learning effects from repeated exposure to the gaze-based selection interfaces did not influence the results.

The initial vergence study laid the technical foundation by assessing the reliability of the gaze direction and depth derived from vergence. The vergence algorithm consistently produced horizontal and vertical gaze errors below 1%, with an average MAE of 0.29% and 0.32%, respectively, both outperforming the native tracking accuracy of the Varjo system (0. 85% and 0. 39%). However, the accuracy of depth estimation decreased with distance, showing a higher error variability for farther targets and between participants. These differences were particularly pronounced among users with vision problems when corrective lenses were removed, underscoring the role of individual visual characteristics and weak spatial cues in depth perception.

Although the gaze direction proved consistently accurate and robust, these depth inconsistencies

revealed a systematic sensitivity to small alignment errors and individual variation. As a result, raw vergence data alone were deemed insufficient for stable selection, leading to the introduction of a margin of acceptable error (MAE) to make depth-based targeting more tolerant and usable. This finding directly informed the design of subsequent techniques by defining the limits of vergence reliability in practical interaction scenarios.

Following the vergence study, the investigation shifted away from dynamic environments. Preliminary explorations indicated that incorporating moving targets or continuously changing scenes would introduce additional variability without yielding new insights into the core performance of vergence-based selection. Consequently, the evaluation was conducted in static yet heavily cluttered environments, providing a controlled setting to stress-test the interaction techniques. Participants were allowed to move laterally within the virtual space to adjust their perspective, but not closer to the targets, ensuring consistent depth conditions while still challenging the robustness, accuracy, and usability of each method under realistic spatial constraints. This approach simplified the experimental design while enabling a rigorous assessment of technique performance in complex 3D scenes.

Given the lack of reliability in depth estimation, four gaze-based selection techniques were developed to explore different ways of compensating for vergence limitations. OverlapSphere was implemented as a baseline technique without explicit depth consideration, providing a point of comparison against depth-aware methods. GazeRay and SphereCast incorporated vergence depth in distinct ways through direct ray projection and volumetric interaction zones, respectively, while ConeCast introduced a two-stage simplified targeting process, designed to enhance stability at the cost of additional user input.

Similarly, because there was no clearly established confirmation method for gaze-based selection, three techniques were designed to address different user needs. Dwell offered a simple, fatigue-free option suited for baseline comparison, Wink provided fast and intuitive activation but required precise eye control, and Double Blink presented a less demanding alternative that traded execution speed for reliability.

Since testing all possible combinations of selection and confirmation techniques simultaneously would have introduced too many variables, the evaluation process was divided into two phases: one dedicated to confirmation methods and another focused on selection performance.

Building on this foundation, the confirmation test focused on developing reliable and comfortable methods to be paired with selections. Quantitative analysis revealed that Wink and Dwell achieved comparable confirmation times, both significantly faster than Double Blink ($p = 0.0029$), while the precision remained similar across the three techniques, around 70%. Despite this similarity in performance, participants consistently rated Wink as the most preferred method, citing its balance of control and comfort, although some of the users reported having an easier time using Double Blink, due to the less demanding method of execution, although it was slower. Subjective ratings reinforced these trends, with

Wink and Dwell receiving average difficulty scores of 2.00 and 1.92, compared to 3.33 for Double Blink (Friedman's test, $p = 0.0056$).

Although Dwell demonstrated competitive timing and comfort, it was ultimately excluded from subsequent stages due to incompatibilities with ConeCast, which would have required distinct confirmation mechanisms and introduced unnecessary cognitive overhead. These results confirmed that Wink offers the most favorable trade-off between responsiveness, precision, and user comfort, while Double Blink remains a suitable fallback. However, it should be noted that these findings are based on a limited participant sample ($n = 12$), and slight deviations from observed trends may emerge in larger or more heterogeneous populations. This phase also emphasized that the effectiveness of gaze-based confirmation depends as much on ergonomic and perceptual factors as on technical accuracy, underscoring the importance of designing interaction mechanisms that align with both user capability and system limitations, rather than optimizing each component in isolation.

Finally, the selection test integrated all prior insights into a complete interaction workflow, evaluating four vergence-based selection techniques. Quantitative results confirmed that ConeCast achieved the highest accuracy (98.2%) and reliability, requiring fewer retries than any other method ($p<0.001$), but at the cost of slightly longer selection times compared to RayCast and SphereCast. These two demonstrated similar high accuracy (92.7%) and faster, more fluid performance when the calibration was stable, confirming that these techniques perform best under precise tracking conditions. OverlapSphere, in contrast, exhibited the lowest accuracy (70.9%) and required significantly more time and attempts to complete selections ($p<0.001$), reaffirming that depth awareness is essential for maintaining performance in cluttered 3D environments.

ConeCast's design also revealed a potential scalability limitation: as target distance increases, the number of intersecting spheres within the selection cone grows exponentially in cluttered scenarios, occasionally resulting in dozens of simultaneous collisions. Implementing a constraint on the maximum number of detected spheres or dynamically adjusting the cone's spread could mitigate this issue and preserve efficiency at larger depths and clarity to the selected targets' recreation.

Subjective ratings mirrored these results: participants overwhelmingly rated ConeCast as the most effective (mean = 2.00), followed closely by RayCast (2.18) and SphereCast (2.14), while OverlapSphere received the poorest evaluation (3.91, $p<0.001$). Fatigue and discomfort remained minimal under all conditions, indicating that the differences in performance stemmed primarily from interaction design rather than usability strain.

In general, these findings validated the design progression—from tolerating vergence imprecision to balancing control and speed—as an effective approach toward robust and adaptable gaze-based interaction in 3D environments. However, since this experiment involved 22 participants, the results should be interpreted with some caution, acknowledging that larger-scale studies may further refine

these observations.

The research demonstrates that no single technique can guarantee optimal performance across all conditions. Instead, gaze-based interaction benefits from adaptive hybrid approaches that adjust parameters dynamically based on calibration quality, gaze stability, and spatial context. This conclusion aligns with recent findings in gaze-interaction literature (e.g., [2, 20]), which emphasize dynamic calibration, gaze–head fusion, and depth-adaptive thresholds as essential components of future VR interfaces.

In summary, this thesis advances the understanding of vergence-based interaction by (1) confirming the robustness of gaze direction but the fragility of vergence depth, (2) identifying Wink as the most effective confirmation technique for immersive contexts, and (3) validating ConeCast as a reliable, high-accuracy selection method built upon these insights. Together, these contributions form a coherent framework for designing gaze-based interaction systems that are precise, user-friendly, and adaptive to individual and environmental variability.

