Coordinated Eye and Head Movements for Gaze Interaction in 3D Environments

Ludwig Sidenmark, BSc, MSc
School of Computing and Communications
Lancaster University

A thesis submitted for the degree of
Doctor of Philosophy

January, 2023

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The work presented in this thesis was done under the guidance of Professor Hans Gellersen at Lancaster University's School of Computing and Communications. The material has not been submitted, either in whole or in part, for a degree at this, or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words including appendices and footnotes, but excluding the bibliography. A rough estimate of the word count is: 54389.

Ludwig Sidenmark

Coordinated Eye and Head Movements for Gaze Interaction in 3D Environments

Ludwig Sidenmark, BSc, MSc.
School of Computing and Communications, Lancaster University

A thesis submitted for the degree of *Doctor of Philosophy.* January, 2023

Abstract

Gaze is attractive for interaction, as we naturally look at objects we are interested in. As a result, gaze has received significant attention within human-computer interaction as an input modality. However, gaze has been limited to only eye movements in situations where head movements are not expected to be used or as head movements in an approximation of gaze when an eye tracker is unavailable. From these observations arise an opportunity and a challenge: we propose to consider gaze as multi-modal in line with psychology and neuroscience research to more accurately represent user movements. The natural coordination of eye and head movements could then enable the development of novel interaction techniques to further the possibilities of gaze as an input modality. However, knowledge of the eye and head coordination in 3D environments and its usage for interaction design is limited.

This thesis explores eye and head coordination and their potential for interaction in 3D environments by developing interaction techniques that aim to tackle established gaze-interaction issues. We study fundamental eye, head, and body movements in virtual reality during gaze shifts. From the study results, we design interaction techniques and applications that avoid the Midas touch issue, allow expressive gaze-based interaction, and handle eye tracking accuracy issues. We ground the evaluation of our interaction techniques through empirical studies.

From the techniques and study results, we define three design principles for coordinated eye and head interaction from these works that distinguish between eye-only and head-supported gaze shifts, eye-head alignment as input, and distinguishing head movements for gestures and head movements that naturally occur to support gaze. We showcase new directions for gaze-based interaction and present a new way to think about gaze by taking a more comprehensive approach to gaze interaction and showing that there is more to gaze than just the eyes.

Contributing Publications

Ludwig Sidenmark and Hans Gellersen. "Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality". In: *ACM Trans. Comput.-Hum. Interact.* 27.1 (Dec. 2019). ISSN: 1073-0516. DOI: 10.1145/3361218

Ludwig Sidenmark and Hans Gellersen. "Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection". In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. UIST '19. New Orleans, LA, USA: Association for Computing Machinery, 2019, pp. 1161–1174. ISBN: 9781450368162. DOI: 10.1145/3332165.3347921

Ludwig Sidenmark et al. "Radi-Eye: Hands-Free Radial Interfaces for 3D Interaction using Gaze-Activated Head-Crossing". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021, pp. 1–11. DOI: 10.1145/3411764.3445697

Ludwig Sidenmark et al. "BimodalGaze: Seamlessly Refined Pointing with Gaze and Filtered Gestural Head Movement". In: *ACM Symposium on Eye Tracking Research and Applications*. ETRA '20 Full Papers. Stuttgart, Germany: Association for Computing Machinery, 2020. ISBN: 9781450371339. DOI: 10.1145/3379155.3391312

Acknowledgements

Behind this thesis lies the support and help of numerous people for whom I wish to express my gratitude. First, I would like to express my sincere thanks to my advisor Prof. Hans Gellersen who invited me into the academic world and for his continuous support of my doctoral studies and related research, for his patience, motivation, and wisdom. Hans' guidance helped me grow to the research that I am today and through the research and writing of this thesis. I could not have had a better advisor and mentor for my Ph.D. journey.

I want to thank my labmates (both old and new) for the long days, the long conversations about life and research, the fun, and the baking. I will fondly look back at the last five years of all the fun and moments we had together. I feel lucky to count you all among my friends who have been there and supported me day after day.

My sincere thanks also goes to Patrick Connor and Jessica Grant, and everyone at Sony Interactive Entertainment Europe who supported my PhD and also provided me with the opportunity to join their team as an intern. Your support and the lessons I have learnt working with you will remain with me. I also thank Dr. Marcello Giordano, Dr. Mark Parent, and all at Meta Reality Labs for the opportunity to intern with their team. My time with you was enormously inspiring and it was a fantastic experience to spend the summer with you.

I could not have reached this without the support and love of my family. Most importantly, I would like to thank my loving partner Boyeun for her continuous support, never ending patience, and always lending an ear and a hand to anything that comes to mind. I would also like to thank my parents Susanne and Johan who have continuously supported me and who encouraged me to take the leap to start this journey. Special thanks to my sister Ellie and my brother Victor for all the laughs and chats whenever I would come home to Sweden. Finally, I would like to thank all of my friends I have made in Lancaster and all my friends back home who I have shared the ups and downs and who have made this journey so special.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis introduces the concept of leveraging eye and head coordination for interaction in 3D environments. Eye and head movements have long been investigated for use in interaction on 2D screens [74] and in 3D interaction within virtual or augmented reality (VR/AR) [87, 138]. However, these works the eyes and head as separate rather than integral. We base our work on previous research in psychology and neuroscience that treats gaze as multi-modal and continuous coordination between the eyes, head, and body [88] discussed more in Section 2.1 and Section 3.1). In this thesis, we review previous work on gaze interaction and eye-head coordination, introduce gaze as multi-modal to the HCI community, investigate the nature of the eye and head coordination in VR, and present three design principles for eye and head coordinated interaction.

1.1 Gaze for Interaction

Our eyes are at the centre of our interaction with the world, where we direct our gaze is a powerful cue for attention, interest, and intent. Gaze is attractive for interaction, as we naturally look at objects we are interested in and consider for manipulation. Furthermore, we can move our gaze to a target faster than any other body part. Gaze shifts are also considered effortless, as significantly less energy is required to move our eyes compared to other modalities such as our head, hands or feet. This aspect makes gaze-based interaction attractive for users with limited motor control [91], or in places where motion is limited, such as public transport [191]. Also, gaze is commonly suggested as a hands-free alternative for manual interaction when the hands are unavailable or busy [55].

Eye tracking for interaction using eye pointing was initially introduced to the Human-Computer Interaction (HCI) field in the eighties to provide users with limited motor control an alternative to mouse input [68, 190]. Since then, researchers have focused mainly on tracking gaze relative to a personal computer display where, at typical viewing distances, the display width is usually within a visual angle of 40° [149]. The content can be viewed and interacted with the eyes alone in such settings and does not require any significant head contribution.

More recently, eye trackers have become common in commercial head-mounted VR (e.g. HTC Vive Pro Eye, FOVE 0) and AR (e.g. HoloLens 2, Magic Leap 1) products, which has enabled researchers and designers to develop gaze-based experiences beyond the desktop screen. These devices support viewing of virtual scenes that are not limited to the field of view (FOV) of the head-mounted display (HMD), but that can extend over a wider field of regard (FOR). For example, if the virtual world completely surrounds the user, the FOR is 360°, while the FOV would be limited to around 100° with current VR HMDs. The display position and orientation in space determine the part of the FOR within the current FOV, based on the premise that users naturally move their head and body to control what they see. Therefore, gaze is central to interaction in AR/VR but it is commonly only approximated by head orientation in space, disregarding the smaller movements made by the eyes. For the rest of this thesis, we will use the term extended reality (XR) as an umbrella term for AR and VR.

In the real world, our eyes are at the centre of our interaction with the world and gaze shifts are achieved through a combination of eye, head, and body movements. We move the eyes in the head, the head relative to the torso, and the torso relative to the world [89]. For example, when we look from a display in front to a second screen in hand, we will lower our eyes in coordination with tilting of our heads; depending on the distance neither head nor eyes will cover the full distance [180]. When we look up and to a person next to us, we will not only move our eyes and head again but may also shift our torso toward the new target. Researchers have investigated these coordinated movements for decades in real-world settings within the field of neuroscience [13, 88]. We use these works as a foundation for our work when adapting gaze interaction to XR.

Within the fields of eye tracking and HCI, the literature on gaze interaction is not consistent in its use of terminology. Where eye trackers are used as an input device, gaze input is commonly treated as synonymous with eye movement [73, 74] while head movement is suppressed, filtered or plain ignored [57, 204]. In contrast, other work has conceived gaze as the direction a user faces, in abstraction from the finer-grained movement of the eyes within the head due to lack of sensors or computational costs [111, 201].

In this thesis, we argue that gaze must be treated as a *multi-modal* input when adapted for XR interaction beyond the screen to use it to its full potential. The eyes have a limited range of motion and require the support of the head and body to reach further, while the head and body lack the fine-grained movements of gaze. By treating gaze as multi-modal, we can capture the fine grained movements and the full range of motion of gaze while leveraging the relative strengths of each movement system. In this work, we use terminology consistent with research on eye movements in natural behaviour [88]. We refer to gaze as the direction in which the eye points relative to the world. As such, gaze presents the sum of the eye-in-head rotation and the head orientation (or head pose) toward the gaze target. The head orientation itself is the sum of the torso orientation toward the gaze target and head-on-torso rotation. Gaze shifts can occur within the user's current FOV (e.g., triggered by a stimulus in the peripheral vision), but can also extend beyond view (e.g. when a user turns to look behind their back).

We aim to address the well-known problems of gaze-based interaction in a new way by treating gaze as multi-modal. Our focus is in the domains of XR, where the eyes and the head and body are required to interact comfortably with the world visually and where our movements and behaviour mimic those of the real world. We aim to design interaction techniques and interfaces that adapt to gaze, i.e. eyes in conjunction with the head, how they coordinate and their physiological capabilities to address gaze-based interaction limitations. We argue that this aspect is vital to consider as gaze-based interaction moves away from the screen to the wider FOVs and space around as made possible with modern XR devices.

1.2 Gaze Interaction Challenges

Although there are many benefits to gaze-based interaction, it also has several inherent interaction problems. In this thesis, we focus on three limitations that have received extensive attention in previous research [43, 74]:

Midas touch With gaze-based interaction users have to use gaze for two purposes. Users mainly rely on gaze primarily for information seeking, and overlaying this with gaze input can cause issues. This issue is commonly referred to as the Midas touch problem, where every gaze movement can trigger interactions [74]. For eyes-only interfaces, this has several practical problems. First, using eyes for hover interaction will trigger an interaction on every gazed-on target [74]. Similarly, if a cursor is used that indicates the pointing position will follow the eyes' every move, distracting users [183]. Furthermore, after starting an interaction with the eyes, the user has to maintain their gaze (i.e. dwell) on the interacted object to maintain the interaction [15]. These aspects restrict users' ability to inspect targets with their gaze while maintaining or triggering new interactions.

Lack of expressive interaction. Gaze lacks an intrinsic mechanism (e.g. a 'click') for target selection or manipulation which limits the available action that can be performed by users. Gaze-only interfaces usually address this issue with dwell-based techniques, where users fixate their gaze on a target for an extended period to perform an interaction [74]. The decision on the required dwell time to perform an interaction is a weighting of speed and accuracy. A short dwell time allows faster selections, but users lose control as they may accidentally select targets while simply looking at them [69]. Meanwhile, a longer dwell time leads to more accurate selection, but is slower and can cause strain because the eyes remain stationary longer than what is natural [26, 173]. Furthermore, another drawback of the dwell mechanism is that users cannot rest their gaze on a target without causing a selection. In addition, users cannot pause a dwell selection to consider other targets without cancelling the current and starting a new selection. All these drawbacks restrict the users gaze movements, as moving the gaze across targets may cause unintended interaction, but also restricts the user to slow simple selection interactions.

Calibration and accuracy issues. Eye tracking and in extension gaze interaction can have significant signal accuracy issues. First, our natural gaze alignment over a target is not precise as the fovea, the small region of highest visual acuity at the centre of our eyes has a 1° diameter [90]. Furthermore, even when fixating on objects, our eyes are never entirely still, as they will perform a constant jitter movement [61]. Measurement limitations of eye tracking further exacerbate these factors [62]. Eye tracking calibration, together with signal filtering, can generally address these issues [61]. However, there is no guarantee that they will altogether remove accuracy limitations and signal noise [43]. Gaze interaction is limited in its use when a high level of granularity is needed [87]. Furthermore, if a gaze-based pointer is off-target after calibration, there is no direct way for the user to nudge the cursor to the actual position [136].

1.3 Research Questions

The overarching goal of this thesis is to position gaze as a multi-modal input, with movement from not just the eyes but also the head and in extension the body for 3D interaction. To achieve this, we aim to extend our understanding of gaze movements in 3D environments and provide novel capabilities for gaze-based interaction by incorporating our natural coordination. With this in mind, we ask two fundamental questions about eye and head coordination in the context of gaze-based interaction:

RQ1: *How do the eyes and head coordinate in VR?* Eye and head coordination has been investigated for decades within neuroscience [45]. However, we are interested in how these findings apply to the specific domain of head-mounted XR interaction.

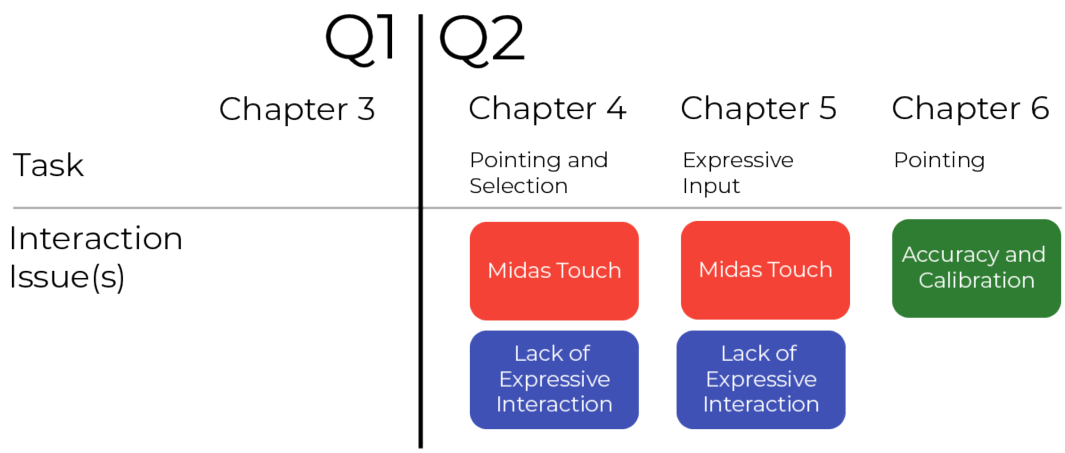

Figure 1.1: The underlying structure of the main chapters of this thesis

RQ2: *How can insights on eye-head coordination improve gaze-based interaction?* Gaze-based interaction has several fundamental interaction problems that are currently being researched [74]. Knowing how the eyes and head coordinate, we investigate how these insights can address these limitations.

1.4 Methodology

Figure 1.1 represents the structure of the work presented in this thesis. To answer the proposed research questions, we iteratively design, develop, and evaluate interaction techniques and systems. Where appropriate, we use application examples to explore the design space and ground our work with empirical data gained from data collection. To establish a deeper understanding of the main research questions, we ask relevant sub-questions within each section of work. Although our work is relevant to the broader space of XR and wearable eye tracking, we use VR as our primary platform. VR allows us to fully control the environments and conditions of study while allowing access to all movement systems in a common environment.

We begin with **RQ1** by conducting a VR study of gaze shifts. We identify general eye, head, and torso coordination patterns and analyse the contribution of relative movements and temporal alignment. We quantify the effects of target distance, direction, and user posture, describe preferred eye-in-head motion ranges, and identify high variability in head movement tendency. The insights from the study lead us to propose gaze zones that reflect different contribution levels from eye, head, and body. We discuss design implications for HCI and VR.

Based on the results of the first user study, we start to address **RQ2** by identifying design principles for eye and head coordinated gaze pointing and selection with the specific focus on the problems of *Midas Touch* and *lack of expressive interaction* for pointing and selection. We introduce three novel techniques that build on the distinction of head-supported versus eyes-only gaze to enable dynamic coupling of gaze and pointer, hover interaction, visual exploration around pre-selections, and iterative and fast confirmation of targets. We demonstrate the techniques on applications in VR, and evaluate them against baselines in pointing and confirmation studies.

We further build on the results of previous designs and continue to address the problems of *Midas Touch* and *lack of expressive interaction* for more expressive input beyond simple pointing and selection. We develop *Radi-Eye*, a novel pop-up radial interface designed to maximise expressiveness with input from only the eyes and head. Radi-Eye provides widgets for discrete and continuous input and scales to support larger feature sets. Widgets can be selected with Look & Cross, using gaze for pre-selection followed by head-crossing for selection and manipulation. The technique leverages natural eye-head coordination where the eye and head move at an offset unless explicitly brought into alignment, enabling interaction without risk of unintended input. We explore Radi-Eye in three AR and VR applications and evaluate the effect of radial interface scale and orientation on performance with Look & Cross.

Finally, we extend our learnings from **RQ1** to address the problem of *accuracy and calibration issues* by classifying head-shifts and introduce BimodalGaze, a novel technique for seamless head-based refinement of a gaze cursor. The technique leverages eye-head coordination insights to separate natural from gestural head movement. BimodalGaze allows users to quickly shift their gaze to targets over larger fields of view with naturally combined eye-head movement and refine the cursor position with gestural head movement. We compare BimodalGaze to a manual baseline in a VR user study and provide an in-depth analysis of user performance to provide insights into the classification of natural versus gestural head movement for future applications.

1.5 Contributions

The work presented in this thesis makes the following contributions:

- **Introducing gaze as multi-modal.** An introduction to the HCI field of fundamental knowledge of eye, head, and body coordination of relevance to any form of visual interaction that involves attention shifts over wider fields of view.

- **A detailed description of eye, head and body coordination patterns during gaze shifts.** We study the nature of gaze shifts in VR. This includes the contribution and temporal alignment of movements, effects of gaze amplitude, gaze direction, user posture, and preferred motion ranges, which establishes a point of reference for design of novel interactions that integrate input from eyes, head, and body.

- **Design guidelines on eye, head, and torso movement and their coordination for gaze and visual attention.** We propose several design guidelines based on eye, head, and torso coordination for XR and interaction; thereby providing designers with practical guidance on eye, head, and torso movement and their coordination for gaze and visual attention.

- **Design, application demonstrations and empirical evaluation of eye and head coordinated interaction techniques.** We propose a set of interaction techniques that utilise the knowledge of eye and head coordination to address limitations of gaze-based interaction.

  - **Eye&Head Interaction:** We present three pointing and selection techniques to address the Midas touch problem and the lack of expressive interaction. *Eye&Head Pointing* uses head movement as a clutch to separate gaze from pointing to allow users to point and inspect with their gaze. *Eye&Head Dwell* separates the dwell mechanism from gaze to allow free exploration of interfaces without triggering accidental interactions. *Eye&Head Convergence* that utilises the natural misalignment between the eyes and head to allow free exploration and movement with gaze without triggering selections.
  - **Radi-Eye:** We describe a radial interface for on-demand and hands-free object control via gaze and head input. The interface is designed to address gaze limitations such as low interaction granularity and lack of expressive controls by distributing the interaction to both the eyes and head while limiting accidental interactions by leveraging eye and head coordination insights.
  - **BimodalGaze:** We contribute a pointing technique that uses head movement to refine the gaze position to address the sensing difficulties of eye tracking. The technique uses eye and head coordination insights to classify head movements as natural parts of gaze shifts or gestural for pointing refinement.

- **Three general design principles for gaze-based interaction using eye and head coordination.**

  - **Differentiate between head-supported and eyes-only gaze shifts.** Gaze shifts can be performed with or without accompanying head movement depending on a wide array of factors such as amplitude, posture, task, etc. We discuss how to use this information for interaction.

- **Use head-eye alignment to signal intent.** The eyes and head are naturally misaligned to preserve energy. This aspect of our gaze behaviour can be utilised for a variety of uses.
- **Differentiate between natural and gestural head movements.** Head movements as part of gaze shifts are commonly used in XR environments, as content is presented around users. These natural head movements should be filtered out if head movements are used for interaction to avoid accidental interactions. Eye and head coordination insights can be used for this purpose.

1.6 Thesis Structure

The thesis is structured as follows:

Chapter 2 begins by detailing the physiology of the eyes and different eye movements. We then provide an overview of the development and state-of-the-art of gaze interaction within desktop and XR settings. We further cover the gaze interaction challenges that we explore in this thesis: Midas Touch, lack of expressive interaction, and eye tracker calibration and accuracy.

Chapter 3 introduces a study of the eye, head, and torso coordination during gaze shifts to a target within and outside the FOV using VR. Based on the data, we provide an analysis of the contribution and temporal coupling of the eye, head, and torso toward target-reaching, and of eye and head motion ranges. The outcome is a detailed description of movement coordination patterns, the effects of amplitude, direction, and user posture on relative movements, preferred ranges for eye-in-head motion, and user variability. We further detail design guidelines based on our findings. The work presented in this chapter was originally published in the ACM Transactions on Computer-Human Interaction (TOCHI) [151].

Chapter 4 introduces *Eye&Head* interaction. We identify design principles for coordinated eye and head movements for gaze-based interaction. We develop three pointing and selection techniques based on the design principles. The techniques are demonstrated in three application examples and two user studies are conducted to validate the principles and techniques in comparison to existing baseline techniques. The work presented in this chapter was originally published in the 32nd ACM User Interface Software and Technology Symposium (UIST '19) [152].

Chapter 5 introduces *Radi-Eye*, a radial interface that extends the principles and techniques of the previous chapter for more expressive hands-free interface control. We discuss the components and capabilities of the proposed interface. Permutations of the interface and their advantages are demonstrated in three AR and VR applications. Components of the interface are then studied and compared in a user study. The work presented in this chapter was originally published in the 2021 ACM CHI Virtual Conference on Human Factors in Computing Systems (CHI '21) [157].

Chapter 6 introduces the concept of using eye and head coordination to distinguish between *natural* and *gestural* head movements. We operationalise this concept with *BimodalGaze* - a head-based refinement technique of a gaze cursor. We evaluate the technique in a VR user study against a manual baseline. Finally, we provide a detailed analysis of participant behaviour with our technique that provides insight into different causes of error and how they relate to design choices. The work presented in this chapter was originally published in the 12th ACM Symposium on Eye Tracking Research and Applications (ETRA '20) [155].

Chapter 7 discusses the work presented here, and specifically details the lessons learnt in designing eye-head coordinated interaction techniques and interfaces. It also provide a reflection on eye the design principles, gaze interaction issues and research questions presented in this thesis.

Finally, **Chapter 8** concludes the thesis.

Chapter 2

Related Work

This work studies the coordination of eye movement with the head from an HCI perspective. To contextualise our work, we first present fundamental eye physiology and types of eye movement. We then review research in HCI and XR, first examining different ways in which work has reflected on gaze and its fundamental problems for interaction and also briefly cover existing challenges for gaze-interaction.

2.1 Eye Physiology and Eye Movements

The human eye consists of multiple connected parts (Figure 2.1). Light enters through the cornea and the pupil, which changes size when the iris contracts or relaxes to regulate the amount of light entering. The light then passes through the lens, converging to hit the retina. To accommodate different depths, ciliary muscles change the curvature of the lens to focus the light rays to hitting the retina. The retina is filled with light-sensitive cells called *"cones"* and *"rods"*. Cones provide us with colour vision, and rods are more sensitive to light and support vision under dim light conditions. A small area of the retina called the "fovea" (Figure 2.1), has an extreme over-representation of cones compared to other parts of the retina [193]. This has the result that we have full aquity only in this small area (about two degrees in diameter) [38]. This means that to see an object sharply (e.g. a word in text), we have to move our eyes so that the light from the word falls directly on the fovea [61]. Therefore, the eyes need to move constantly to explore a scene and acquire a sharp image of the surroundings. This is achieved through three pairs of opposite extra-ocular muscles that perform different movements [61, 88], mainly:

Saccades. These are the fast movements the eyes make constantly to explore an environment. Saccades are extremely quick and can reach up to 500 degrees/second. They are ballistic movements, which last between 20 to 120

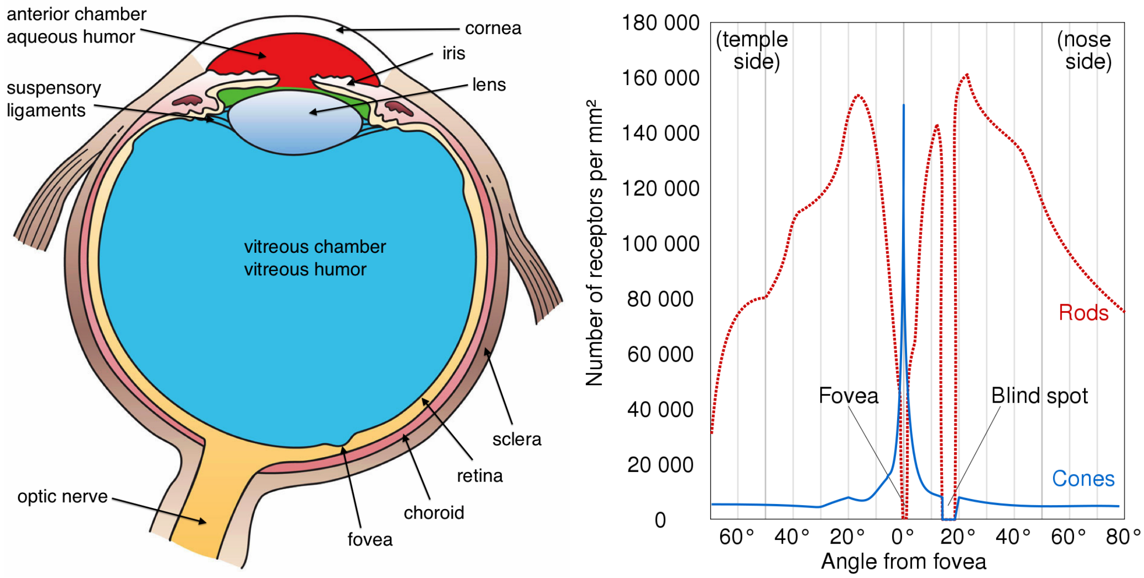

Figure 2.1: Left: Details of the Human Eye (from Wikimedia Commons). Right: distribution of rods and cones over the retina. Created by Cmglee, from Wikimedia Commons.

ms depending on their amplitude. Once a saccade starts, the eyes cannot stop moving until it reaches its landing position.

Fixations. These are the periods of time during which the eyes remain stationary on a target. During fixations, the brain acquires a still image of the scene for neural processing and typically last between 200 and 600 milliseconds. Fixations are considered to be movements as the eyes constantly perform micro-saccades around the point of focus and exhibit very small tremor.

Smooth pursuits. Smooth pursuits are smooth movements performed by the eyes when following a moving object, such as a bird. Smooth pursuits have a maximum velocity of 90 visual degrees/second; beyond this speed, the pursuit is interrupted by catch-up saccades to keep up with the moving object.

Vergence. This movement adjusts the eyes to avoid double vision when the foveated object moves closer or farther away. The eyes can converge (rotating towards each other) to adjust to a closer object, and diverge (rotating away from each other) to adjust to an object farther away. Vergence movements are the only movement in which the eyes move in opposite directions.

Vestibulo-ocular reflex (VOR). This compensatory movement readjusts the position of the eyes while the head is moving to stabilise the line of sight and remain gaze on an object. This maintains a clear image on the retina, which would otherwise be blurred whenever the head is moving.

As our visual organ, the eyes are the main sensor available to humans to comprehend the surrounding environment. A sequence of fixations reveals a person's region of attention, and can indicate a person's intention and future task. For example, a person looks at a kettle before reaching for it while making tea [90]. The need to move our eyes toward a target is the basis for eye tracking: it is possible to deduct the gaze vector by observing the "line-of-sight".

2.2 Gaze-based Interfaces

Gaze-based interactive systems involve interpreting user eye movements and mostly focus on the user's ability to look at a point (fixation), and between points (saccades) on a virtual screen. Gaze-based interaction was first introduced in the eighties, mainly to provide users with limited motor control an alternative to mouse input [19, 68, 190]. Since then eye tracking has been used for a wide variety of applications and contexts for both explicit and implicit user input. Explicit gaze-based interaction techniques use gaze as the main modality of input, where the user controls the user interface by explicitly looking at widgets and controls. The user's eyes act as a pointer similar to a cursor controlled by a computer mouse. In these techniques, the eyes can also be supplemented with other modalities for unique interaction behaviours. Implicit gaze-based interaction techniques use the information from eye movements in the background to make decisions. In contrast to explicit interaction where the eyes act as a replacement to manual input, implicit interaction use the eyes as a complement to the main manual modality to improve the task. We further look at interactive systems that leverage gaze-based gestures instead of the gaze point, and techniques that leverage additional modalities together with gaze.

2.2.1 Gaze Pointing and Dwell

The first system that used gaze for input was *"Gaze-orchestrated Dynamic Windows"* by Bolt [19] in 1981. The user interface consisted of multiple windows (i.e., screens), each gaze-interactive. The system was designed to help users cope and manage the "onslaught of events" that was predicted to accompany the digital adoption by professionals. For example, a video was active/playing as long as the user looked at it. At the same time, other videos were automatically paused to reduce the visual and cognitive overhead that comes with having multiple available windows while also enabling rapid access to any video. Bolt discussed multiple challenges with using gaze for interaction, such as the system's trade-off of *"to be sure that you want to do [the* *action], yet [the action] shouldn't hesitate too long."* A dwell-time technique where the user fixates at a window for a set time (e.g., 300 ms) to trigger an action was considered as a method based on the eyes while using manual input or speech was proposed to allow users to trigger gazed-on actions explicitly.

Ware and Mikaelian [190] presented the first study of the performance of gaze interaction. This study aimed to investigate the limitations of gaze for interaction and was mainly concerned with target size and confirmation of selection. Comparison of target sizes was motivated by the inherent jitter in eye movements and the accuracy of the eye tracking calibration to find appropriate sizes for gaze-selected targets. In addition, different confirmation mechanisms (click, dwell, and a saccadic gesture to a button after fixating on the target) and their impact on user performance were evaluated and compared. The study results highlighted two main points that have proven relevant for a large body of gaze-based interaction. First, the gaze techniques proved faster in performing selection than other modalities, highlighting the main advantage of gaze being faster than other modalities. Second, gaze input is only fast for large targets where the accuracy and noise limitations of the eyes and trackers are not factors. If targets are small, selection speed is prone to decline, and errors increase.

In 1990, Jacob [74] investigated gaze input in conventional user interfaces with the principle "What You Look at is What You Get". Jacob argued that the main barrier to introducing gaze-based interaction to the wider public is not related to the technology but rather making gaze-based interaction natural and unobtrusive. In this seminal work, Jacob introduced the *Midas Touch* problem that describes the fundamental ambiguity of using gaze for selection. When a user looks at a target, it is unclear whether they want to select it or look at it. As there is no inherent "confirmation technique" to gaze (e.g., click), finding a mechanism to distinguish between these two intentions has since then been a central part of eye tracking research, as will be further discussed in the following sections. Like Bolt in his early work, Jacob suggests either using dwell-time or a manual confirmation for selection, depending on the application scenario. A manual button press can be faster and appropriate for manual tasks, while dwelling is entirely hands-free, reducing workload and effort.

During this time, eye tracking also received significant attention as an accessibility tool for users with severe motor disabilities who could not rely on the typical mouse and keyboard. Notably, *ERICA* (Eye-gaze Response Interface Computer Aid), was iteratively developed during the eighties and nineties to provide a holistic integration of eye gaze input into the Microsoft Windows operating system [68, 91]. ERICA provided clicking and typing with gaze-only input and enabled the control of various tasks within the operating system. Clicking and typing was provided by dwelling where users had to fixate on a button or virtual key. To select small targets, the first fixation on an area expanded the area while the second dwell could be used for target selection. Overall, ERICA provided the first interaction style to holistically interact with an operating system.

These early works have in common in that they originate from the cursor metaphor and use gaze as a replacement for the mouse to integrate with existing windowing systems. Saccades are used to move the cursor and fixations to keep the cursor still. This research has highlighted the potential of gaze-based interaction as a quick and hands-free alternative for interaction. However, researchers also uncovered fundamental challenges to the widespread adoption of gaze as an interaction modality.

2.2.2 Implicit Gaze Interaction

In conjunction with work on explicit gaze interaction, researchers investigated implicit interaction. Starker and Bolt [164] conducted early work in 1990 with *Gaze-Responsive Self-disclosing Display*. Their system operated *interpretively* by aggregating user fixations to make inferences about what part of the screen holds the most interest to the user for interaction. Zhai et al. [202] first argued for the implicit use of gaze in combination with a second modality with their paper on Manual and Gaze Input Cascaded (MAGIC) pointing. The authors argued that it is unnatural to overload the eyes' visual channel with motor control and that the eyes are fundamentally limited in their accuracy. They further say that interaction should mainly be manual while using the information from the eyes to support interaction to avoid these issues. MAGIC pointing achieves this by leveraging the natural eye-hand coordination where the eyes are used to guide hand movements (Figure 2.2). Warping the cursor to the gaze position when the mouse moves. Users then perform fine-grained movements and selection with manual input. The MAGIC work showed that considering the natural coordination of the eyes and other modalities could address limitations of gaze-based interaction.

Furthermore, using eyes implicitly to adapt interactions and interfaces to the user's attention has been explored in work on Attentive User Interfaces [185]. We can find these and similar work in application scenarios such as automatic scrolling [85], translation of gazed-on text [70] or with appliances activated by gaze such as attentive cell phones, TVs, light bulbs, or sound systems [148, 187].

These works highlight how our eyes are implicitly used in coordination with other modalities to guide our interaction with the world. As we reach for and interact with objects via our hands, our eyes are used to guide the hands. Furthermore, as we explore our surroundings, gaze is directed towards objects of interest that can be leveraged for interaction. All these works show how gaze interaction can be extended beyond eye movements for simple cursor control to be leveraged in coordination with other modalities or implicitly through inferring attention and interest.

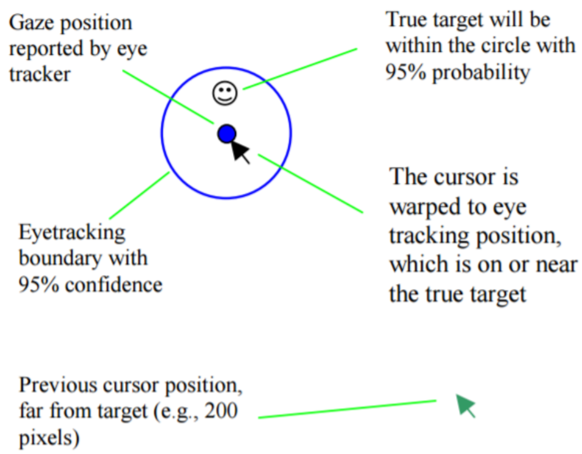

Figure 2.2: MAGIC pointing: the cursor warps the the target area (blue circle) followed by manual position refinement. Image from Zhai et al. [202].