7

Conclusions

Contents

This chapter concludes the thesis by incorporating the main findings and contributions of the investigation. It is organized into three sections: the first discusses the limitations identified during the development and evaluation of the system, the second presents the overall conclusions drawn from the study, and the final section outlines possible avenues for future work to expand upon the results achieved here.

7.1 Limitations

Despite the positive outcomes observed across the different experimental stages, some limitations should be acknowledged. The first concerns the number of participants in each study, with eleven in the vergence study, twelve in the confirmation test and twenty-two in the selection test. Although these

samples were sufficient for exploratory analysis, they restrict the statistical power and potential to generalize the findings. Participant diversity was also limited, particularly regarding visual conditions and VR experience, which may have influenced vergence stability and selection accuracy.

The second limitation relates to accuracy and stability of the calibration. The precision of vergence estimation was inherently bounded by the HMD's internal calibration and the tracking quality of the Varjo system. Even minor drift or misalignment during extended use could have affected depth estimation, especially for distant targets. A more refined or adaptive calibration method could help mitigate these issues and improve depth consistency across sessions.

Hardware and software constraints also played a role. The implementation relied on a specific HMD model and the Unity engine, both of which impose limitations in sampling frequency, data synchronization, and gaze model transparency. As such, the results may not fully generalize to other devices with different tracking capabilities or internal gaze estimation algorithms.

Another limitation stems from the experimental scope. All studies were conducted in controlled laboratory conditions using static tasks and predetermined object layouts. Although this ensured consistency among participants, it limited the evaluation of dynamic or real-world interaction scenarios.

Finally, the design of the interaction techniques themselves introduced certain trade-offs. For example, ConeCast may require additional constraints to limit the number of simultaneously detected targets, as distant objects can produce excessive overlap within the cone selection volume. Similarly, confirmation techniques such as Wink and Double Blink depend on individual oculomotor control and camera sensitivity, which may vary between users or hardware platforms.

7.2 Conclusion

This work set out to investigate how vergence-based gaze interaction could enhance precision, comfort, and robustness in virtual reality environments, particularly under visually cluttered and depth-ambiguous conditions. Through a structured sequence of studies, beginning with vergence evaluation, followed by confirmation and selection testing, the research systematically explored the reliability of vergence estimation, the effectiveness of gaze confirmation strategies, and the overall usability of depth-aware gaze interaction models.

The findings of the vergence evaluation demonstrated that while vergence can provide valuable depth cues, its accuracy remains sensitive to calibration precision, viewing distance, and individual visual differences. Nevertheless, the adopted vergence model, adapted from Duchowski's approach [5], proved to be sufficiently consistent to support the development of stable gaze-based selection techniques. This directly addressed my first research question, showing that vergence techniques can indeed be refined to improve depth estimation when appropriately filtered and interpreted, even in complex 3D layouts.

Building upon these results, the second and third stages of experimentation focused on interaction and confirmation techniques. The comparative evaluation revealed that depth-aware strategies such as SphereCast and ConeCast enhanced target selection reliability without compromising usability, while simpler confirmation methods such as Wink and Double Blink offered effective trade-offs between speed and user comfort. Together, these results addressed my second research question, demonstrating that carefully designed interaction techniques can mitigate inaccuracies inherent to current eye-tracking systems and substantially improve targeting precision.

Finally, analysis of user performance and subjective feedback provided insights into visual fatigue and comfort. Although fatigue increased with task complexity, as expected, the introduction of stable confirmation-based selection techniques reduced the perceptual strain associated with gaze instability and repeated corrections. These outcomes contributed to the third research question, highlighting how interaction design choices can help maintain user comfort even when gaze data are imperfect or environments are visually demanding.

Overall, this work aims were achieved. The research established a reliable vergence-based interface capable of supporting accurate gaze depth estimation and robust object selection within cluttered virtual environments. Furthermore, it identified how different interaction strategies influence precision, speed, and fatigue, providing empirical guidance for the design of future gaze-based systems.

To conclude, this thesis contributes to the advancement of gaze-driven interaction by demonstrating that vergence, despite its inherent limitations, can serve as a viable and effective foundation for precise, controller-free interaction in VR. By combining methodological rigor with practical interface design, the work bridges the gap between theoretical gaze estimation models and their real-world application, paving the way for more adaptive and comfortable gaze-based interaction paradigms in extended reality.

7.3 Future Work

Although the present research established a functional and reliable vergence-based interaction interface, several promising directions remain open for future exploration.

A natural extension of this work is to test the proposed techniques in dynamic and ecologically valid environments. The present experiments deliberately employed heavily cluttered and static scenes to serve as a rigorous stress test for the developed methods. Although this design was effective in exposing the limits of vergence-based interaction, future studies could explore more natural but still moderately complex and dynamic environments, featuring moving targets or subtle environmental changes. Such scenarios would provide a more realistic assessment of robustness and adaptability under everyday VR conditions, balancing experimental control with ecological relevance.

Another avenue concerns hardware diversity. The experiments in this thesis were performed using

a Varjo headset, which provides high-quality eye tracking compared to most commercial systems but still operates at limited sampling rates. Testing the same techniques on professional-grade eye trackers, such as the 1000 Hz EyeLink series or other emerging research-oriented devices, could help determine how much higher frequency tracking improves vergence depth precision and stability.

In addition, future work should explore enhanced calibration and personalization. The current calibration process, though functional, may not fully account for individual differences in oculomotor behavior or IPD. A more refined calibration model, potentially adaptive or user-specific, could substantially improve the accuracy of vergence-based depth estimation and reduce fatigue over longer sessions.

Finally, there is room to further refine the techniques developed. In spite of the fact that ConeCast and Wink provided a strong performance, iterative adjustments, such as an improved dynamic interaction thresholds or limiting the number of objects considered at long distances, could enhance usability and reduce computational overhead. Such refinements could contribute to the design of next-generation gaze interfaces capable of maintaining high precision across both static and dynamic environments.