2.2.3 Behaviour-based Gaze Interaction

The calibration and accuracy limitations of gaze pointing discovered by Ware and Mikaelian [190] spurred research that used relative eye movements to avoid the need for good accuracy. Early work on *gaze gestures* looked at users performing saccades between different points on the screen and used the saccade directions to infer the gesture performed [37]. This concept has since been further investigated as a replacement or complement to manual input for contexts where an accurate calibration may be difficult to achieve, such as mobile phones [35, 81] and wearables [21].

In addition to saccades, researchers have investigated eye movements other than saccades and fixations for gaze interaction. Vidal et al. [188] developed Pursuits in their seminal paper where users follow a moving target with smooth pursuit movements, and a correlation between the movements infers intention. Pursuits enable gaze selection of targets without prior calibration, as demonstrated for walk-up-and-use gaze interaction with public displays [188], gaze input at a glance on smartwatches [40], and gaze control across ambient devices [184]. In particular, pursuits avoid the Midas Touch problems of fixation-based gaze techniques, as the eyes only exhibit smooth pursuit when the user attends to a moving object.

The work on gaze gestures and pursuit interaction highlight how gaze interaction extends beyond saccades and fixations. The eyes perform a wide variety of movements that interaction designers can leverage in new ways to tackle interaction challenges. We draw inspiration from these works when considering how gaze interaction can extend beyond leveraging only the eyes during gaze shifts.

2.2.4 Multi-modal Interaction

The introduction of MAGIC pointing [202] inspired multiple works where gaze is combined with a second, more accurate modality to circumvent accuracy limitations and avoid Midas touch issues. Researchers have expanded the MAGIC technique to include implicit calibration [42] and other modalities such as the head [86]. Furthermore, a wide addition of techniques have been developed that investigated the use of a second modality in conjunction of gaze. Touch has received notable attention from researchers where techniques utilised our natural hand-eye coordination [12, 90, 124]. Touch and gaze techniques have been developed for remote display interaction with a hand-held smartphone and gaze [165, 166, 167, 168], interaction across displays or for out-of-reach parts of displays where gaze is for pointing and touch for confirmation [175, 176, 177, 178, 179], or to enhance direct touch manipulation with a combination of gaze and touch on tablets [126, 127, 128, 130].

The head has also been commonly combined with eye tracking in personal computer display settings where, at typical viewing distances, the display width is usually within 40° visual angle [149]. In such a setting, the display is viewable with eye-in-head rotations of up to 20° from a central position, and therefore does not require any significant head movement. Head movements performed by the user can therefore be assumed to be gestural. Early work compared eye tracking against head tracking for desktop pointing interaction as an assistive technology and found that the head was more accurate for pointing than the eyes but the eyes were faster [15]. Researchers further investigated techniques that combined the two modalities for interaction. Špakov and Majaranta [161] investigated simple head gestures as an alternative to dwell confirmation. These head movements are detected via an eye tracker where changes in eye position relative to the camera position are inferred as a head movement for interaction. Similarly, Mardanbegi et al. [104] implicitly leveraged eye-head coordination by developing a system that infers head movements for interaction through eye tracking based on VOR movements performed by the eyes to keep fixating on a target during head movement. Špakov et al. [160] further developed the concept to infer head movement directions and amplitudes to use as a pointing refinement when an eye tracker may be too inaccurate or noisy for pointing and selection. Nukarinen et al. [119] further extended these concepts and applied them to continuous interaction in which users turn their heads to manipulate the values of, for example, a slider.

These works show how combining gaze with other modalities can leverage the strengths of gaze-based interaction, allowing quick and effortless pointing while avoiding its known weaknesses such as accuracy and the Midas Touch. The choice of modality also gives users different affordances. Combining gaze with the hands allows total control with the hands but with less required hand movement, while combining gaze with the head enables more accurate and controlled hands-free interaction.

Furthermore, the work also shows that the setting impacts the assumptions and movements that interaction designers can leverage. In a desktop setting, head movements as part of gaze shifts are rare and can be assumed to be part of the interaction. The implicit assumption for all these head-based works is that head movements and eye movements are considered to be separate and not related.

2.3 Gaze Interaction in Virtual 3D Environments

Concurrently with research on gaze-based interaction, researchers within the field of XR have investigated the use of gaze for head-mounted displays. Eye tracking has been stated to be a vital technology for XR HMDs to increase graphical performance through foveated rendering [122], increase presence and collaboration in multi-user settings [48, 79, 118, 133], infer attention and intention [4, 31, 49, 107], and most important for this thesis a means for effortless hands-free interaction. For a complete review of the use of eye tracking in XR, see the work by Plopski et al. [135].

2.3.1 Comparison of Gaze and other Modalities

XR has access to a multitude of modalities for interaction. Early gaze interaction work compared the performance of gaze with other modalities. In these works gaze is used as a *ray-casting* technique, where a virtual ray is created from the eye position and gaze direction in the virtual world to point at targets. Early works compared gaze pointing with controller-based pointing where the user points at targets by creating a ray based on the controller position and direction. In the first comparison of gaze and controller pointing in 3D environments, Tanriverdi and Jacob [170] found that gaze pointing was faster than controller-based pointing, especially for distant targets where hand jitter has an increased impact. However, gaze-based selection worsened participants' ability to recall spatial information of targets, presumably because of the less body movement required for gaze pointing. On the contrary, the follow-up work by Cournia et al. [29] showed that gaze-based pointing was, in fact, slower or equal to hand-based pointing. Furthermore, Hülsmann et al. [67] found similar results and identified gaze-based pointing issues mainly relate to calibration and accuracy. These early works showed the promise of gaze-based pointing as a hands-free and effortless interaction modality. However, they also exposed challenges for future work.

Early work on interaction in VR also proposed gaze-directed input but treated gaze as synonymous with head orientation [111, 201]. State of the art interactive devices similarly associate head pose and gaze. For example, the Smart Eye Pro 3D eye tracker supports a tracking mode based on a head model instead of eye movement, and Microsoft's HoloLens 1 assumed head pose as the user's gaze vector. This reflects the major role head movement plays in contexts where gaze is not focussed on small viewing areas, but is engaged with wider environments. However, the general assumption that we look where our head points is problematic. We address this in future chapters with an in-depth analysis of how head and gaze relate.

The introduction of integrated eye tracking into consumer version HMDs in the 2010s spurred a resurgence for gaze interaction in the VR community. Multiple studies compared gaze with a wider array of modalities, in particular the head, in VR and wearable eye tracker settings. In these works, gaze-based pointing is usually described as *eye* pointing, ignoring the contribution of the head and body towards gaze. Qian and Teather [138] performed the first study comparing Head-only, Eye-only (disengaged head-tracking) and *Eye+head* (gaze) pointing and found that Head-only pointing was significantly more performant than using Eyes-only or gaze. Furthermore, multiple studies compared gaze in settings where head-tracking was disabled thus only allowing eye movements as gaze. In these studies, gaze tended to perform worse than head pointing [77, 110, 138, 139]. Meanwhile, other studies that allowed free head movement during gaze interaction found that gaze pointing was faster than head pointing, while the head was more accurate [18, 56, 121].

An interesting observation from these comparison studies is that gaze pointing and head pointing are consistently treated and discussed as separate or competing modalities, rather than coordinating.

2.3.2 3D Gaze Interaction

Researchers have also investigated applying gaze to different contexts such as gaze typing [141], and object manipulation [97, 98] or subtitles [153]. Furthermore, a myriad of gaze techniques have been developed beyond basic ray-cast pointing. These techniques are commonly inspired from work on desktop settings, and either leverage eye movements other than fixations and saccades, incorporate additional steps to the interaction, or pair gaze with other modalities.

2.3.2.1 Gaze-only Interaction

In contrast to desktop settings, XR introduces elements displayed at different depths. This aspect has been leveraged for multiple interaction techniques. Vidal et al. [189] first proposed using vergence for interaction by detecting via vergence whether the user is looking at the virtual screen at a close distance, or the world through the screen, thus hiding the screen content. Similarly, Hirzle et al. [60] used the same principle to propose x-ray vision, where users can look through objects by focussing at a depth farther away. Furthermore, Ahn et al. [3] used vergence movements for selection of moving targets as an alternative to dwell. Researchers also leveraged depth to distinguish between targets or perform commands using VOR movements [103, 105, 134]. These works implicitly leverage eye-head coordination for interaction, but only use sensing of one for interaction.

Researchers have also investigated the transfer of existing gesture-based and dwell-based techniques from desktop to 3D interaction. Piumsomboon et al. [134] developed a dwell-based technique that uses a delayed cursor that follows gaze movements and an alignment of gaze and cursor lead to a dwell-bed selection. Furthermore, the virtual button technique originally developed by Ware and Mikaelian [190] was redeveloped and evaluated in VR [112]. Ahn et al. [2] investigated the use of saccadic gestures for marking-menu-based interaction. Delamare et al. [33] and Jungwirth et al. [80] re-explored gaze-gestures in AR settings to activate physical appliances and virtual objects. Furthermore, several work investigated smooth pursuit gestures in VR where findings from desktop settings are applied and transferred to 3D interaction [83, 134, 150].

The transition from 2D to 3D environments has extended the use of gaze to include eye movements that are reactive to depth changes. Furthermore, 3D environments allow researchers to leverage well-established eye movements for interaction in novel ways. However, a common theme among these works is that while the interaction has been transferred from desktop to XR, interaction designers' mental model of gaze has not changed. Just as in desktop settings, gaze is only considered as eye movements, omitting the movements performed by the head or body.

2.3.2.2 Multimodal Interaction

Finally, taking inspiration from multi-modal work on desktops and touch interfaces, multiple researchers have investigated the use of gaze combined with the hands. Zeleznik et al. [201] originally developed gaze and hand techniques that were aimed at minimising hand movement and allowing "lazy" hand interaction, where the hands can be used from a resting position together with gaze for granular movements and interaction confirmation. Since then, established techniques such as MAGIC pointing have been developed for 3D gaze and hand interaction [76], and the combination of gaze and hand interaction has been explored for contexts such as menus [131, 132], object manipulation [131, 200], travel [131], and pointing refinement [87]. These works all show the potential and increased control and expression afforded by leveraging our natural coordination between systems. Recent work by Kytö et al. [87] has started to explore the usage of gaze and head in interaction techniques for pointer refinement (Figure 2.3 right).

These works show, just as in 2D interaction, that combining gaze with other modalities can improve the affordances of gaze-based interaction and control while minimising strain caused by large body movements (Figure 2.3 left). Recent work has started to consider the head in combination with gaze. However, these works

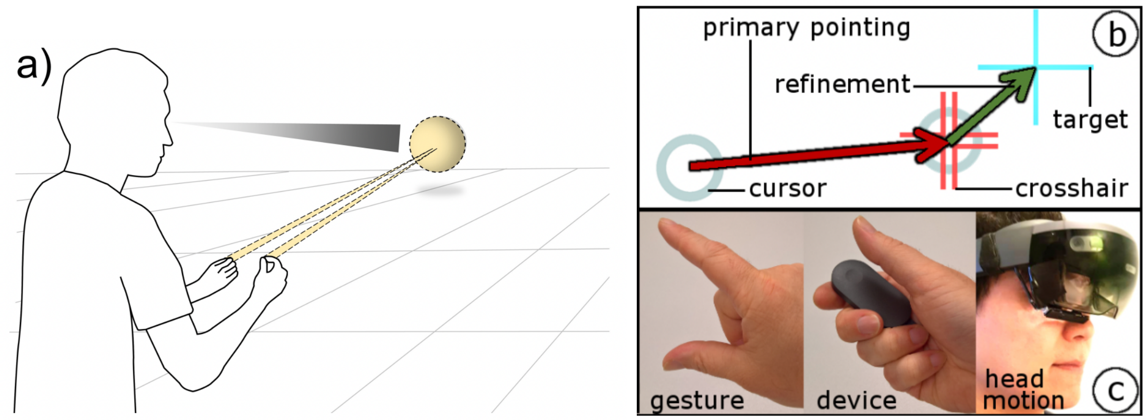

Figure 2.3: Left: Gaze + Pinch interaction where users first look at a target and then manipulates it with hand interaction. Image from Pfeuffer et al. [131]. Right: Pinpointing techniques consist of a primary pointing motion (e.g. gaze) plus secondary refinement (e.g. hand or head motion). Image from Kytö et al. [87].

apply gaze and head pointing sequentially and the synergistic relationship between the movements is not considered. By drawing inspiration from the extensive work on eye and hand coordination, we can leverage the coordinated relationship between our eyes and head to tackle gaze interaction challenges in novel ways.

2.4 Gaze Interaction Challenges

While significant research has been made on gaze interaction in both desktop and 3D setting, challenges that were first discovered and discussed in the eighties and nineties are still hindering the widespread usage of gaze-based interaction by the wider public. This section highlights gaze interaction challenges and previous research on tackling them.

2.4.1 Midas Touch

While conventional pointing with a mouse uses a cursor metaphor, it is less clear how best to provide feedback with gaze as eye movement is primarily engaged in information seeking rather than target identification [75]. Coupling gaze with continuous feedback supports target selection but can be distracting when it follows every gaze movement. Furthermore, using a cursor metaphor can also be distracting as the cursor will be always be in front of the user's gaze [15] and may occlude regions

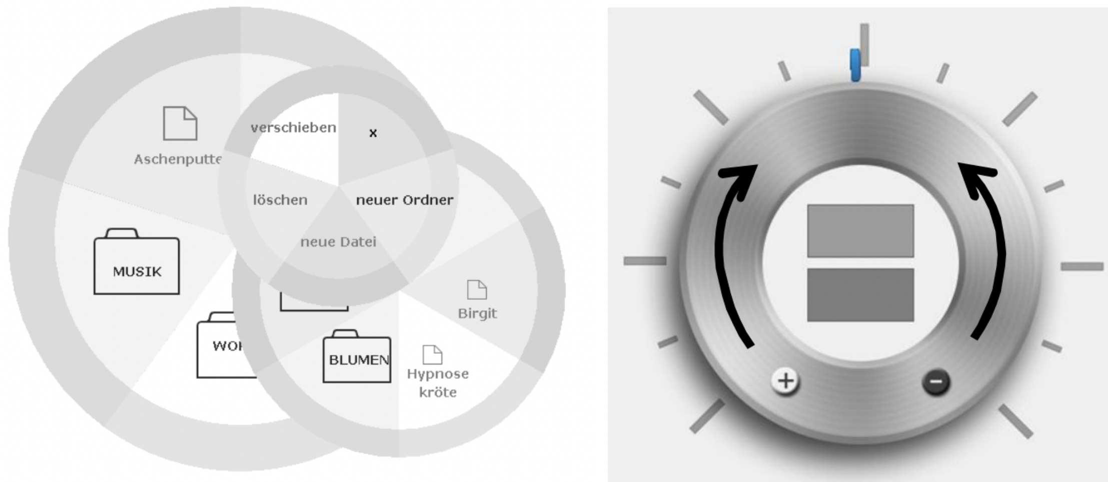

Figure 2.4: Left: In pEyes, users traverse through menu levels through fixations on widgets enabling nested interaction. Image from Huckauf and Urbina [66]. Right: PursuitAdjuster enables continuous interaction by users following the movement of the "-" and "+" widgets to change the handle value (blue). Image from Špakov et al. [159].

of interest. Research has attempted to address these issues through dwell methods where no interaction is performed and no target feedback executed until the user has fixated on a target for a significant time [69], or through combining gaze with a keyboard or button click to decouple the confirmation from gaze [186]. These techniques are especially common in 3D environments where gaze is used for pointing and a controller is used for confirmation [138, 139]. However, such methods remove the benefits of hands-free interaction. Furthermore, such techniques also do not solve the fundamental issue that the user is unable to decouple the cursor from gaze. Hands-free techniques that give the user opportunity to gaze on objects without activating interactions is still an open research topic.

2.4.2 Lack of Expressive Interaction

Gaze is "always on", necessitating a confirmation mechanism to distinguish between information seeking and input and perform fundamental actions such as selections and manipulation. However, in contrast to hand interaction, gaze lacks an explicit "click". For the confirmation phase, the most common eyes-only technique is dwell selection by prolonged fixation of a target [55, 74, 190], while it is also possible to use eye gestures [113, 142], smooth pursuit if targets are in motion [188], or additional confirmation buttons [101]. Eyes-only techniques need to be based on gaze behaviours distinct from natural viewing and consequently can be experienced as awkward and tiring. Alternatively, eye gaze can be combined with a separate modality for confirmation, such as key, mouse or button click, which also enable higher throughput [203]. However, such techniques do not provide fully hands-free interaction. Purely gaze-based interactions are limited in the control and expressiveness they afford and tend to be limited only toward discrete input.

Research has attempted to extend the capabilities of gaze-based interfaces by including more expressive and advanced widgets. Nested interaction has been investigated by fixation on widgets to enter submenus (Figure 5.5 left) [66]. Alternatively, entering a submenu can be performed by saccading across widget borders [3]. Furthermore, Špakov et al. [159] studied the adaptation of smooth pursuit widgets for continuous control of dials and sliders (Figure 5.5 right). They found that while participants could control the widgets, the interaction caused eye strain and was slow. Furthermore, the interaction used audio-based feedback as the eyes were used for extensive widget control and visual feedback could distract the user and interrupt the interaction. Gaze-based interactions that are fully hands-free and provide free viewing of the environment and expressive input remain an open challenge.

2.4.3 Eye Tracker Calibration and Accuracy

Accuracy issues and the requirement for user calibration are well recognised as critical issues for the wider adoption of eye tracking for interaction. Morimoto and Mimica [114], in their survey, point specifically to calibration and accuracy issues as a critical usability issue. Furthermore, it is common to find studies where users had to be discarded from results due to calibration issues or poor accuracy [3, 9, 36, 43, 105, 118, 121]. Specifically for AR and VR HMDs, slippage of the HMD relative to the user's head is a prevalent cause of accuracy error [117].

Consequently, handling accuracy errors and calibration are central parts in eye tracking research. A common approach for gaze-only interactive systems are to develop interaction techniques that only rely on relative movements such as smooth pursuits [188] or vergence [103, 105] to circumvent the need for an accurate calibration, or through "error-aware" interactive systems that can adjust widget sizes [43] or the gaze point [14, 42, 156] to cope with existing errors. Furthermore, as highlighted in previous sections, a large body of research has investigated combining gaze with a second, more accurate modality such as the hands [42, 87, 130, 166] or head [86, 87, 160] for accurate interaction.

2.5 Conclusion

As we have seen, the eyes contain a lot of information that has been leveraged in various ways. However, key issues remain for gaze-based interaction that inhibit their broader adaption. Furthermore, the adaptation of eye tracking into 3D environments brings a set of new challenges that need to be overcome for gaze to be fully leveraged in these settings. However, as eye tracking moves from stationary desktop environments to mobile settings where gaze behaviour mimics the real world, opportunities arise to leverage new user behaviours and insights for interaction.

We find several knowledge gaps. First, the mental model of gaze consists of only eye movements, even though XR enable full-body exploration. Second, the eyes and head have been extensively studied separately and in succession. However, their coordination has so far been ignored. Third, the specifics of eye and head coordination in 3D environments and how to best leverage these for interaction remain unknown. In this thesis, we investigate the coordination between eye and head that arise as we explore our surroundings in VR settings and leverage these behaviours to tackle fundamental gaze-based interaction challenges.

Chapter 3

Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality

As a first step to investigate eye and head coordinated interaction, we need to identify the common patterns of eye and head coordination performed during gaze shifts in XR (RQ1). This chapter studies the coordination of eye movement with head and body within XR from an HCI perspective. We achieve this by conducting a study where participants performed gaze shifts towards targets at different directions, amplitudes and postures and provide common movement patterns and design implications towards the design of gaze-based interfaces. To contextualise our study, we first review fundamentals of eye-head gaze shifts established in neuroscience, and existing understanding of factors in the head contribution to gaze. Based on this, we review research in HCI on eye, head and body coordination related to gaze and XR.

3.1 Eye-head gaze fundamentals

The motion range of the eyes is approximately 50° in any direction for a healthy adult [88, 162]. The total FOV of human vision is 210° in the horizontal plane and 120° in the vertical plane (50° up, 70° down). However, gaze shifts are not limited by the FOV and can also be performed to targets beyond the FOV by turning the head. The neck has a motion range of about 80-90° for head movement in horizontal directions and 60-70° in vertical directions [44]. Eyes and head together thus provide a movement range of around 130-140° in horizontal directions and 110-120° in the vertical directions. Gaze shifts with a larger amplitude are only possible with additional use of the torso. With a combination of eyes, head and torso movement, humans can reach gaze targets on the full range surrounding them (unless torso motion is constrained, as for instance in seated positions).

The relationship of eye and head movement is complex. During a gaze shift, head movement augments the saccadic movement of the eye, such that the movements are additive toward reaching the target. The faster the head moves in synchrony with the eyes, the smaller the required saccade [54]. When a gaze target has been reached, the head will typically continue to move while the eyes fixate the target by performing compensatory eye movement in the opposite direction. This allows the eyes to rotate back into a more central and comfortable position relative to the head [181], mediated by the VOR which stabilises the visual axis on the target while the head is still in motion [13]. There is a similar interaction between head and torso rotations, with a corresponding vestibulo-collic reflex [89].

Eye-head gaze shifts and their relative timing were first described in the 1930s [115], and have been studied under controlled conditions since the 1960s [13]. A conventional study paradigm is to have light stimuli arranged in a semicircle or hemisphere around the subject, and task participants to initially fixate in the centre from where they shift to a flashing target. Subjects are seated and their body movement is suppressed by a belt, while bite bars and brakes control available head movement [54]. Since the 1990s, gaze has also been studied in naturalistic contexts, enabled by wearable eye-trackers that capture eye-in-head orientation relative to the visual scene which is recorded from a head-centred perspective [88]. These studies lack in procedural control but draw attention to the role the body plays in active gaze, in addition to eye and head [92].

Our study design is a lab-based experiment that reproduces a conventional hemispherical arrangement of stimuli in VR for comparability with prior studies. However, we are not constraining head and body as our goal is to inform design by capturing natural behaviour, in both seated and standing postures. By using VR we are able to record eye, head and body motion and their relative timing at higher accuracy than in previous studies. Head-mounted VR, however, limits the FOV in comparison with real-world studies. How this limitation affects gaze is of explicit interest, as one of our goals is to inform adoption of gaze for interaction in XR.

3.1.1 Factors affecting a head contribution to gaze

The amplitude of a gaze shift has been shown to have a significant effect on whether and how much the head contributes. Gaze shifts extending less than 20° are almost entirely made by eye movement [45, 52, 88]. While the eyes have a physical range of 50°, it has been found that they rarely rotate beyond 30° relative to the head [88]. In natural tasks, the head contributes about $\frac{1}{3}$ to gaze shifts of up to 30°, and more to gaze shifts that are larger [144]. Large gaze shifts can be achieved in a single-step head-supported saccade but may also exhibit stepping movements [89]. These observations provide us with a baseline for assessing whether the head contribution is comparable when people interact in a virtual rather than natural environment.

Gaze shifts made within the limits of the eye-in-head range may or may not be associated with a head movement, as observed in the lab [47, 120, 162] as well as naturalistic settings [39, 172]. The decision whether to move the head involves an internal weighing of costs and benefits, specifically energy required to accelerate and decelerate the mass of the head in a short time versus fixation accuracy and stability which decrease at far-eccentric eye-in-head positions [163]. Amplitude, as noted, is a significant factor but others have been identified, including the initial eye-in-head position [46, 47, 162], the expected duration of maintaining gaze in the general vicinity of the current target, and the position of the next target [120].

There are also multiple external factors that can affect head or torso movements. External factors such as environmental light, noise, or heavy equipment that cause strain can cause changes in head movement [108, 197]. User-specific factors, such as neck flexibility or disabilities, can significantly impact the amount of movement performed by users [109, 197]. Vision loss or vision aids, such as glasses, may also need to move their head further to see objects [28]. It has also been shown that there are individual differences in how much head movement is used [47, 120, 162, 171, 172]. Fuller proposed a notion of "head-movers" versus "non-head movers" [47] and other work suggests that observation of different types of head movement tendency carries over from controlled to real world contexts [171]. This prompts us to look closely at eye-head gaze patterns in our study, as differences in tendency to move the head are of obvious importance for head-mounted display paradigms.

3.1.2 Eye and head movement in VR, Head-mounted Displays and Virtual Characters

Several studies provide insight related to gaze behaviour and eye and head movement in VR and HMD contexts [84, 125]. In contrast to natural viewing, HMDs limit the user's view, blocking out the wider visual periphery as if wearing blinkers. This was observed to lead to less eye rotation and consequently more head rotation to achieve the same lateral shift of gaze direction [84]. However, the magnitude of eye rotations was still considerable when the HMD limits the FOV (12° in any direction), showing that head orientation alone is limited for predicting attention. A recent study comparing eye-head coordination in virtual versus physical reality likewise observed that display limitations induce more head movement in VR [125].

Related work has used head-mounted VR to study visual attention to content that surrounds the viewer and requires head movement for exploration [20, 64, 65, 140, 158]. Analogies have been drawn between eye movement in desktop viewing and head movement in VR viewing, as the latter exposes similar relationships between amplitude, duration and peak velocity [65]. In visual exploration, "head fixations" were observed during which subjects only made eye movements over a mean range of 18° horizontally, while larger shifts usually involve head movement [64]. Other work on visual saliency observed coupled head and eye movement, suggesting the head following the eye with an average delay of 58ms on the basis of cross-correlation of movements, with a mean gaze direction of around 14±12° [158] relative to the head orientation. Another study found that the distribution of gaze fixations did not peak at the centre of the view-port where they align with head orientation, but at a distance of about 14 degrees from the display centre, due to the exploratory nature of image viewing [140]. These studies provide insight into general correlation patterns of eye and head movement in visual exploration, whereas we focus on the systematic analysis of explicit gaze shifts and the relative movements by which they are accomplished.

Finally, there is also a substantial body of work on eye-head coordination in gaze animation for virtual characters and embodied conversational agents. Gaze models in the graphics literature reflect insights from neuroscience, for example rendering gaze shifts as eyes-only when their amplitude falls below a threshold of 10-15° [143], and accounting for factors such as target predictability [5]. These works are aimed at generating realistic gaze [72] or aiding animators in creating specific communicative effects (e.g., glances out of the corner of eye) [123]. In contrast, our focus is on gaze as input, and understanding of eye, head, and body coordination for the design of interaction techniques.

3.2 Study design and methodology

The aim of our study was to understand how eye, head and body movement is coordinated during gaze shifts in VR (RQ1), and how this depends on a variety of factors. Our main questions were:

- How are eye, head and body movement aligned in time?
- What is the relative contribution of each of the movement systems to reaching gaze targets?
- How much do head and torso move in total toward a target?
- What are preferred motion ranges for eye and head?

We identified four principal factors of interest, for investigation in our study:

Amplitude. The significant effect of gaze shift amplitude on eye and body movements is well established and our objective was to systematically quantify it for gaze shifts in VR. In our study, we cover a range of target amplitues from 5° to 100° in visual angles from a central position (Figure 3.1).

Direction. The mechanics of eye and head imply differences in effort and range for horizontal, upward and downward movement, raising the question how this affects the composition of gaze shifts. We study this on the basis of gaze shifts in the four cardinal directions, as well as diagonal directions (Figure 3.1).

Visibility. In VR, gaze targets can be visible within the initial FOV (within-view), or invisible in the wider FOR (beyond-view). In comparison with natural vision, HMDs provide a limited FOV (around 100° in contemporary devices) in which objects become more easily hidden from view in the course of interaction. We are therefore interested in how gaze shifts compare for within-view versus beyond-view targets (Figure 3.1).

Posture. Head-mounted VR can be used standing with more freedom of body movement, or seated for safety or comfort. Existing platforms and experiences differ in how they are geared for standing versus seated usage. For example, Playstation VR advocates seated use, whereas HTC Vive promotes room-scale interaction for which users typically need to be standing. We study gaze shifts in both postures as we expect a significant effect on the performance of gaze shifts (Figure 3.2).

The experimental design for our study was inspired by prior work on eye-head coordination in other fields, where gaze targets were arranged on a hemisphere, equidistant from the user, with gaze shifts initiated from a central position [52, 120, 162, 163, 171]. As shown in Figure 3.1, we placed gaze targets in VR around the user, along the cardinal axes as well as diagonally in between. The targets were small (1.5°) and placed in spherical coordinates at 5°, 15°, 25°, 35°, 45°, 60°, 80° and 100° visual angle from the centre of the display, at a fixed depth of 2 meters from the viewer. The spread of amplitudes was chosen to have a clear separation of within-view (5°–45°) and beyond-view targets (60°–100°).

3.2.1 Tasks

We constructed two tasks for our user study, one for gaze shifts within-view, and one for gaze-shifts to targets placed beyond-view. In both cases, a trial started with participants aligning their eyes, head and torso straight towards a central target used as a starting point for all gaze shifts. Colour feedback on the central target indicated when the participant had aligned correctly. Trial initiation was complete when participants had been in the correct position for 1.2 seconds. In the first task, the central target then disappeared, and a new within-view target appeared. The participants were instructed to shift their gaze towards the new target as accurately and quickly as possible and dwell on it. The participants were allowed to use their eyes,

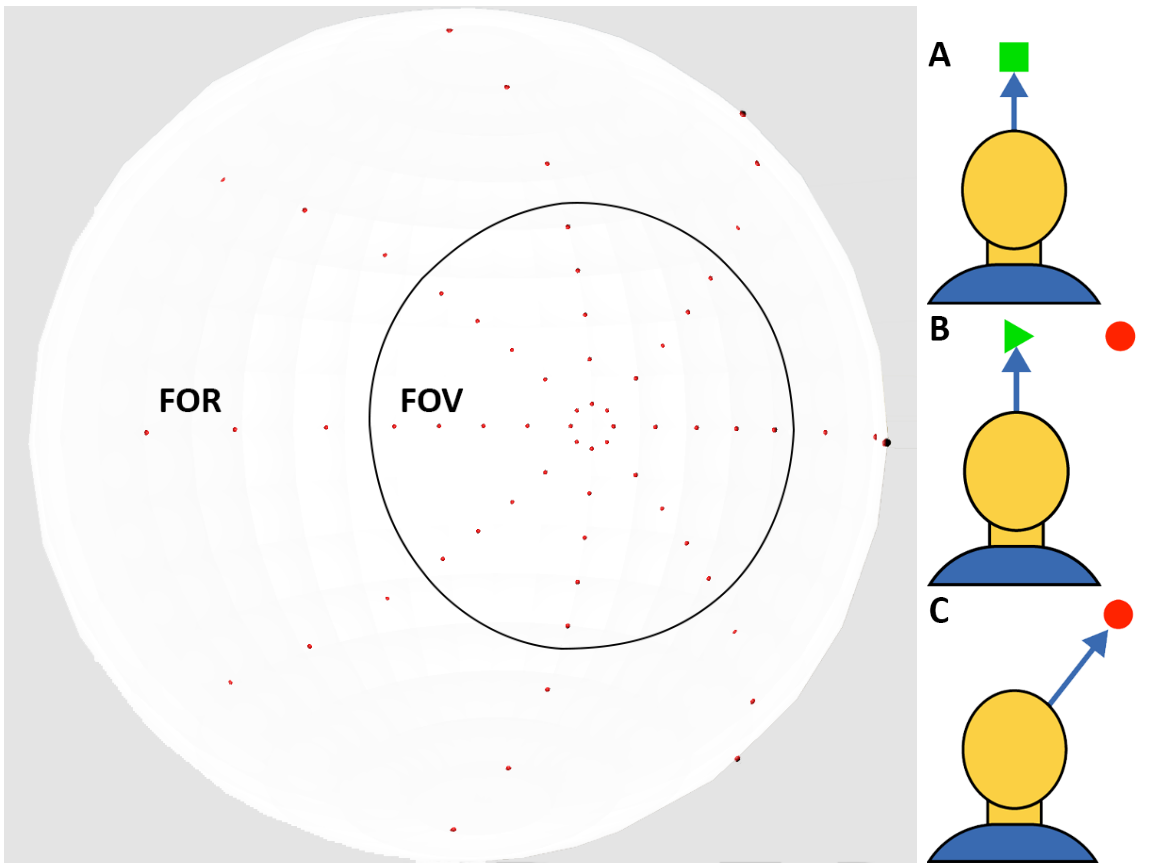

Figure 3.1: Left: Virtual environment used in the study, showing the spherical layout of gaze targets around the user. The black circle represents the participants' FOV at the start of a trial. Right: Participants started each trial by aligning with a square shown in the display centre (A); the square disappears and a gaze target is presented; for targets outside the FOV, an arrow replaces the square pointing in the direction of the target (B); the participant shifts their gaze towards the target and dwells on it until it disappears (C), upon which the next trial is initiated.

head and body at their discretion. The target disappeared when the participant had dwelled within ±2.5° of it for 1.4 seconds, upon which the central target reappeared. The same procedure then repeated for the next target (Figure 3.1). Each within-view target on our spherical grid was presented three times, where every target was presented once in random order before a target could reappear, for a total of 120 trials (5 amplitudes x 8 directions x 3 trials).

The first task is a conventional pointing task, whereas the second task (beyond-view) can be described as *peephole pointing*, as the target is not initially visible and only revealed by moving the display [23]. In our study, this is initiated by the user aligning centrally, as in the first task. After initiation, the central target was replaced by an arrow showing the direction towards a target beyond-view (Figure 3.1). The task was otherwise organised in the same way as first one, for a total of 72 trials (3 amplitudes x 8 directions x 3 trials).

3.2.2 Apparatus

Both tasks used in the study were developed in Unity Version 2017.1.0f3. The position of the head was placed in the zero vector position with the central target placed along the z-axis. An HTC Vive with an integrated Tobii Pro Eye Tracker and data output frequency of 120Hz was used to record eye and head movement. We were able to record data at full frame rate and mean gaze accuracy of 1.533°. Eye and head data were synchronised by the Tobii SDK. The HTC Vive has a FOV of 100° in the horizontal plane, 110° in the vertical plane and a frame rate of 90Hz. A "SparkFun 9DoF Razor IMU M0" attached to the centre of the torso recorded torso movement at 50Hz. Torso data was recorded with eye and head data during run-time in the Unity application. To account for sensor drift by the IMU, the IMU forward direction was reset at the start of each trial (∼ every 5s). Sensor latency adjustment between gaze and head data was provided by the Tobii SDK. Timestamps were marked and participants were asked to perform synchronous gestures with their torso and head at regular intervals to align IMU and head, and gaze data during data analysis. The full setup is shown in Figure 3.2. All sensors were filtered with a 3-point moving average.