Bibliography

[1] C. Nitschke, A. Nakazawa, and H. Takemura, “Corneal imaging revisited: An overview of corneal reflection analysis and applications,” *IPSJ Transactions on Computer Vision and Applications*, 2013. [Online]. Available: https://www.jstage.jst.go.jp/article/ipsjtcva/5/0/5_1/_pdf/-char/ja

[2] M. S. AREFIN, J. E. S. II, R. A. C. HOFFING, and S. M. THURMAN, “Estimating perceptual depth changes with eye vergence and interpupillary distance using an eye tracker in virtual reality,” *Symposium on Eye Tracking Research and Applications*, 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3517031.3529632

[3] E. J. Norling, *Modelling Human Behaviour with BDI Agents*. University of Melbourne, Department of Computer Science and Software Engineering, 2009. [Online]. Available: https://www.researchgate.net/publication/229069973_Modelling_Human_Behaviour_with_BDI_Agents

[4] E. Gray, “Understanding depth of field – a beginner’s guide,” Online article, 2024, updated: 10th October 2024. [Online]. Available: https://photographylife.com/what-is-depth-of-field

[5] A. T. Duchowski, K. Krejtz, M. Volonte, C. J. Hughes, M. Brescia-Zapata, and P. O. Pilar Oreroe Brescia-Zapata e, “3d gaze in virtual reality: Vergence, calibration, event detection,” *Procedia Computer Science*, 2022. [Online]. Available: https://www.researchgate.net/publication/364426988_3D_Gaze_in_Virtual_Reality_Vergence_Calibration_Event_Detection

[6] M. R. Rebsdorf, T. Khumsan, J. Valvik, N. C. Nilsson, and A. Adjorlu, “Blink don’t wink: Exploring blinks as input for vr games,” in *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*, 2023, pp. 1–8. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3607822.3614527

[7] X. Xu, Y. He, Y. Ge1, and Z. Zheng, “Eyeexpand: A low-burden and accurate 3d object selection method with gaze and raycasting,” in *Computer Graphics Forum*. Wiley Online Library, 2025, p. e70144. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.70144

[8] L.-H. Fan, W.-C. Huang, X.-Q. Shao, and Y.-F. Niu, “Design recommendations for voluntary blink interactions based on pressure sensors,” *Advanced Engineering Informatics*, vol. 61, p. 102489, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10494120

[9] M. Burgess, “Apple vision pro’s eye tracking exposed what people type,” Online Article, 2024, accessed: 9th September 2025. [Online]. Available: https://www.wired.com/story/apple-vision-pro-persona-eye-tracking-spy-typing/

[10] X. Liu, L. Wang, Y. Liu, and J. Wu, “Automatic portals layout for vr navigation,” *Vol.:(0123456789)Virtual Reality*, vol. 28, no. 1, p. 9, 2024. [Online]. Available: https://doi.org/10.1007/s10055-023-00897-7

[11] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurs, “Pinpointing: Precise head- and eye-based target selection for augmented reality,” in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 2018, pp. 1–14. [Online]. Available: https://3dvar.com/Kyt%C3%B62018Pinpointing.pdf

[12] T. Piumsomboon, G. Lee, R. W. Lindeman, and M. Billinghurst, “Exploring natural eye-gaze-based interaction for immersive virtual reality,” in *2017 IEEE symposium on 3D user interfaces (3DUI)*, 2017, pp. 36–39. [Online]. Available: https://www.researchgate.net/publication/315852071_Exploring_natural_eye-gaze-based_interaction_for_immersive_virtual_reality

[13] A. T. Duchowski, E. Medlin, N. Cournia, A. Gramopadhye, B. Melloy, and S. Nair, “3d eye movement analysis for vr visual inspection training,” in *Proceedings of the 2002 symposium on Eye tracking research & applications*, 2002, pp. 103–110. [Online]. Available: http://andrewd.ces.clemson.edu/research/vislab/docs/etra02.pdf

[14] B. A. Myers, R. Bhatnagar, J. Nichols, C. H. Peck, D. Kong, R. Miller, and A. C. Long, “Interacting at a distance: Measuring the performance of laser pointers and other devices,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2001, pp. 33–40. [Online]. Available: https://www.researchgate.net/publication/234814854_Interacting_at_a_distance_Measuring_the_performance_of_laser_pointers_and_other_devices

[15] S. Zhai, C. Morimoto, and S. Ihde, “Manual and gaze input cascaded (magic) pointing,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1999, pp. 246–253. [Online]. Available: https://www.allpsych.uni-giessen.de/journalclub/pdf/ZhaiEtAl.Chi.1999.pdf

[16] A. T. Duchowski, D. H. House, J. Gestring, R. Congdon, L. Swirski, N. A. Dodgson, K. Krejtz, and I. Krejtz, “Comparing estimated gaze depth in virtual and physical environments,” in *Proceedings of the Symposium on Eye Tracking Research and Applications*, 2014, pp. 103–110. [Online]. Available: https://dl.acm.org/doi/10.1145/2578153.2578168

[17] S. Wibirama and K. Hamamoto, “3d gaze tracking system for nvidia 3d vision®,” in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 3194–3197. [Online]. Available: https://www.researchgate.net/publication/257601428_3D_gaze_tracking_system_for_NVidia_3D_VisionR

[18] P. Bourke. (1988, October) Points, lines, and planes. [Online]. Available: https://paulbourke.net/geometry/pointlineplane/

[19] A. Duchowski, “Eye tracking methodology; theory and practice,” *Qualitative Market Research: An International Journal*, vol. 5, pp. 51–59, 2007. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-84628-609-4_5

[20] K. McAnally, P. Grove, and G. Wallis, “Vergence eye movements in virtual reality,” *Displays*, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141938224000477?via%3Dihub

[21] E. G. Mlot, H. Bahmani, S. Wahl, and E. Kasneci, “3d gaze estimation using eye vergence,” in *International Conference on Health Informatics*, vol. 6. Scitepress, 2016, pp. 125–131. [Online]. Available: https://dl.acm.org/doi/10.5220/0005821201250131

[22] P. B. Hibbard, L. C. van Dam, and P. Scarfe, “The implications of interpupillary distance variability for virtual reality,” in *2020 International conference on 3D immersion (IC3D)*, 2020, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9376369

[23] N. A. Dodgson, “Variation and extrema of human interpupillary distance,” in *Stereoscopic displays and virtual reality systems XI*, vol. 5291, 2004, pp. 36–46. [Online]. Available: https://www.researchgate.net/publication/229084829_Variation_and_extrema_of_human_interpupillary_distance

[24] L. A. Remington and D. Goodwin, *Clinical Anatomy and Physiology of the Visual System E-book: Clinical Anatomy and Physiology of the Visual System E-book*. Elsevier Health Sciences, 2021. [Online]. Available: https://books.google.pt/books?id=gQ01EAAAQBAJ