3.2.3 Procedure

The experiment took place at Lancaster University, approved by the Lancaster University FST Research Ethics Committee. Twenty participants (14 male, 6 female, ages 26.7 ± 3.6) participated in the study. After giving informed consent and answering a basic demographic pre-study questionnaire, the participant put on the HMD and body tracking device. The participants would always perform both tasks in one posture (sitting or standing). Participants always started with the first task

Figure 3.2: Set-up of the experiment and each posture. Left: The IMU attached to the participant's torso. Middle: The participant is seated on a non-swivel chair in seated posture. Right: The participant stands in a natural position when in the standing posture.

containing within-view targets. Posture order was counterbalanced with a Latin square. The participants sat in a non-swivel chair when in sitting posture, and was asked to stand in a neutral position when in standing posture (Figure 3.2). Participants started each task by conducting an initial eye tracking calibration. The participant would then perform the task while eye, head and torso movement was tracked for each trial. Each task took 5-12 minutes to complete, and the participant was asked to take off the HMD and rest in-between tasks until ready to continue. The whole study took 30-45 minutes to complete for each participant, during which 384 gaze shifts were recorded.

3.2.4 Measures and Data Analysis

For each gaze shift we measured the total gaze movement, and the movement of head, torso, and eye as the contributing movement systems:

- **Gaze movement** during a shift was found by retrieving both eyes' combined directional gaze vector in the VE from the eye tracker.
- **Torso movement** associated with a gaze shift was found by retrieving the directional vector of the torso by calculating the quaternion retrieved from the IMU. Since the tracking of the torso was in a different coordinate space than the VE, the initial direction of the torso was always set to forward in the z-axis at the time of a new target appearing.
- **Head movement** associated with a gaze shift was found by retrieving the directional vector of the HMD in the VE. The head-in-torso movement was found by subtracting the head vector with the torso vector.
- **Eye movement** (eye-in-head rotation) associated with a gaze shift was found by subtracting the gaze vector with the head vector.

For the data analysis, we split all gaze shifts into their respective target amplitude, direction and posture. We divided the data analysis into four main parts:

Temporal Coupling. We analysed how the three movement systems aligned in time by comparing the start times of the eyes, head and torso movements. Only trials with an eventual head movement were used when comparing the coupling between the eyes and the head movement. The eyes were considered used if their velocity exceeded 100°/sec. A head movement was considered used if head velocity exceeded 20°/sec. Similarly, only trials where both head and torso movements were found was used when comparing the coupling between the head and torso. A torso movement was considered used when torso velocity exceeded 10°/sec. All thresholds were found through data testing. The eye threshold was chosen by incrementally increasing the threshold (starting at 10°/sec) until eye tracking noise and other eye movements such as corrective saccades are not detected as a gaze shift [61]. The effect of thresholds was examined through visual inspection of individual trials and changes in the values of dependent variables. The final threshold was also examined to ensure that gaze shifts to the smallest amplitudes (5°) were still detectable. We used a similar approach for head and torso velocity thresholds, where they were incrementally increased (starting at 5°/sec) until sensor noise and small involuntary movements did not impact detection.

System Contribution. We determined the contribution of each system towards a gaze shift by comparing the eye-in-head, head-in-torso and torso directional vectors with the directional gaze vector at the point in time where a gaze shift first reached the target. Statistical analysis was performed on the relative contribution (percentages) of each system's contribution via an *Amplitude* × *Direction* × *Posture* three-way repeated measures ANOVA, Greenhouse-Geisser-corrected in cases where Mauchly's test indicated a violation of sphericity to evaluate the effect of amplitude, posture and direction on each system contribution. Bonferroni correction was used for pair-wise comparisons. An alpha value of 0.05 was used for all tests. The analysis was done on all targets together ($8 \times 8 \times 2$) as well as within-view ($5 \times 8 \times 2$) and beyond-view ($3 \times 8 \times 2$) targets separately.

Relative Total Movement. Head and torso can continue to move toward a target after it has been reached by gaze. We captured the total head movement by retrieving the directional vector of the head at the end of the head movement, and similarly the total torso movement by retrieving the directional vector of the torso. For head movements, the reported values are from the head direction and not from the head-in-torso direction. This allows us to see the total amplitude of the head movement from its initial position independently of whether the torso was used. All analyses are then based on the total movement relative to gaze amplitude, as we are interested in how far head and body are turned toward a gaze target. Statistical analysis was performed via an *Amplitude* × *Direction* × *Posture* three-way repeated measures ANOVA, Greenhouse-Geisser-corrected in cases where Mauchly's test indicated a violation of sphericity to evaluate the effect of amplitude, posture and direction on each system's total movement. Bonferroni correction was used for pair-wise comparisons. An alpha value of 0.05 was used for all tests. The analysis was done on all targets together ($8 \times 8 \times 2$) as well as within-view ($5 \times 8 \times 2$) and beyond-view ($3 \times 8 \times 2$) targets separately.

Preferred Movement Range. Any movement of the systems after first reaching a gaze target serves to reach a preferred eye-in-head and head-on-torso position. We captured preferred ranges using the Customary Ocular Motor Range (COMR) and Customary Head Orientation Range (CHOR) as used in previous studies of eye-head coordination [120, 162, 163, 171, 172].

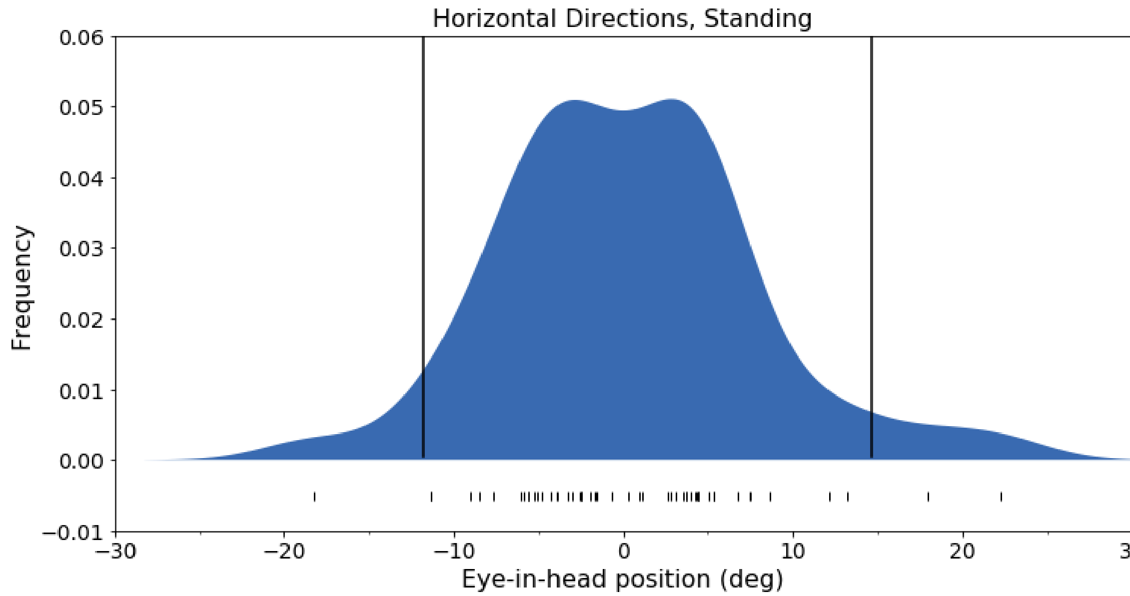

Figure 3.3: The distribution of eye-in-head positions for one participant at the end of the head movement. The centre 90 % consists of the area between the vertical lines.

We used a Gaussian kernel method with a standard deviation of 3° and width of 20° to plot a frequency curve of the relative eye-in-head position at the end of the head movement of a gaze shift to calculate the eye-in-head range (COMR) as in previous work [120, 162, 171]. The curve in (Figure 3.3) is generated by summing multiple Gaussian curves (the kernels), one for each saccade centred over the relative eye-in-head position at the end of the head movement of a gaze shift. If no head movement was registered, then the end of the eye movement was used instead. The preferred eye-in-head range was then defined as the central 90% of the area under a curve plotting the distribution of eye angle relative to the head (Figure 3.3).

The preferred head-in-torso range (CHOR) was calculated by applying the same principle to the frequency of the head position relative to the torso at the end of the torso movement. If no torso movement was registered, then the end of the head movement was used, and if there was no head movement either, then the end of the eye movement was used. Note, that the values calculated represent the whole range in a particular axis, for example, left and right for the horizontal axis.

A disadvantage of the COMR is that it only captures the eye-in-head range after completion of all movements associated with a gaze shift. However, we observed that long shifts to targets beyond the FOV did not conform to the straight ballistic movement to within-view targets, and therefore also analysed eye-in-head range during gaze shifts for beyond-view targets. We defined the preferred eye-in-head position as the mean eye-in-head position between the start of the eye movement until the target was reached. For COMR and CHOR we performed a *Direction* $\times$ *Posture* (8×2) two-way repeated measures ANOVA. However, for the eye-in-head range we performed a *Amplitude* $\times$ *Direction* $\times$ *Posture* $(3 \times 8 \times 2)$ three-way repeated measures ANOVA. We used Greenhouse-Geisser correction in cases where Mauchly's test indicated a violation of sphericity. Bonferroni correction was used for pair-wise comparisons. An alpha value of 0.05 was used for all tests.

3.3 Study results

The 20 participants performed gaze shifts towards each stimulus three times in sitting and standing posture which resulted in a total of 7680 gaze shifts among all participants. 4800 gaze shifts were performed towards within-view targets, and 2880 gaze shifts towards beyond-view targets. 87 gaze shifts (1.1%) were removed from the data analysis due to failed tracking or the participant not finding the target. All measures showed that gaze behaviour was symmetrical in horizontal directions, both upwards diagonal directions and both downwards diagonal directions. However, measure significant differences were found between the vertical directions. Therefore, the results presented in this section have been categorised into upwards, downwards, horizontal, upwards diagonal and downward diagonal gaze shifts for readability.

3.3.1 Temporal coupling and patterns of movement

For each gaze shift, we analysed the start- and end-time for associated eye, head and torso movement, to identify general patterns of coordination and temporal relationships. The reaction time before any movement was registered was on average 200ms for within-view targets, and 375ms for beyond-view targets. The additional reaction time for beyond-view targets is explained by participants needing to process the directional arrow.

3.3.1.1 Patterns of movement

Figure 3.4 illustrates the different gaze shifts that participants performed during the study. The choice of movements during the gaze shift was dependent on multiple factors; amplitude, direction, posture as well as individual differences. The eyes were generally first to move toward a target, followed by optional head and torso movement. Head movement generally preceded any torso movement.

Smaller gaze shifts at amplitudes of 15° or less were generally performed by the eyes only, without any significant head or torso movement. This corresponds with

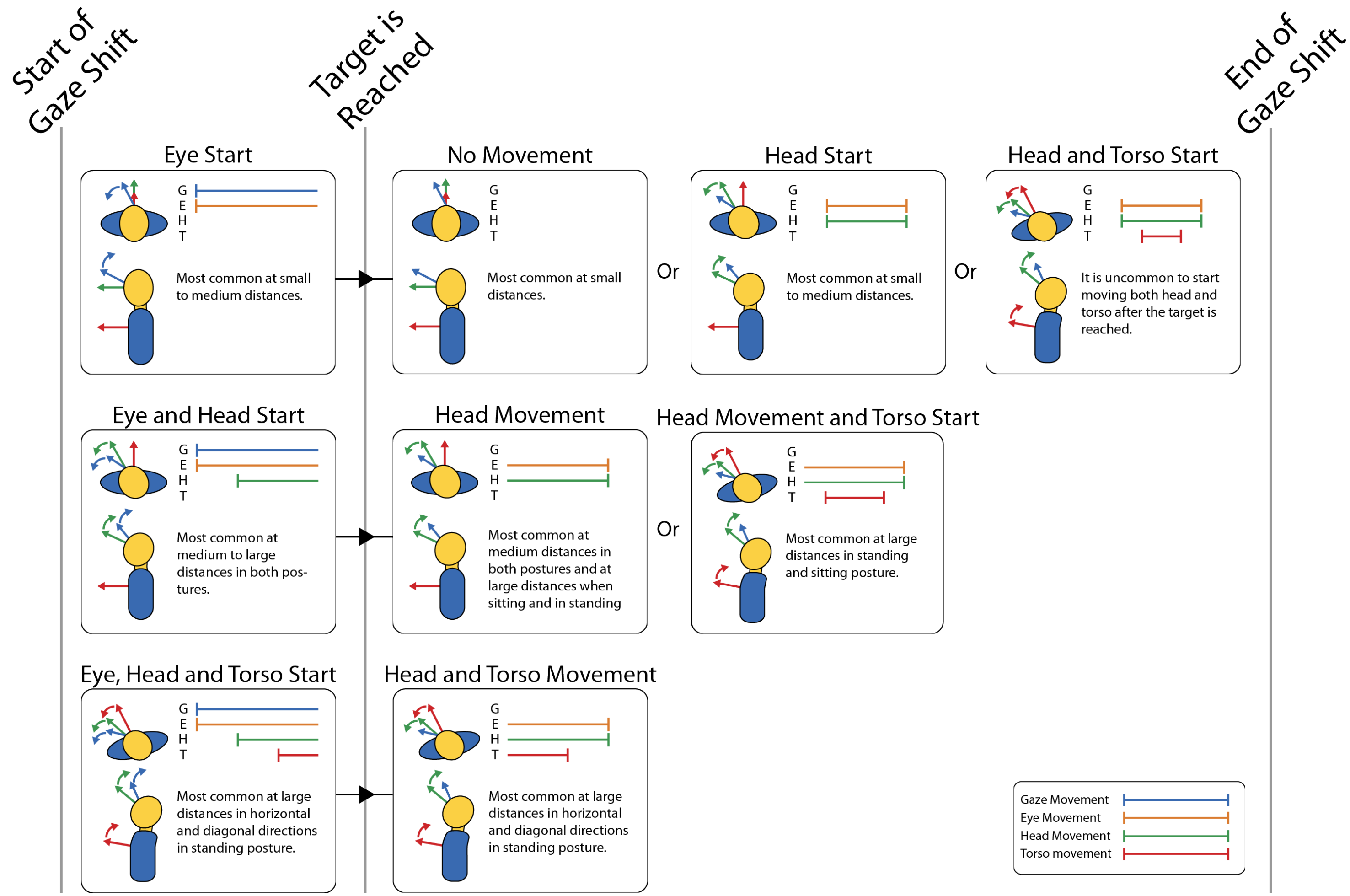

Figure 3.4: Patterns of eye, head and torso movement associated with gaze shifts showing possible movements performed during different stages of a gaze shift. Gaze shifts start with eye movement. Head and torso can contribute toward reaching a target when the gaze shift is longer, and start or continue move after a target was reached. Any eye movement after a target was reached is in the opposite direction to stabilise the eye on the target.

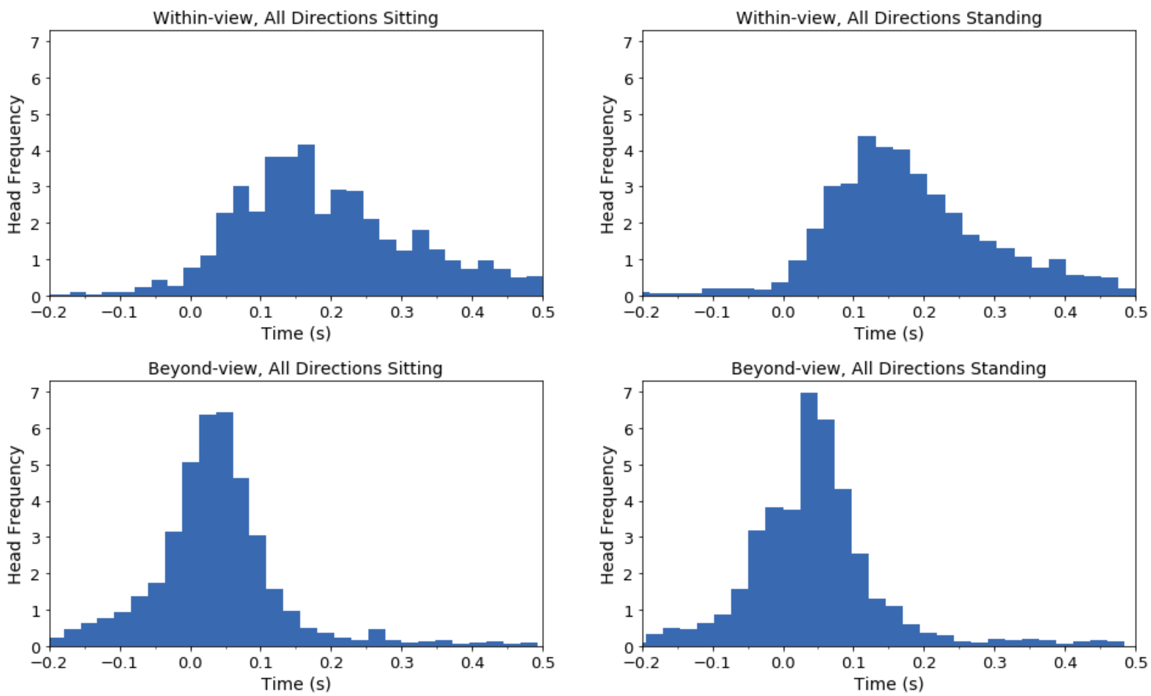

Figure 3.5: Distribution of head movement start-times relative to the start of the eye movement, shown for the different combinations of posture (left: sitting; right: standing) and target visibility (top: within-view; bottom: beyond-view).

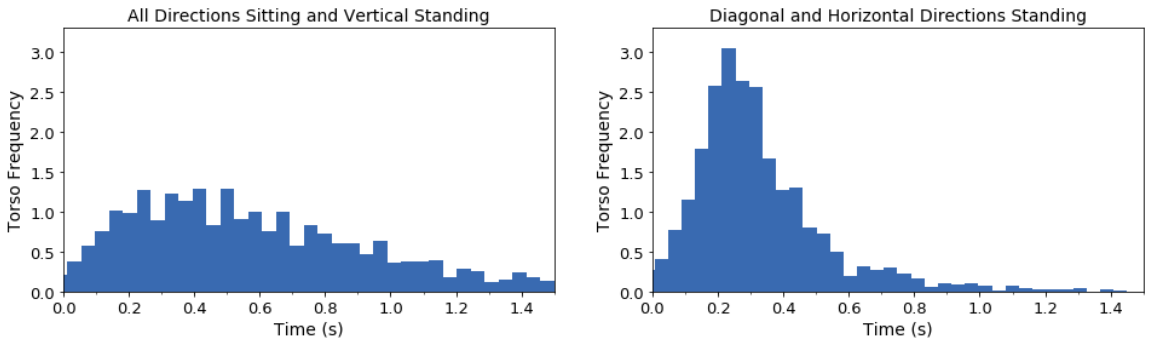

Figure 3.6: Distribution of eventual torso movement start-times relative to start of the head movement. The torso is used more and follows the head with less delay when users stand and the gaze shift involves a horizontal component.

observations reported for gaze in the real world [45, 52, 88]. Head movement that supported larger gaze shifts generally started before the gaze target was reached, whereas any torso movement was further delayed, and frequently only occurred after a gaze target was first reached.

Both the head and torso would often continue to move or start moving after the target had been reached by the eyes, to which the eyes responded with stabilising VOR movement in the opposite direction. These movements were made by the participants to reach an eye-in-head and head-in-torso position that is more comfortable over an extended period. We observed these movement patterns across all amplitudes and postures in our VR study, and they correspond with gaze behaviour observed in real-world studies [144, 181].

3.3.1.2 Lag between eye, head and torso

For any gaze shifts supported by head movement, we observed that the lag between eye and head differed depending on visibility of the target. For within-view targets, participants started a head movement on average at 150ms after the start of the eye movement, whereas the difference was only 30ms on average for beyond-view targets (Figure 3.5). These results seem to confirm previous research suggesting that there is an earlier head movement when the gaze shift is the result of top-down goal-driven behaviour, as is the case for beyond-view targets, as opposed to a bottom-up reaction to a target appearing, which may be a factor for within-view targets [34].

The torso would rarely move independently and would mainly start after the head movement. Torso movement was generally not used in conjunction with the gaze shift for amplitudes under 45°. Additionally, torso movement was rarely used in vertical

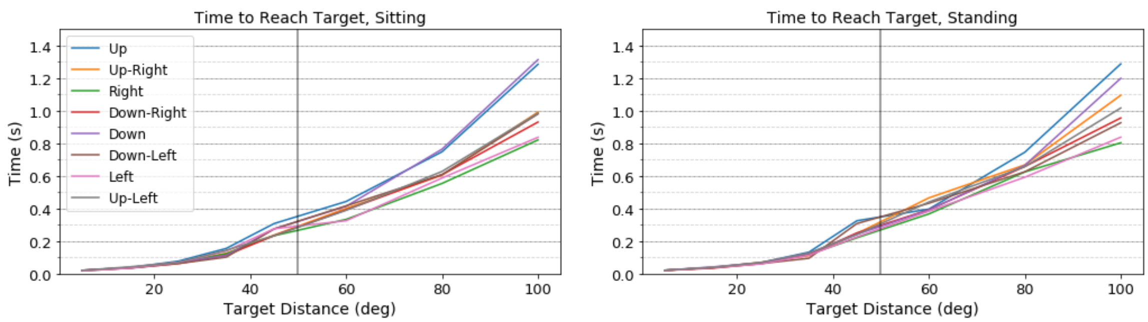

Figure 3.7: Median time to reach the target as a function of amplitude for each direction and posture, excluding reaction time. The vertical line denotes the boundary between within-view and beyond-view gaze shifts. Gaze shifts in the horizontal direction are fastest, and vertical gaze shifts are slowest.

directions or in sitting posture. If a torso movement was used, then it was often used in a later period during the gaze shift, on average 550ms after the head movement. However, torso movement was found for 85% of all horizontal and diagonal shifts in the standing posture at amplitudes of 45° or larger, and in these cases followed the head more quickly, with a lag of only 300ms on average (Figure 3.6).

3.3.1.3 Time to Reach Target

Figure 3.7 shows the median time to reach the target. For small amplitudes, the reach time is quick since only the fast eyes are needed to reach the target. However, the reach time increases for amplitudes larger than 35° due to that the slower head needs to be used to be able to reach the target. We observed no effect of targets being within-view or beyond-view. Once a gaze shift had been initiated, there was no marked slowing down in reaching targets just beyond view, compared to just within view.

We also see a pattern in both postures where the vertical gaze shifts require a longer time to reach the target compared to diagonal and horizontal gaze shifts. Horizontal gaze shifts were shown to be the fastest. No large differences were found when comparing between postures. Also, note that differences between directions mainly appear at the largest amplitudes (80° and 100°).

3.3.1.4 Single Trial Examples

Figure 3.8 shows single-trial examples of typical gaze shifts performed by participants during the study. The top row shows gaze to a target within-view at a amplitude of

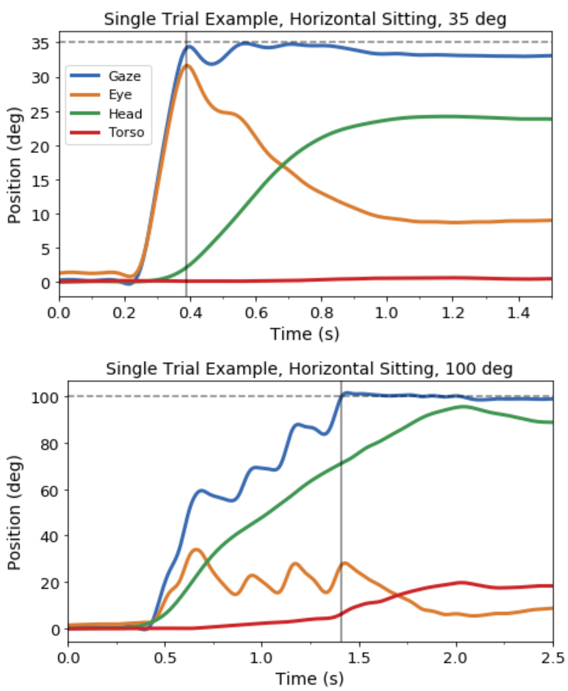

Figure 3.8: Single-trial examples of gaze shifts showing the movement of each system as a function of time. Eye movement is shown relative to the head; all other movements are shown relative to the environment. Note the ballistic movement to targets in closer range versus stepped movement to targets over a larger amplitude, in both seated and standing posture.

35°. The target is reached in one ballistic movement, largely based on eye movement with some support by the head. Upon reaching the target, head and eye movement become opposing and the gaze is stabilised. The pattern is equivalent for sitting versus standing posture.

The bottom row in Figure 3.8 shows a typical example of a larger gaze shift, to a target presented beyond-view at a amplitude of 100°, exposing a strikingly different movement pattern. Large gaze shifts would commonly start with the eyes and head moving towards the target. However, as the eyes are much faster than the head, the eyes would move and then wait for the head to catch up before making another movement. The eyes would in general stay closer to the head compared to what we initially expected. We thought that the eyes would continuously stay ahead of the head at larger eccentricity, gazing toward the edge of the HMD screen where the target is expected to appear, whereas what we found is a distinct stepped behaviour of less eccentric eye shifts and VOR eye movements until the gaze target is reached. This pattern started to appear at amplitudes of 45° and became more prevalent when the amplitude increases. We observed this behaviour in both sitting and standing posture, for all gaze directions.

The single-trial examples also illustrate the absence of torso movement for smaller gaze shifts versus their contribution to larger shifts. Note that the head position shown in these plots represents the combination of torso orientation and head-on-torso relative rotation.

3.3.2 System contributions towards reaching the gaze target

Our next analysis examined the relative contribution of eye, head and torso movement toward *first* reaching the gaze target, i.e. which proportion of the amplitude each of the systems covered. Figure 3.9 provides a complete overview of the system contributions, depending on amplitude, and shown for the different directions and postures.

3.3.2.1 Eye Contribution

Amplitude had a main effect on eye contribution, but direction and posture also had an influence. The eyes perform the significant majority ($> 90\%$) of the gaze shift towards the target for amplitudes of 25° or less, with the head and torso contributing a minimal amount. These amplitudes are easily reached by the eyes alone, and these results were seen in both postures and all directions. However, the eyes contribute less towards the gaze shift as the amplitudes increase beyond the preferred eye range, down to 30-35% for targets at a 100° amplitude.

In general, there was no interaction between amplitude and posture. However, the results showed a significant interaction between amplitude and direction. The

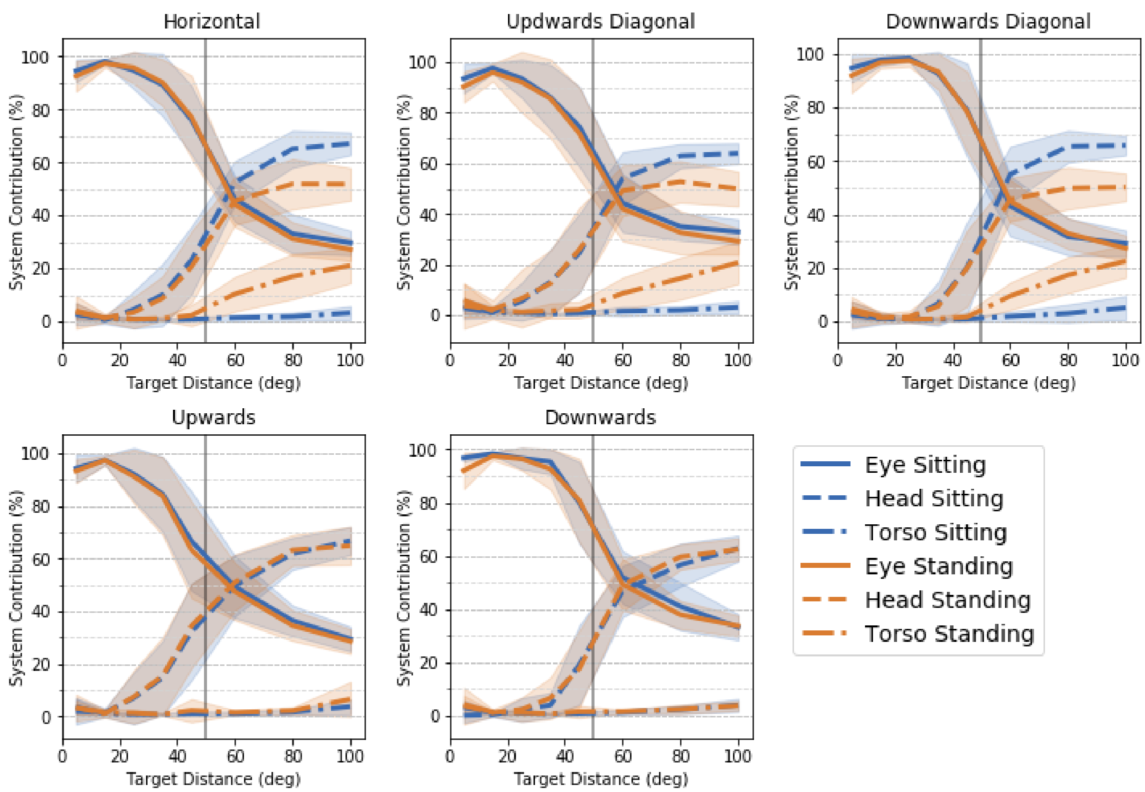

Figure 3.9: The average contribution from eye, head and torso toward first reaching the target, depending on target amplitude, and shown different directions and postures. Shaded areas indicate one standard deviation from the mean.
The vertical line indicates the edge of the participants' initial FOV.

direction had a significant effect on eye contribution for gaze shifts of 25° and larger, in both postures. Participants would, in general, have a larger eye contribution for downwards shifts compared to the other directions. The results showed that eye contribution was symmetrical within the horizontal, upwards diagonal and downwards diagonal directions respectively but this was not the case for vertical directions, where eye contribution were larger in downwards direction than upwards. This can be explained by many tasks naturally require us to gaze downwards at our hands and the limited visual FOV upwards.

Posture did not have any significant effect on eye contribution and no significant interaction with either amplitude or direction. For complete statistical analysis (interactions and main effects of repeated measure ANOVAs on eye contribution), see Appendix Section A.1.

3.3.2.2 Head Contribution

Unsurprisingly, amplitude had a significant effect on head contribution. The head provided a minimal contribution for small amplitudes in both postures and all directions ($< 10\%$). However, the head contribution became more substantial as the target amplitude increased, reaching up to 60% contribution at 100°. The results also showed a significant interaction between amplitude and direction for both postures and both within-view and beyond-view amplitudes. In sitting posture, participants tended to use their head less downwards for within-view amplitudes compared to the other directions. In standing posture, the head contributed more in vertical directions due to less torso movement. Head contribution was symmetrical within each posture when comparing the two horizontal, upwards diagonal and downwards diagonal directions respectively. However, the head would contribute slightly less in downwards direction compared to the downwards direction. Posture and amplitude had a significant interaction for beyond-view amplitudes in horizontal and both diagonal directions due to the added torso movement in standing posture, where the difference in head contribution between the postures could be up to 20% of the gaze shift. Posture showed no significant effect in vertical directions. Posture and direction showed to have a significant interaction for amplitudes at 80° and 100° due to the added torso movement. For complete statistical analysis (interactions and main effects of repeated measure ANOVAs on head contribution), see Appendix Section A.2.

3.3.2.3 Torso Contribution

Torso contribution was in general only prevalent in standing posture in horizontal and diagonal directions for beyond-view amplitudes where it would reach up to 20%. Amplitude and posture only had a significant interaction for horizontal, and both diagonal directions as the torso was rarely used for vertical directions. No significant interaction was found for within-view amplitudes as participants rarely used their torso at such small amplitudes. Amplitude and direction only had significant interactions in standing posture. However, amplitude showed to have a main effect on torso contribution in all directions and postures. Posture and direction had a significant interaction at beyond-view amplitudes. Additionally, the direction had a significant effect on torso contribution in standing posture for beyond-view amplitudes. Posture had a significant effect on horizontal and both diagonal directions for amplitudes at 45° and above. Just as with the eyes and head, torso contribution was symmetrical horizontally and diagonally. For complete statistical analysis (interactions and main effects of repeated measure ANOVAs on torso contribution), see Appendix Section A.3.

3.3.3 Total movement of systems

As illustrated by the single-trial examples shown above (Figure 3.8), much of the head and torso movement associated with gaze shifts occurs after a target is first reached by the eyes. Our next analysis thus considered the total movement of head and torso toward the gaze target. The results are summarised in Figure 3.10, showing how far head and torso turned toward a target as a percentage of the gaze amplitude.

3.3.3.1 Total Head Movement

The results show that the relative total head movement was larger than the head and torso contribution together, as participants generally continued to move their head after the target was reached in order to obtain a more comfortable eye-in-head position. We observed a high variance in the amount of head movement from participant to participant, with differences of up 30° in how much they oriented their head toward a target at the end of gaze shift. This indicates that different tendencies in using head movement, previously observed in real-world studies, also show in VR.

We further observed that the total head movement was generally short of the full gaze amplitude. While participants often continued to rotate their head toward a target after first reaching it with their eyes, their head did not follow their eyes all the way. This behaviour was found in both postures and all directions. It can be explained by the relative effort required for head versus eye movement, and relative saving of energy by only moving the head as far as reaching a comfortable as opposed to central eye-in-head position.

The total head movement was symmetrical in horizontal, upwards diagonal and downwards diagonal directions but participants tended to move their head further upwards compared to downwards. Statistical analysis showed that not only amplitude but also direction and posture affected total head movement.