[25] W. S. Tuten and W. M. Harmening, “Foveal vision,” *Current Biology*, vol. 31, no. 11, pp. R701–R703, 2021. [Online]. Available: https://www.cell.com/current-biology/fulltext/S0960-9822(21)00470-X

[26] K. Holmqvist, M. Nyström, R. Andersson, and R. Dewhurst, *Eye Tracking: A Comprehensive Guide To Methods And Measures*. oup Oxford, 2011. [Online]. Available: https://www.researchgate.net/publication/254913339_Eye_Tracking_A_Comprehensive_Guide_To_Methods_And_Measures

[27] S. P. Liversedge and J. M. Findlay, “Saccadic eye movements and cognition,” *Trends in cognitive sciences*, vol. 4, no. 1, pp. 6–14, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1364661399014187

[28] A. G. Lee, A. Kini, N. Al-Zubidi, and B. A. Othman, “Saccade,” Online article, 2022, reviewed: June 13, 2025; American Academy of Ophthalmology. [Online]. Available: https://eyewiki.org/Saccade

[29] J. R. Landreneau, N. P. Hesemann, and M. A. Cardonell, “Review on the myopia pandemic: epidemiology, risk factors, and prevention,” *Missouri medicine*, vol. 118, no. 2, p. 156, 2021. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8029638/

[30] T. Holmes, “Eye tracking study recruitment — managing participants with vision irregularities,” Blog post on Tobii website, feb 2019, accessed: 15 September 2025. [Online]. Available: https://www.tobii.com/blog/eye-tracking-study-recruitment-managing-participants-with-vision-irregularities

[31] I. B. Adhanom, P. MacNeilage, and E. Folmer, “Eye tracking in virtual reality: a broad review of applications and challenges,” *Virtual Reality*, vol. 27, no. 2, pp. 1481–1505, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s10055-022-00738-z

[32] D. Mardanbegi, C. Clarke, and H. Gellersen, “Monocular gaze depth estimation using the vestibulo-ocular reflex,” in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 2019, pp. 1–9. [Online]. Available: https://dl.acm.org/doi/10.1145/3314111.3319822

[33] S. Weber, R. S. Schubert, S. Vogt, B. M. Velichkovsky, and S. Pannasch, “Gaze3dfix: Detecting 3d fixations with an ellipsoidal bounding volume,” *Behavior research methods*, vol. 50, pp. 2004–2015, 2018. [Online]. Available: https://link.springer.com/article/10.3758/s13428-017-0969-4

[34] P. Kellnhofer, P. Didyk, K. Myszkowski, M. M. Hefeeda, H.-P. Seidel, and W. Matusik, “Gazestereo3d: Seamless disparity manipulations,” *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–13, 2016. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/2897824.2925866

[35] F. Monier, L. Hertel, S. Droit-Volet, and P. Chausse, “Ocular vergences measurement in virtual reality: A pilot study,” *Vision Research*, vol. 234, p. 108658, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0042698925001191

[36] X. M. Wang, M. Prenevost, A. Tarun, I. Robinson, M. Nitsche, G. Resch, A. Mazalek, and T. N. Welsh, “Investigating a geometrical solution to the vergence-accommodation conflict for targeted movements in virtual reality,” *arXiv preprint arXiv:2505.23310*, 2025. [Online]. Available: https://arxiv.org/abs/2505.23310?utm_source=chatgpt.com

[37] L. Sidenmark, C. Clarke, J. Newn, M. N. Lystbæk, K. Pfeuffer, and H. Gellersen, “Vergence matching: Inferring attention to objects in 3d environments for gaze-assisted selection,” in

Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–15. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3544548.3580685

[38] I. Schuetz and K. Fiehler, "Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability," *Journal of Eye Movement Research*, vol. 15, no. 3, pp. 10–16 910, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10136368/

[39] S. Mall, P. C. Brennan, and C. Mello-Thoms, "Modeling visual search behavior of breast radiologists using a deep convolution neural network," *Journal of medical imaging*, vol. 5, no. 3, pp. 035 502–035 502, 2018. [Online]. Available: https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC6086967&blobtype=pdf

[40] M. Boss, A. Engelhardt, A. Kar, Y. Li, D. Sun, J. T. Barron, H. P. A. Lensch, and V. Jampani, "Samurai: Shape and material from unconstrained real-world arbitrary image collections," *Advances in Neural Information Processing Systems*, vol. 35, pp. 26 389–26 403, 2022. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/file/a8f2713b5c6bdcd3d264f1aa9b9c6f03-Paper-Conference.pdf

[41] G. Wang, H. Zheng, and X. Zhang, "A robust checkerboard corner detection method for camera calibration based on improved yolox," *Frontiers in Physics*, vol. 9, p. 819019, 2022. [Online]. Available: https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2021.819019/full

[42] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, no. 1, p. 7068349, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1155/2018/7068349

[43] J. Singh, Urvashi, G. Singh, and S. Maheshwari, "Augmented reality technology: Current applications, challenges and its future," in *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2022, pp. 1722–1726. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9985665

[44] A.-L. von Behren, Y. Sauer, B. Severitt, and S. Wahl, "Cnn-based estimation of gaze distance in virtual reality using eye tracking and depth data," in *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*, 2025, pp. 1–7. [Online]. Available: https://dl.acm.org/doi/10.1145/3715669.3723122

[45] J. S. Stahl, "Eye-head coordination and the variation of eye-movement accuracy with orbital eccentricity," *Experimental brain research*, vol. 136, no. 2, pp. 200–210, 2001. [Online]. Available: https://www.researchgate.net/publication/12125492_Eye-head_coordination_and_the_variation_of_eye-movement_accuracy_with_orbital_eccentricity

[46] T. Pejsa, S. Andrist, M. Gleicher, and B. Mutlu, “Gaze and attention management for embodied conversational agents,” *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 5, no. 1, pp. 1–34, 2015. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/2724731

[47] Y. Wei, R. Shi, D. Yu, Y. Wang, Y. Li, L. Yu, and H.-N. Liang, “Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions,” in *Proceedings of the 2023 CHI conference on human factors in computing systems*, 2023, pp. 1–14. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3544548.3581042

[48] Y. Wang, G. Zhai, S. Chen, X. Min, Z. Gao, and X. Song, “Assessment of eye fatigue caused by head-mounted displays using eye-tracking,” *Biomedical engineering online*, vol. 18, no. 1, p. 111, 2019. [Online]. Available: https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/s12938-019-0731-5

[49] S. Kang, J. Jeong, G. A. Lee, S.-H. Kim, H.-J. Yang, and S. Kim, “The rayhand navigation: A virtual navigation method with relative position between hand and gaze-ray,” in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–15. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3613904.3642147

[50] D. Tao, X. Ren, K. Liu, Q. Mao, J. Cai, and H. Wang, “Effects of color scheme and visual fatigue on visual search performance and perceptions under vibration conditions,” *Displays*, vol. 82, p. 102667, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0141938224000313

[51] M. Vinnikov, R. S. Allison, and S. Fernandes, “Impact of depth of field simulation on visual fatigue: Who are impacted? and how?” *International Journal of Human-Computer Studies*, vol. 91, pp. 37–51, 2016.