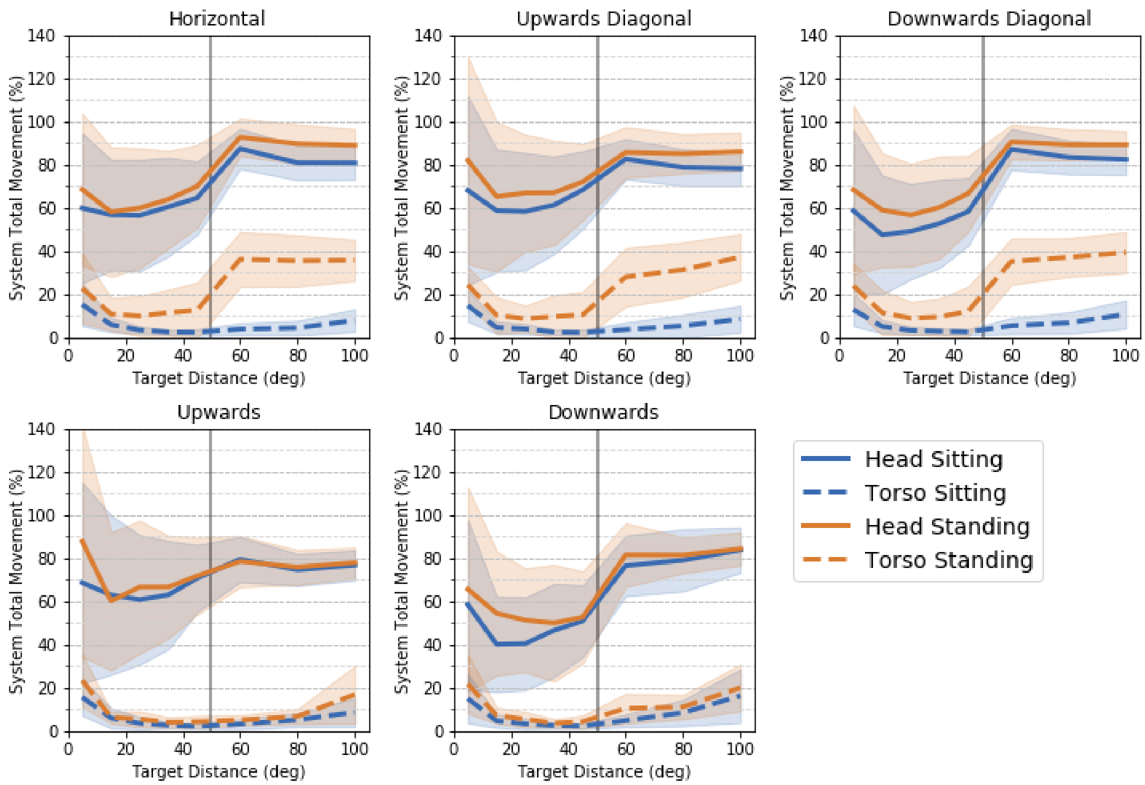

Figure 3.10: The average total movement of the head and torso toward a gaze target as a percentage of the target amplitude, depending on target amplitude, and shown for different directions and postures. Shaded areas indicate one standard deviation from the mean.
The vertical line indicates the edge of the participants' initial FOV.

Over shorter amplitudes within-view, head movement is not necessarily needed to achieve a gaze shift, but given the short range it can be performed in any direction without strain. The amount of head movement can be highly varied, and we observed relative head movement from none to almost 90% of the gaze amplitude. However for beyond-view, head movement was essential to bring the target into the FOV and we therefore saw less variation in the amount of relative head movement. Direction had a main effect in both postures for beyond-view amplitudes where more head movement was generally made in horizontal and diagonal directions. Posture only had a significant main effect in horizontal and both diagonal directions where the extra torso movement would help the head move further. Amplitude only showed a main effect for movements with horizontal component in the seated posture, where the relative total head movement peaks at 60° and then decreases as the torso is not able to help the head to travel further. For complete statistical analysis (interactions and main effects of repeated measure ANOVAs on total head movement), see Appendix Section B.1.

3.3.3.2 Total Torso Movement

Just as with the head, we saw torso movement continued or started after a target was reached, but to a much smaller extent. Participants would generally only use the torso for beyond-view amplitudes, in horizontal and diagonal directions in standing posture during which the torso would turn horizontally. The torso was rarely used in vertical directions or sitting posture. For these conditions, torso movement was generally only seen at the largest amplitudes. The results indicate that torso movement may not be necessary to reach a position of comfort toward targets in the range we studied. However, participants would still support any non-vertical head movement with their torso when they were not constrained by sitting.

Statistical analysis again showed that amplitude, posture and direction had a significant effect on total torso movement. Large differences in total torso movement were found for beyond-view amplitudes. In this range, a significant interaction was found between amplitude, posture and direction. Also, significant interactions were found between amplitude and direction in both postures as well as amplitude and posture in horizontal directions. Additionally, amplitude had a main effect on total torso movement regarding percentages in both postures and all directions except for horizontal directions in standing posture. Posture and direction showed to have significant interaction and main effects for all beyond-view amplitudes. For complete statistical analysis (interactions and main effects of repeated measure ANOVAs on total torso movement), see Appendix Section B.1. Note that the full analysis shows significant effects also for within-view gaze shifts. However the total torso movement in that range was minimal and observed effects pertain to small differences only.

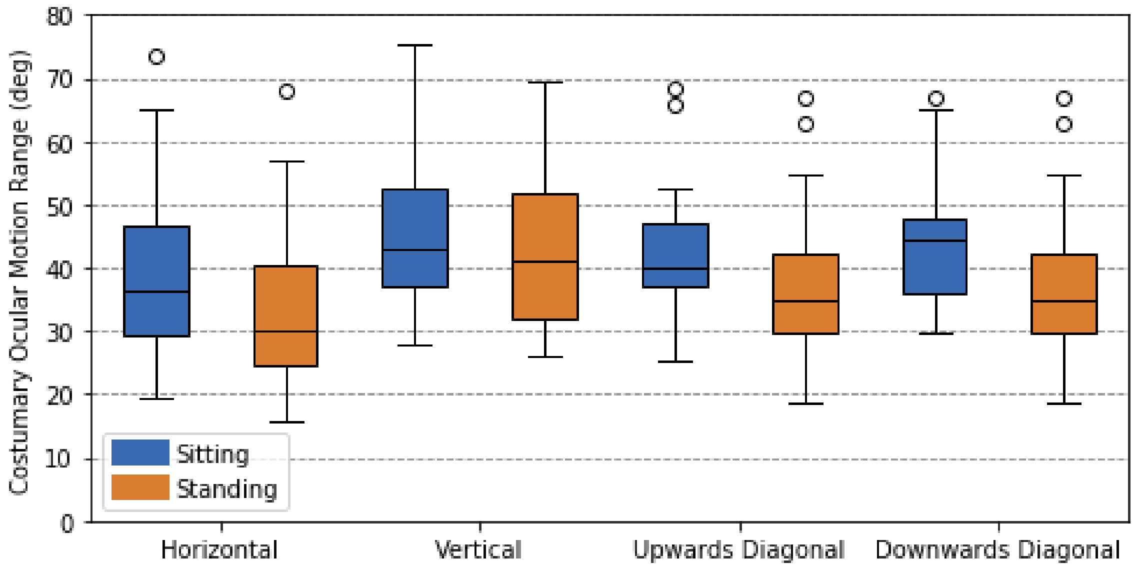

Figure 3.11: The Customary Ocular Motion Range among all participants, with the median value marked.

3.3.4 Preferred Motion Ranges

The next set of analyses was on the range of eye and head motion used, to gain insight into preferred ranges.

3.3.4.1 Customary Ocular Motion Range

As shown above, the head continues to move toward a gaze target after it has been reached by the eyes, so to reach a preferred eye-in-head position. Figure 3.11 illustrates the range of final eye-in-head positions observed in the study. The eyes' motion range varied from 20°-70° among the participants for all postures and directions showing large individual differences. In range and variance, the results are comparable with findings in real-world studies of eye-head coordination [47, 120, 162, 171, 172]. The results show that the preferred range is significantly smaller than the physiologically possible range of eye movement (100°), and also significantly smaller than the FOV of contemporary HMDs.

Statistical analysis showed that the COMR was significantly higher in sitting posture for horizontal ($F_{1,19}$=9.23, p<.05), downwards diagonal ($F_{1,19}$=9.78, p<.05),and upwards diagonal ($F_{1,19}$=7.56, p<.05) directions compared to standing posture. These results align with our observation that participants would move their heads further in standing posture than when seated, for directions with a horizontal component.

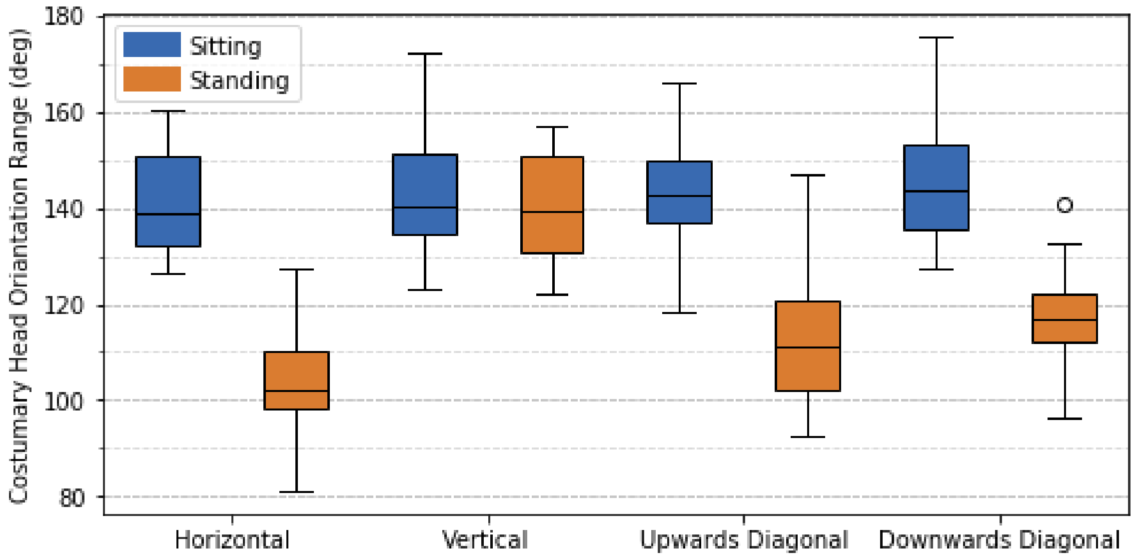

Figure 3.12: The Customary Head Orientation Range among all participants, with the median value marked.

Direction had a significant effect on COMR in both sitting ($F_{3,57}$=7.00, p<.001) and standing posture ($F_{1.80,34.28}$=18.93, p<.001) where vertical directions had a significantly larger COMR compared to horizontal directions. The larger COMR can be explained by the more limited range of the head vertically versus horizontally [44], and lack of support by the torso for vertical head orientation shifts.

3.3.4.2 Customary Head Orientation Range

As shown in Figure 3.12, the preferred head-in-torso range was between 130°-180° in sitting posture and 80-160° in standing posture, showing large individual differences. We observed a significant interaction between posture and direction on CHOR ($F_{1.81,34.36}$=27.32, p<.001). CHOR was significantly higher in sitting posture for horizontal ($F_{1,19}$=170.03, p<.001) downwards diagonal ($F_{1,19}$=85.48, p<.001) and upwards diagonal directions ($F_{1,19}$=96.37, p<.001) compared to sitting posture. No significant difference was found between postures in vertical directions as participants tended to not use their torso. Direction showed a significant effect in standing posture ($F_{2.15,40.93}$=40.32, p<.001) where vertical directions had a significantly higher CHOR compared to the other directions, and the diagonal directions had a significantly higher CHOR compared to horizontal directions. No significant differences were found in the sitting posture as there was minimal torso movement in all directions.

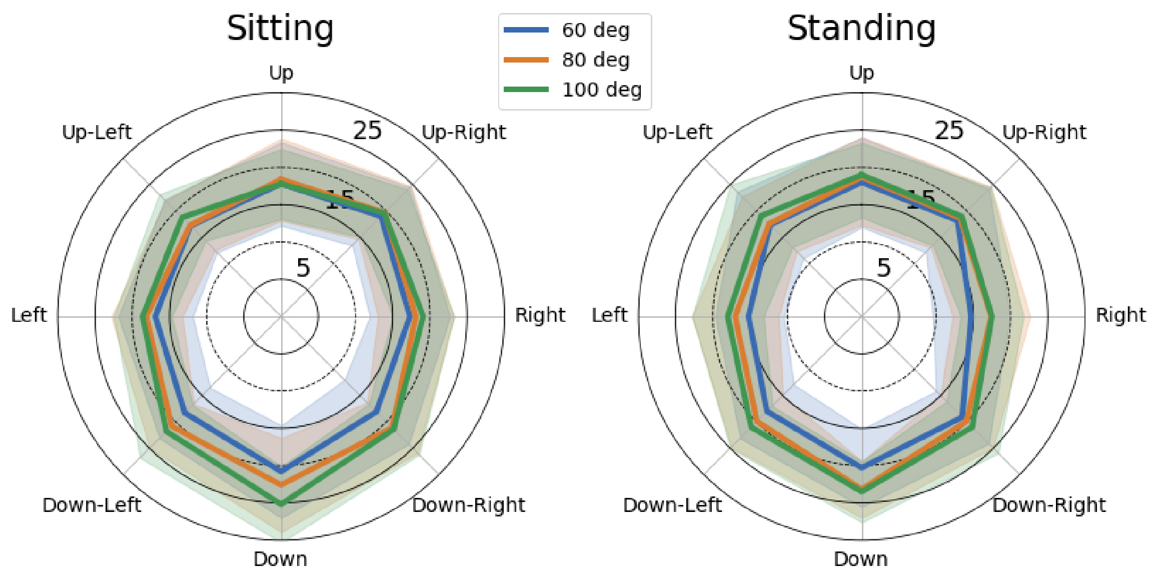

Figure 3.13: Average eye-in-head position from start of a gaze shift to reaching the target, as a function of target direction and amplitude, shown for larger amplitudes and sitting versus standing posture.

3.3.4.3 Preferred Eye Range During Gaze Shifts

The COMR analysis above is based on eye-in-head position upon completion of a gaze shift. We additionally analysed the range within which the eyes moved relative to the head during the movement to a gaze target. Figure 3.13 shows the average preferred eye-in-head positions from the start of the gaze shift to the first hit on the target, for shifts towards beyond-view targets (60°, 80° and 100°). As shown, the range is not symmetric but larger in downward direction. We observed large individual differences where the standard deviation was around was around 5°.

Statistical analysis showed a significant interaction between amplitude, posture and direction on eye-in-head range during gaze shifts ($F_{10.44,553.12}$=1.93, p<.05). Amplitude and direction had significant interaction in sitting posture ($F_{10.17,549.23}$=3.26, p<.001), and in standing posture ($F_{10.31,587.88}$=2.81, p<.05). The results showed that an increase in amplitude would lead to a significantly larger eye-in-head range for participants, except for upwards and upwards diagonal directions in both postures. The direction also had a significant effect, with a larger eye-in-head range downwards compared to other directions, coinciding with our observation of a larger eye contribution downwards compared to other directions. Participants also tended to have a larger eye-in-head range for sitting posture compared to the standing posture.

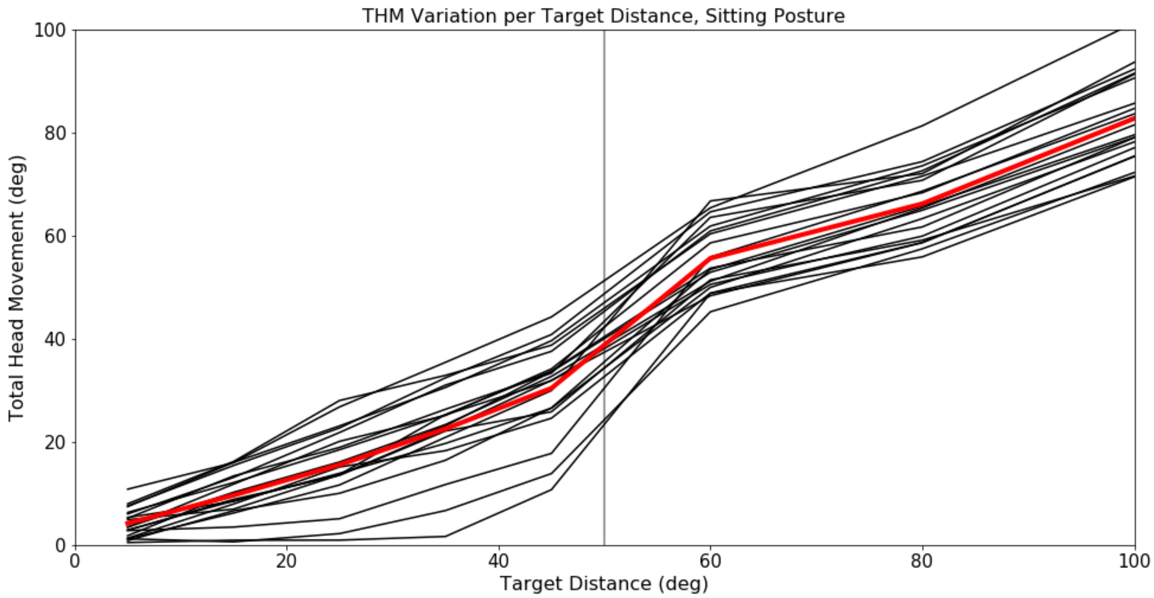

Figure 3.14: Total Head Movement (THM) for horizontal gaze shifts in sitting posture. The red line indicates the mean THM across all participants, while each black line represents the mean THM for a single participant. The vertical line indicates the edge of the participants' initial FOV. Note that some participants displayed distinctly less head movement for targets within-view.

A possible explanation is that the standing posture affords more freedom of head and torso movement to the effect that participants have less need to "stretch" their eyes.

3.3.5 Participant Variability

A common theme for all results was the high variability. Gaze shifts made by the same participant showed consistent behaviour, but gaze shifts between participants showed high variability. The results showed that this was primarily due to differences in the amount of head movement used.