[52] Y. Ji, “Evaluation of eye movement features and visual fatigue in virtual reality games.” *International Journal of Advanced Computer Science & Applications*, vol. 16, no. 1, 2025. [Online]. Available: https://openurl.ebsco.com/contentitem/gcd:182970577?sid=ebsco:plink:scholar&id=ebsco:gcd:182970577&crl=c

[53] J. Orlosky, C. Liu, K. Sakamoto, L. Sidenmark, and A. Mansour, “Eyeshadows: Peripheral virtual copies for rapid gaze selection and interaction,” in *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2024, pp. 681–689. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10494120

[54] “Performance analysis of saccades for primary and confirmatory target selection,” in *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, 2022, pp. 1–12. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3562939.3565619

[55] A. R. R. Gomez, C. Clarke, L. Sidenmark, and H. Gellersen, “Gaze+ hold: eyes-only direct manipulation with continuous gaze modulated by closure of one eye,” in *ACM symposium on eye tracking research and applications*, 2021, pp. 1–12. [Online]. Available: https://dl.acm.org/doi/10.1145/3448017.3457381

[56] A. K. Mutasim, A. U. Batmaz, and W. Stuerzlinger, “Pinch, click, or dwell: Comparing different selection techniques for eye-gaze-based pointing in virtual reality,” in *Acm symposium on eye tracking research and applications*, 2021, pp. 1–7. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3448018.3457998

[57] HTC Corporation, “Htc vive base station,” Product webpage, 2025, accessed: 9 September 2025. [Online]. Available: https://www.vive.com/eu/accessory/base-station/

[58] F. Nusrat, F. Hassan, H. Zhong, and X. Wang, “How developers optimize virtual reality applications: A study of optimization commits in open source unity projects,” in *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*. IEEE, 2021, pp. 473–485. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3448018.3457998

[59] W. J. Lee, J. H. Kim, Y. U. Shin, S. Hwang, and H. W. Lim, “Differences in eye movement range based on age and gaze direction,” *Eye*, vol. 33, no. 7, pp. 1145–1151, 2019. [Online]. Available: https://www.nature.com/articles/s41433-019-0376-4

[60] A. R. Pijpaert, H. H. L. M. J. Goossens, B. W. van Dijk, L. J. B. Roetman, R. M. A. van Nispen, and L. J. R. van Rijn, “A validation study on the accuracy and precision of gaze and vergence using stereoscopic eye-tracking technology,” *Behavior Research Methods*, vol. 57, no. 8, p. 214, 2025. [Online]. Available: https://link.springer.com/article/10.3758/s13428-025-02731-1

[61] CitrusBits, “Choosing the right sampling rate for eye tracking in vr headsets,” Jul. 2025, accessed: 2025-10-03. [Online]. Available: https://citrusbits.com/choosing-the-right-sampling-rate-for-eye-tracking-in-vr-headsets/

[62] K. Holmqvist, S. L. Örbom, I. T. Hooge, D. C. Niehorster, R. G. Alexander, R. Andersson, J. S. Benjamins, P. Blignaut, A.-M. Brouwer, L. L. Chuang *et al.*, “Retracted article: Eye tracking: empirical foundations for a minimal reporting guideline,” *Behavior research methods*, vol. 55, no. 1, pp. 364–416, 2023. [Online]. Available: https://link.springer.com/article/10.3758/s13428-021-01762-8

[63] Unity Technologies, “Physics.overlapsphere (unity scripting api),” https://docs.unity3d.com/6000.2/Documentation/ScriptReference/Physics.OverlapSphere.html, 2025, accessed: 2025-26-09.

[64] M. Lamb, M. Brundin, E. P. Luque, and E. Billing, “Eye-tracking beyond peripersonal space in virtual reality: validation and best practices,” *Frontiers in Virtual Reality*, vol. 3, p. 864653, 2022. [Online]. Available: https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2022.864653/full

[65] Unity Technologies, “Physics.spherecastall (unity scripting api),” https://docs.unity3d.com/6000.2/Documentation/ScriptReference/Physics.SphereCastAll.html, 2025, accessed: 2025-26-09.

[66] Varjo, “Industrial-strength eye tracking in varjo headsets,” https://varjo.com/blog/industrial-strength-eye-tracking-in-varjo/, 2025, accessed: 2025-09-27.

[67] Y. T. Paulus and G. B. Remijn, “Usability of various dwell times for eye-gaze-based object selection with eye tracking. displays, 67, article 101997,” 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141938221000123

[68] X. Yi, L. Qiu, W. Tang, Y. Fan, and Y. S. Hewu Li, “Deep: 3d gaze pointing in virtual reality leveraging eyelid movement,” in *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, 2022, pp. 1–14. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3526113.3545673

[69] F. Milanese, “Dna molecules (3d model),” https://superhivemarket.com/products/dna-molecules, royalty Free license; accessed: 2025-09-27.

Profile Questionnaire for Pilot Test

Experiência de seleção de objetos com o olhar em realidade virtual

Questionário de perfil anónimo do participante
Este formulário busca recolher informações demográficas das pessoas que contribuírem para os testes em questão. Ao preencher este formulário, permito que as respostas aqui dadas sejam utilizados no contexto deste projeto de forma anónima e para fins estatísticos apenas.

* Indica uma pergunta obrigatória

1. ID do participante (preenchido pelo investigador com garantia de anonimato) *

2. Marque a opção abaixo se você concorda em participa na experiência. *

Marcar tudo o que for aplicável.

- [ ] Aceito participar na experiência. Fui devidamente informado(a) pelo investigador sobre a experiência e os procedimentos nela envolvidos. Foi-me garantido o sigilo das informações e que posso interromper a minha participação a qualquer momento.

Triagem demográfica

3. Faixa etária *

Marcar apenas uma oval.

- ◯ 18 a 24 anos
- ◯ 25 a 35 anos
- ◯ 35 a 45 anos
- ◯ 45 a 55 anos
- ◯ 56 anos ou mais

4. Sexo *

Marcar apenas uma oval.

- ◯ Masculino
- ◯ Feminino
- ◯ Prefiro não responder
- ◯ Outra:

5. Escolaridade *

Marcar apenas uma oval.

- ◯ Ensino básico completo (até 9º ano)
- ◯ Ensino secundário completo (até 11º ano)
- ◯ Ensino pós secundário não-superior
- ◯ Ensino superior incompleto
- ◯ Licenciatura completa
- ◯ Mestrado completo
- ◯ Doutoramento completo
- ◯ Outra: _____

6. Possui algum problema de visão ou dificuldades motoras nos olhos?

Marque todas que se apliquem

Marcar tudo o que for aplicável.

- ☐ Miopia
- ☐ Astigmatismo
- ☐ Hipermetropia
- ☐ Daltonismo/Discromatopsia
- ☐ Visão monocular (visão muito reduzida num olho, por estrabismo, ambliopia ou outra condição que impeça a visão estereoscópica e afete a capacidade de avaliar distâncias e profundidades)
- ☐ Tenho dificuldades para piscar um ou os dois olhos controladamente
- ☐ Outra: _____

7. Teste do olho dominante (fazer o teste) *

Marcar apenas uma oval.

- ◯ Esquerdo
- ◯ Direito

Realidade Virtual

Realidade virtual é um paradigma e estilo de interação com interfaces onde o usuário vê apenas elementos virtuais, uma vez que sua visão do mundo real é bloqueada pelo uso de óculos especiais. Chamamos a isso: imersão. Os objetos virtuais representam um mundo onde sua exploração e manipulação são úteis para a realização de uma determinada atividade, seja ela de entretenimento, treino, ou análise de dados.

8. Já experimentou interagir com **realidade virtual** usando quaisquer óculos VR antes de participar deste experimento? *

Marcar apenas uma oval.

- ◯ Nunca
- ◯ Menos de 5 vezes
- ◯ Mais de 5 vezes
- ◯ Faz parte do meu dia-a-dia

9. Com que frequencia joga jogos digitais 3D? *

Marcar apenas uma oval.

- ◯ Nunca joguei
- ◯ Menos de uma vez por ano
- ◯ Cerca de uma vez por mês
- ◯ Pelo menos uma vez por semana
- ◯ Quase todos os dias

B

Post Questionnaire for Pilot Test

Experiência de seleção de objetos com o olhar em realidade virtual

Questionário de opinião sobre a experiência

* Indica uma pergunta obrigatória

1. ID do participante (preenchido pelo pesquisador com garantia de anonimato) *

_____

Perceção ao utilizar o sistema

Responda em referência à sua experiência com as tarefas que você acabou de realizar

2. Havia esferas de tamanhos diferentes. Classifique a dificuldade em selecionar conforme o tamanho. *

Marcar apenas uma oval.

	1	2	3	4	5	
Pequ	◯	◯	◯	◯	◯	Grandes mais fácil

3. Havia esferas a diferentes distâncias. Classifique a dificuldade em selecionar conforme a distância. *

Marcar apenas uma oval.

	1	2	3	4	5	
Mais	◯	◯	◯	◯	◯	Mais distantes mais fácil

4. Você achou que conseguiu confirmar a seleção da esfera desejada... *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Nun	◯	◯	◯	◯	◯	◯	◯	Sempre

5. Nas vezes em que teve dificuldade em posicionar o contorno e fazer a seleção, sentiu que a seleção recaia mais predominantemente para uma direção. *

Marcar apenas uma oval.

- ◯ Para cima ↑
- ◯ Para baixo ↓
- ◯ Para a esquerda ←
- ◯ Para a direita →
- ◯ Sem direção preferencial

6. Comparando Double Blink e Wink, foi mais fácil confirmar a seleção ao usar...

Marcar apenas uma oval.

- ◯ Double Blink
- ◯ Wink

7. Comparando Dwell e Wink, foi mais fácil confirmar a seleção ao usar...

Marcar apenas uma oval.

- ◯ Dwell
- ◯ Wink

8. Comparando Dwell e Double Winking, foi mais fácil confirmar a seleção ao usar...

Marcar apenas uma oval.

- ◯ Dwell
- ◯ Double Winking

9. Comentários livres (partilhe qualquer ideia ou impressão que lhe ocorreu sobre a experiência.

Muito obrigado por suas respostas!

C

Profile Questionnaire for Final Test

Experiência de seleção de objetos com o olhar em realidade virtual

Questionário de perfil anónimo do participante

Este formulário busca recolher informações demográficas das pessoas que contribuírem para os testes em questão. Ao preencher este formulário, permito que as respostas aqui dadas sejam utilizados no contexto deste projeto de forma anónima e para fins estatísticos apenas.

* Indica uma pergunta obrigatória

1. ID do participante (preenchido pelo investigador com garantia de anonimato) *

2. Marque a opção abaixo se você assinou o termo de consentimento informado: *

Marcar tudo o que for aplicável.

- [ ] Aceito participar do experimento e assinei o termo de consentimento. Fui devidamente informado(a) pelo investigador sobre o experimento, os procedimentos nele envolvidos, assim como os possíveis riscos e benefícios decorrentes de minha participação. Foi-me garantido o sigilo das informações e que posso retirar meu consentimento a qualquer momento.

Triagem demográfica

3. Faixa etária *

Marcar apenas uma oval.

- ◯ 18 a 25 anos
- ◯ 26 a 35 anos
- ◯ 36 a 45 anos
- ◯ 46 a 55 anos
- ◯ 56 anos ou mais

4. Sexo *

Marcar apenas uma oval.

- ◯ Masculino
- ◯ Feminino
- ◯ Prefiro não responder
- ◯ Outra:

5. Escolaridade *

Marcar apenas uma oval.

- ◯ Ensino básico completo (até 9º ano)
- ◯ Ensino secundário completo (até 11º ano)
- ◯ Ensino pós secundário não-superior
- ◯ Ensino superior incompleto
- ◯ Licenciatura completa
- ◯ Mestrado completo
- ◯ Doutoramento completo
- ◯ Outra: _____

6. Possui algum problema de visão ou dificuldades motoras nos olhos? *

Marque todas que se apliquem

Marcar tudo o que for aplicável.

- ☐ Miopia (longe)
- ☐ Hipermetropia (perto)
- ☐ Astigmatismo (ambos)
- ☐ Daltonismo/Discromatopsia
- ☐ Visão monocular (visão muito reduzida num olho, por estrabismo, ambliopia ou outra condição que impeça a visão estereoscópica e afete a capacidade de avaliar distâncias e profundidades)
- ☐ Tenho dificuldades para piscar um ou os dois olhos controladamente
- ☐ Outra: _____

7. Usa lentes de contacto, se sim de que tipo são? *

Marcar apenas uma oval.

- ◯ Não
- ◯ Lentes suaves
- ◯ Lentes tóricas (para astigmatismo)
- ◯ Lentes rígidas (RGP)
- ◯ Lentes esclerais

8. Já realizou alguma cirurgia de correção visual?

Marcar apenas uma oval.

- ◯ Não
- ◯ LASIK / PRK / LASEK
- ◯ Lentes intraoculares (ICL)
- ◯ Cirurgia de cataratas

9. Teste do olho dominante (fazer o teste) *

Marcar apenas uma oval.

- ◯ Esquerdo
- ◯ Direito

Realidade Virtual

Realidade virtual é um paradigma e estilo de interação com interfaces onde o usuário vê apenas elementos virtuais, uma vez que sua visão do mundo real é bloqueada pelo uso de óculos especiais. Chamamos a isso: imersão. Os objetos virtuais representam um mundo onde sua exploração e manipulação são úteis para a realização de uma determinada atividade, seja ela de entretenimento, treino, ou análise de dados.

10. Você já experimentou **realidade virtual** usando quaisquer óculos VR antes de participar deste experimento? *

Marcar apenas uma oval.

- ◯ Nunca
- ◯ Menos de 5 vezes
- ◯ Mais de 5 vezes
- ◯ Faz parte do meu dia-a-dia

11. Com que frequencia você joga jogos digitais 3D? *

Marcar apenas uma oval.

- ◯ Nunca joguei
- ◯ Menos de uma vez por ano
- ◯ Cerca de uma vez por mês
- ◯ Pelo menos uma vez por semana
- ◯ Quase todos os dias

D

Post Questionnaire for Final Test

Experiência de seleção de objetos com o olhar em realidade virtual

Questionário de opinião sobre a experiência

* Indica uma pergunta obrigatória

1. ID do participante (preenchido pelo pesquisador com garantia de anonimato) *

Mal-estar

Marque a opção que mais se aproxima do seu nível atual para cada sintoma

2. Desconforto geral *

Marcar apenas uma oval.

	1	2	3	4	5	
Nen	◯	◯	◯	◯	◯	Severo

3. Vista cansada *

Marcar apenas uma oval.

	1	2	3	4	5	
Nen	◯	◯	◯	◯	◯	Severo

4. Náusea *

Marcar apenas uma oval.

	1	2	3	4	5	
Nen	◯	◯	◯	◯	◯	Severo

5. Tontura *

Marcar apenas uma oval.

	1	2	3	4	5	
Nen	◯	◯	◯	◯	◯	Severo

Percepção ao utilizar o sistema

Responda em referência à sua experiência com as tarefas que você acabou de realizar

6. Você executou seleções em 4 blocos de 10 esferas cada. Pelo que pode se lembrar, marque o bloco que achou **melhor**. *

Marcar apenas uma oval.

- Bloco 1
- Bloco 2
- Bloco 3
- Bloco 4

7. Agora, pelo que pode se lembrar, marque o bloco que achou **menos bom ou pior**. *

Marcar apenas uma oval.

- Bloco 1
- Bloco 2
- Bloco 3
- Bloco 4

8. Havia esferas de tamanhos diferentes. Classifique a dificuldade em selecionar conforme o tamanho. *

Marcar apenas uma oval.

	1	2	3	4	5	
Peq	◯	◯	◯	◯	◯	Grandes mais fácil

9. Havia esferas a diferentes distâncias. Classifique a dificuldade em selecionar conforme a distância. *

Marcar apenas uma oval.

	1	2	3	4	5	
Mai	◯	◯	◯	◯	◯	Mais distantes mais fácil

10. Sobre a **precisão da seleção**, você sentiu que a esfera que assumia o contorno era a mesma para a qual olhava... *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Nun	◯	◯	◯	◯	◯	◯	◯	Sempre

11. Nas vezes em que teve dificuldade em posicionar o contorno e fazer a seleção, sentiu que a seleção recaia mais predominantemente para uma direção. *

Marcar tudo o que for aplicável.

- [ ] Para cima ↑
- [ ] Para baixo ↓
- [ ] Para a esquerda ←
- [ ] Para a direita →
- [ ] Sem direção preferencial

12. Comentários livres (partilhe qualquer ideia ou impressão que lhe ocorreu sobre a experiência.

Muito obrigado por suas respostas!

E

Informed Consent

CONSENTIMENTO INFORMADO PARA ATUAR COMO PARTICIPANTE EM EXPERIMENTO DE INTERAÇÃO COM COMPUTADORES

Título do projeto: Estudo sobre uma técnica para selecionar objetos com o olhar em realidade virtual

Investigador Principal: Prof. Dr. Anderson Maciel, IST, anderson.maciel@tecnico.ulisboa.pt

Está convidado(a) a participar num estudo científico. Por favor, leia este documento com bastante atenção antes de assiná-lo. Caso haja alguma palavra ou frase que não consiga entender, converse com o investigador responsável pelo estudo ou com um membro da equipa de investigação para esclarecê-los. O propósito deste termo de consentimento é explicar tudo sobre o estudo e solicitar a sua concordância em participar. Os detalhes sobre este estudo são apresentados neste termo de consentimento. É importante que você entenda essas informações para que possa tomar uma decisão informada sobre sua participação.

A sua participação no estudo é voluntária. Você pode optar por não participar ou pode retirar o seu consentimento para participar do estudo, por qualquer motivo, sem sofrer penalidades, a qualquer altura. Os estudos científicos são projetados para obter novos conhecimentos. Essas novas informações podem ajudar as pessoas no futuro. Pode não haver nenhum benefício direto para você ao participar do estudo.

Todos os estudos com seres humanos envolvem algum tipo de risco. No nosso estudo, será utilizado um sistema de exibição imersivo (óculos de realidade virtual). Os possíveis riscos ou desconfortos decorrentes do uso desses dispositivos são a ocorrência dos seguintes efeitos colaterais temporários, que podem ser leves, moderados ou graves: tontura, náusea, desorientação, fadiga, cansaço visual, dificuldade de concentração, aumento da salivação, sudorese, vertigem. Esses efeitos são semelhantes aos do enjoo ao viajar de carro ou barco. Se ocorrerem, espera-se que passem alguns minutos após o término da experiência. O efeito mais provável é de que ocorra ligeira fadiga ocular, embora a maioria das pessoas não sinta nada ao usar sistemas semelhantes. Formas graves de qualquer um desses sintomas são menos prováveis. Se você optar por não participar do estudo ou sair do estudo antes de sua conclusão, isso não afetará seu relacionamento com as pessoas e instituições envolvidas.

Estudos como este também podem trazer benefícios. Os possíveis benefícios resultantes da sua participação, são o desenvolvimento de novas tecnologias que permitirão a médicos e outros profissionais de saúde produzir ferramentas de modo a promover uma melhor experiência hospitalar.

Você receberá uma cópia deste termo de consentimento. Se tiver alguma dúvida sobre este estudo a qualquer momento, você deve perguntar aos investigadores mencionados neste termo. As informações de contato estão abaixo:

Prof. Dr. Anderson Maciel, IST, anderson.maciel@tecnico.ulisboa.pt
Prof. Dr. Joaquim Jorge, IST, jorgej@tecnico.ulisboa.pt
David Martins Correia, IST, david.martins.correia@tecnico.ulisboa.pt

Sobre o que é o estudo?

O objetivo deste estudo é avaliar diferentes métodos de seleção de objetos através da direção do olhar usando óculos de realidade virtual equipados com câmaras de rastreio ocular. Através de experiências práticas, pretende-se determinar o desempenho das diferentes técnicas com relação à eficiência e precisão. Os testes incluirão tarefas de seleção de objetos simples como esferas pela simples ação de olhar para eles usando o equipamento de captura. Os participantes também serão convidados a responder perguntas sobre suas preferências no uso das diferentes técnicas. Estas experiências permitiram avaliar os efeitos de cada uma com relação à fadiga e experiência de usuário, para que futuramente sejam usadas no design de aplicações mais ergonómicas e produtivas.

Por que estou a ser convidado?

Por ser um adulto saudável (com pelo menos 18 anos de idade), que não possui uma condição médica relacionada à visão, sensibilidade à luz ou propensão a convulsões, como epilepsia. Também por possuir visão normal ou corrigida para o estado normal e poder perceber imagens estereoscópicas 3D. Essas condições de saúde são essenciais para esta experiência e para a investigação em realidade virtual em geral. A sua capacidade de processar imagens estereoscópicas 3D pode beneficiar a pesquisa e fornecer resultados imparciais informativos. Por favor, informe o pesquisador se tiver alguma das condições médicas mencionadas ou outras preocupações médicas.

SIM	**NÃO**	**Responda à todas as questões abaixo marcando SIM ou NÃO.**
▢	▢	Marque SIM se tem 18 anos de idade ou mais.
▢	▢	Marque SIM se possui visão normal ou corrigida para o estado normal.
▢	▢	Marque SIM se tem uma condição médica relacionada à fotofobia (sensibilidade à luz)
▢	▢	Marque SIM se já sofreu convulsões.
▢	▢	Tive algum sintoma grave ao usar óculos de realidade virtual no passado. Qual? _____
▢	▢	Marque SIM se tem outra condição médica que desaconselha a participação em jogos digitais. Qual? _____

O que terei de fazer se eu concordar em participar do estudo?

Uma vez que tenha concordado em participar, você realizará uma tarefa interativa com o dispositivo fornecido pelo investigador. O objetivo será selecionar alvos representados por esferas destacadas com uma cor diferente dentro de um conjunto de esferas similares. A seleção é feita ao olhar para a esfera, observar se ela ficou marcada com um contorno de pré-seleção, e confirmar a seleção piscando os olhos. Para isto será explicado como usar e calibrar os óculos de realidade virtual. Em seguida, receberá instruções e uma demonstração de como utilizar a técnica de seleção. Após estar ambientado com a ferramenta, poderá praticar com a interface até estar seguro para prosseguir e

iniciar a coleta de dados.

Haverá alguma gravação de áudio/vídeo?

Não. A aplicação registará apenas as ações vistas no ecrã imersivo (óculos), que são todas representações virtuais e sem identificação. Nenhuma imagem sua ou amostra de sua voz será registrada.

Quais são os riscos para mim?

A participação neste estudo apresenta riscos físicos mínimos aos participantes. Os únicos riscos mínimos previsíveis são ligeiro cansaço visual temporário, tontura e leve náusea durante as tarefas com a aplicação. Usaremos as respostas que deu na página anterior para determinar se está apto a participar e você será dispensado caso considere que tem um risco moderado ou alto de ter algum sintoma grave. Se você acredita que não seria capaz de concluir as tarefas realizadas através do manuseamento de um computador, por favor, não concorde em participar deste estudo. Durante o estudo, você terá permissão para interromper ou parar a tarefa. Se tiver dúvidas, desejar mais informações ou tiver sugestões, entre em contato com os investigadores apresentados em cima.

Existem benefícios para a sociedade decorrentes de minha participação nesta investigação?

Os resultados podem contribuir para o desenvolvimento de técnicas de interação em ambientes virtuais, que no futuro podem ser usados em aplicações reais, como simuladores de treino, procedimentos terapêuticos e jogos, além de avançar o campo da realidade virtual e aumentada.

Existem benefícios para mim por participar neste estudo de pesquisa?

Embora você possa não se beneficiar diretamente deste estudo, os resultados obtidos podem ajudar no design de futuros sistemas interativos dos quais poderá indiretamente usufruir.

Receberei algum pagamento por participar do estudo? Terei algum custo?

Este estudo não terá nenhum custo para si e você não receberá nenhum tipo de compensação monetária ou equivalente.

Como você manterá as minhas informações confidenciais?

Nenhum dos dados do estudo será registado com informações que identifiquem os participantes individualmente. Em vez disso, atribuiremos um código numérico (ID) às suas respostas que não estará associado ao seu nome. Se os resultados do estudo forem publicados em um periódico revistado por pares ou apresentados em conferências, apenas dados do grupo ou respostas individuais anonimizadas serão apresentados, e nenhum dado pessoal ou ID de estudo será incluído.

A confidencialidade absoluta dos dados fornecidos através da Internet não pode ser garantida devido às proteções limitadas do acesso à Internet. Certifique-se de fechar seu navegador quando terminar, para que ninguém possa ver o que você estava a fazer. As respostas coletadas serão