Figure 3.14 shows the difference between participants in total head movement in the seated condition. The majority of participants used their head to support gaze at all target amplitudes, but a number of participants tended to do so only for larger target amplitudes where head movement became unavoidable. For sitting, horizontal gaze shifts of 35°, the total head movement had a mean of 21.19° (s.d. 8.02°) but there were clear outliers. Participant P20 reached all targets at this amplitude without any head movement, and P1 exhibited only little head movement (5.60°, s.d. 1.89°). We found corresponding results for the COMR measure of eye movement range. The mean COMR was 39.56° (s.d. 14.33°) but P20 and P1 had much larger COMR values (P20: 73.47° and P1: 65.21°), highlighting their more extensive use of eyes.

The observation that some participants would use their head frequently, while others would not move their head unless it was needed for reaching a gaze target, aligns with findings in fundamental neuroscience research [47, 120, 162, 171, 172]. However, it is not clear whether there are distinct groups of "head-movers" versus "non-head-movers" as proposed in past work [47], or whether there is a continuum in individual tendency to support gaze with head movement. Furthermore, it is also unclear whether these differences are specific to the task (i.e. gaze shifts back and forth from a central position) and participants task approach, or indicative of their inherent head movement tendencies.

3.4 Discussion

The study results provide in-depth insight into eye, head and torso coordination of gaze shifts. Here, we discuss the results with regards the factors of interest we had identified, starting with a general comparison of our findings in VR with findings of studies in real environments. Note that we cover design implications and limitations separately, following the discussion.

3.4.1 Virtual Reality vs. Real World

Our study indicates that gaze shifts are generally performed in the same way in VR as in the real world. Just as in the real world, smaller gaze shifts up to 25° were mostly performed with the eyes while the head and torso contributed more when gaze shifts were larger [45, 52, 88]. Our results on eye-in-head motion are also similar to customary ocular motion ranges observed in the real world (COMR means in the range 45–55°, with standard deviations 7–24° [120, 162, 163, 171, 172]), and we found comparable variability in head movement tendency as previously observed in real world gaze [47]. The correspondence of our results with prior observations is of importance as it implies that general knowledge of visual attention and gaze can be transferred to head-mounted VR. This was not necessarily expected, given the limited FOV of HMDs and additional weight they impose on head movement. However, our study did not directly compare virtual and physical reality, and a recent direct comparison observed that a higher proportion of gaze shifts were supported by head motion when users were in VR [125].

Our study had more extensive coverage than prior studies of eye-head coordination, which primarily focussed on horizontal gaze shifts, in seated position, and frequently over shorter ranges. Given the similarities observed for horizontal gaze shifts, we expect that observations on gaze shifts in other directions and in standing versus seated posture are generalisable from our study in VR to gaze in the real world. Note, the head orientation ranges (CHOR) we observed were considerably higher than reported by prior studies but this is explained by the wider range of gaze shifts we covered (±100° compared to, e.g., ±45° in [163]).

We observed a distinct pattern of stepped gaze shifts where the eyes performed a series of saccadic shifts toward the target interleaved with VOR movement to let the head catch up (Figure 3.8). We have not found any detailed description of this behaviour in prior literature, although Land noted that large gaze shifts can occur in steps [89]. It is reasonable to assume that people perform gaze shifts similarly in the real world, to keep their eyes within a preferred range during transition to a gaze target. However, the question arises whether the limited FOV of the HMD affects how frequently the eyes wait for the head. The further the eyes rotate from the head, the more of the viewer's peripheral vision lies beyond the displays boundary without stimulation by the virtual scene, and this might trigger the eyes to wait. It might also be that gaze is stepped so to stay within a visual range of the display that is reachable by both eyes. This could also explain the amplitude of the stepped behaviour (10-15° from the head direction) as the stepped behaviour would occur far within the range of the eyes and the HMD limits (∼50°).

3.4.2 Amplitudes and Gaze Zones

Amplitude has a critical effect on eye, head and torso coordination. This effect is natural as both the eyes and head have a limited range of motion and therefore need to coordinate, also with the torso, to achieve gaze shifts over larger amplitudes. However, we saw that head and torso supported gaze not only to extend the reach of the eyes, but also to stabilise gaze in a comfortable position. This effect becomes more apparent for larger gaze shifts. Based on our results, we propose three gaze zones: gaze shifts up to 25° that can be comfortably achieved eyes-only; gaze shifts in the range 25–50° where eyes and head together reach targets comfortably; and gaze shifts over amplitudes larger 50° where eyes, head and also torso need to work together for comfortable viewing. The larger the amplitude, the more apparent are also differences between directions and postures.

3.4.3 Within-view vs. Beyond-view

Visibility effects overlap with amplitude effects, as the targets were within-view in our study when they were within the first two gaze zones proposed above, and beyond-view when they were in the third gaze zone. However, there was a distinct effect on the relative starting times of head and eye movement, with the head movement more quickly initiated when targets were beyond-view. The effect is explained by significant differences in the two tasks we used. Within-view targets required the user to perform a *reactive* gaze shift, in response to a stimulus appearing in their peripheral vision. Beyond-view targets, in contrast, required the user to process a directional cue and perform a *predictive* gaze shift toward the direction in which the target is expected. Prior work in neuroscience suggests that reactive gaze is led by the eyes followed by the head, whereas predictive gaze involves earlier head movement [34]. The results in our study appear to support these prior observations.

However, in our study, participants only had knowledge of the target direction but not of how far beyond the view it was positioned, adding to relative complexity of the task. Search for the target may have played a role in the stepped gaze behaviour we observed for beyond-view targets. We saw more frequent interleaving of the saccadic shift with VOR eye movement than we had expected and this could be explained by the eyes intermittently stabilising on the virtual scene to be able to assess whether the target has come into view.

3.4.4 Sitting vs. Standing

The results showed that the head and torso were used significantly more in standing posture compared to sitting posture for large gaze shifts. Standing posture offers more freedom of movement for the torso, and in extension the head, compared to sitting posture and therefore induces a different user behaviour. The head moves further from its initial position in standing posture compared to sitting posture (see total head movement, Figure 3.10), but the head-in-torso range is significantly larger in sitting position where the torso is not able to provide as much support (see CHOR, Figure 3.12). This shows that users prefer to support head movement with torso rotation for head shifts even when these are well within head motion range.

3.4.5 Comparing Directions

The coordination of eye, head and torso is also influenced by direction of gaze shifts. We found no differences between the horizontal directions, but significant differences between the vertical directions. As a consequence, differences were also found between upwards versus downwards diagonal directions. In general, participants would use their eyes more and their head less for downwards shifts compared to upwards shifts, explained by the asymmetric structure of the head where the visual range upwards is more constrained in range.

In the standing posture, participants tended to have different coordination for vertical shifts compared to horizontal and diagonal shifts. These differences are mainly due to the torso not being involved in vertical shifts but in other directions. In the sitting posture, the different directions showed less of an effect, with exception of downwards shifts for which the eyes were used more in comparison with other directions.

3.5 Design Implications

The insights we gained into eye, head and torso coordination lead to a range of implications for VR and interaction design. We reflect implications pertaining to eye, head and torso movement, in each case identifying factors to take into account in design, and suggesting how to reflect them. We also reflect on limitations of our study, and validity of findings.

3.5.1 Eye Movement

The eyes can be used for interaction in many VR design areas. However, certain factors should be taken into account when designing these interactions, namely:

- The eyes contribute more than 90% of the gaze shift for targets at 25° or less amplitude.
- The contribution of the eyes is not affected by the posture.
- The eyes move further downwards relative to the head compared to other directions.
- The eyes prefer to remain within a certain range relative to the head and will wait for the head to catch up in order to remain within that range.
- The eyes' range of motion relative to the head is dependent on the user.

The preferred eye-in-head range during gaze shifts and at the end of gaze shifts (COMR) can be used for the placement of non-diegetic UI elements. The preferred eye-in-head range indicates where to place UI elements in the FOV so that they are not in the area where the user spends most of their time looking. However, while it is possible to place UI elements further out than the average user will naturally look, it may cause long-term strain on the user as they then have to reach eye-in-head positions that they would normally not do. Additionally, placing elements too far out in the FOV may expose limitations of current VR technology at the edges of the screen such as chromatic aberration or reduced eye tracking accuracy.

A common theme in the results was that participants tended to use their eyes more downwards compared to other directions. Therefore, placing objects below the current object-of-interest should be more suitable if the aim is to reduce head movement. For example, when placing subtitles that are attached to the speaker, it may be more suitable to place them below the speaker to decrease the likelihood of unnecessary head movement when users are shifting between the subtitles and speaker. The placement naturally depends on the user's relative position to the speaker, but placing subtitles under the current object-of-interest should leave a larger eye comfort range.

The factors can also be used for the design of future eye tracking applications in VR. Eye tracking in VR offers multiple advantages over common screen-based eye tracking such as the blocking of external light, the eye tracker being closer to user's eyes, the eye tracker being less affected by user's head movement [61]. However, VR applications induce different user behaviour that often requires more body movement compared to interaction with regular desktops, and it is vital to understand the gaze behaviour in VR to transfer relevant design knowledge from desktop applications to VR applications. First implementations of gaze interaction techniques in VR have already been investigated [134, 138, 141, 170] but we expect that better understanding of eye, head and body coordination can inspire new techniques.

3.5.2 Head Movement

In VR, the head position determines the view that is exposed to the user. In existing work, it tends to be assumed that this coincides with where the user looks. However, our results show that head orientation is problematic as an approximation of gaze. There are multiple factors regarding head movement that should be taken into consideration when designing VR experiences:

- Users rarely shift their head fully towards an object to which they have shifted their gaze.
- The amount of head movement towards the target is dependent on the user.
- A significant part of the head movement is performed after the gaze has reached the target, especially for targets reachable by the eyes only.
- The amount of head movement is dependent on the direction. Head movement is symmetrical horizontally, upwards diagonally and downwards diagonally, but not vertically. People use less head movement downwards compared to upwards for within-view amplitudes.
- Standing posture offers more freedom of movement compared to sitting posture and therefore more movement towards the target in horizontal and diagonal directions.

- Users perform head movement earlier, relative to the eyes, if the gaze target is outside their current FOV.

The knowledge of these factors can be used for many fundamental design considerations. For example, the factors can be used to better place upcoming objects of interest. If the approximate position of the FOV at the end of a gaze shift is known, then the next object of interest can be placed within that theoretical FOV to maximise the likelihood of the object being visible upon completion of the gaze shift. This information can also be used ergonomically, where objects can be placed in a manner that decreases unnecessary head movement and reduces strain on the user. For example, if the user is expected to shift back and forth multiple times between objects of interest then it would be advantageous to place the items within 20° of each other where the eyes can do most of the work. On the other hand, if the user is expected to shift once between two objects of interest, then it may be advantageous to have the objects of interest further apart in order to increase the likelihood of head movement so that more of the upcoming objects of interest appear in the FOV.

The large variability among users in head movement tendency is important to reflect in VR design. Head pointing is widely used for interaction in contemporary VR applications [8] but may be unnatural and more straining for users who tend to avoid head movement when gaze targets are attainable by the eyes only. Comparative studies of head versus gaze pointing generally find head pointing preferred [138], but this fails to account for individual differences and growing evidence that a significant proportion of people can be regarded as "non-head movers" [47, 120, 162, 171, 172]. Another aspect to consider is that since the HMD is attached to the head, the amount of head rotation a user performs will, in turn, impact what they will see in the VE. As a consequence, different users may have completely different visual experiences. The knowledge of a user's head movement tendencies can have broad implications for tailoring individual experiences or identifying users [49, 96, 129].

It also important to consider the posture of the user when designing experiences in VR. Whether the user is standing or sitting affects the head movement of the user, which in turn will influence the user's FOV and overall visual experience. Additionally, VR experiences that require much turning may prove strenuous in seated posture, as they induce head movement over larger ranges than users would choose if their torso movement was not constrained.

3.5.3 Torso Movement

Torso movement is used to reach further than is possible by the eyes and head alone as well as to reach more comfortable eye-in-head and head-in-torso positions for the user. Forcing users to overextend their head in order to reach targets may be unsustainable ergonomically and may negatively affect the user experience. Therefore it is important to understand factors that affect torso movement:

- Torso movement is rarely used for amplitudes up to 50°
- Torso movement is highly dependent on posture and direction where torso movement is generally only performed for horizontal and diagonal shifts in standing posture.
- Participants perform torso movement even if it is not necessary to reach the target.
- A significant part of the torso movement is performed after the target has been reached by the gaze.

Knowledge of the torso movement behaviour can be used to roughly approximate the torso's current position based on the position of the head and gaze. In this user study we attached an IMU to the torso to record the current torso movements, however such devices are generally not available in commercial VR devices and it is common to assume that the rest of the body is aligned with the HMD. The findings from our user study could be applied to better estimate torso orientation relative to the virtual environment, as a possible context or input for interaction.

3.5.4 Limitations

Any study design represents choices that may pose limitations on the validity of results. First to consider is the abstract nature of the tasks we chose, in which participants were instructed to shift their gaze to a unique target, in the absence of any distractors. Prior eye-head coordination research, in particular the extensive work of Land and colleagues [88, 89], has shown that fundamental insights gained under such a paradigm generalise well to naturalistic use of gaze. This gives us confidence that the results we discussed hold in principle, but other types of tasks, for example visual search or free-viewing, might expose specific differences.

One of our key observations, of the head typically following the eyes only to within 80–90%, is corroborated by recent works on visual exploration of 360° images and provides an explanation for the head-to-eye fixation offset those work reported [140, 158]. In our study, there was always only one target to reach, after which to return to the display centre. Users might use their head and body more if they have further targets to reach in the same direction. We suspect that this might affect gaze shifts that are otherwise not supported by the head. It is also possible that the total head movement would increase but we consider this less likely given that other work also observed a general offset between head and eye fixations.

Although we used a 3D virtual environment, we chose to present all stimuli at a fixed depth to ensure that our results would not be confounded by accommodation effects. However, it will of course also be interesting to study depth as a factor in gaze shifts. We speculate that there is not much interaction between depth accommodation and eye-head-torso contribution, although it is possible that depth conflicts could induce more central alignment of the head with gaze targets.

With N=20 participants, our study was significantly larger than most prior studies on eye-head coordination, however not of sufficient scale to attain conclusive results on head movement variability. We observed participants who displayed distinctly less head movement than the majority and this may indicate that there are users for whom head-based control may be less natural than generally assumed for interaction in VR and HMDs. The neuroscience literature lends credibility to a possible differentiation of users with more head movement versus users with wider eye movement ranges [47, 171, 172]. However, it may also be that some participants chose to use their head less than they otherwise would as the task required them to return their gaze to the display centre after they dwelled briefly on a target.

Our study was conducted in a head-mounted VR environment. However, the subset of our results for which we found comparable data in real world studies of eye-head coordination shows a strong correspondence. We therefore suggest that most results, such as on proposed gaze zones and preferred motion ranges, will be valid and relevant also beyond VR and HMD-based interaction, especially for settings that involve gaze over wider fields of view (e.g., display walls, multi-device environments, and wearable AR). However, for some of the behaviours we observed, for instance stepped eye movement in large gaze shifts, it is not clear how much they may have depended on the nature of the VR environment with limited FOV.

The specific choice of apparatus also presents potential limitations. Eye, head and body tracking all can involve measurement error but we regard this as negligible in our study, given the use of high-end sensors and careful calibration. The display technology can also influence gaze behaviour. We suspect that the FOV might be a factor in how large gaze shifts are performed, but other potential factors include the weight added to the head, and optical features such as pincushion distortion with lower resolution toward the edges (e.g., Playstation VR HMD). The HMD we used had a FOV that is representative for contemporary devices, but at 100° it implied that we cannot know whether the effects observed for gaze shifts larger 50° from the centre were more due to increased amplitude, or to targets not being visible at the onset.

3.6 Conclusion

The study we reported in this chapter provides detailed insight into eye, head and torso coordination during gaze shifts, thus answering RQ1. We learned that posture and direction in addition to amplitude has an effect on gaze shifts, identified gaze zones reflecting different levels of eye, head and torso contribution, and observed preferred motion ranges and how these vary for different user groups. There are a number of fundamental conclusions we draw from the work:

First of all, gaze is multimodal. That gaze involves not only eye but also head and body movement is not new knowledge as such. However, the HCI field has generally treated gaze as unimodal, most commonly associated with eye movement only. We argue that understanding gaze as multimodal is critical as we move to forms of interaction that expose wider fields of view, including display walls, room-scale interactions, head-mounted displays, virtual environments and mixed realities. We propose gaze zones to guide design of gaze interaction beyond single screens, with each zone relating to a range over which we interact, and movements on which we draw for visual attention.

Secondly, eye, head and body movement are connected. All three movement systems, the eyes, head and body, have been considered separately (and extensively) as input, control or cue for human-computer interaction. Our study shows that they are in fact closely coupled as we shift our gaze and attention. This is fundamental for any visual forms of interaction. We argue that knowledge of the underlying interactions between eyes, head and body holds rich potential for design of novel interactions techniques.

Thirdly, eye and head do not compete but cooperate. Eye and head orientation have been compared and contrasted as computer input in a range of studies over the last 20 years [15, 18, 56, 87, 138]. We argue that this dichotomy is unhelpful as eye and head are naturally coupled. Existing ideas of combing eye and head for interaction reinforce their treatment as separate (e.g. with mapping to different steps in an interaction), whereas we suggest to leverage them as integral to take advantage of natural eye-head coordination behaviour.

Finally, there are significant individual differences in head movement associated with gaze. While the majority of users display head movement that complements eye movement during a gaze shift to different extents, the data also suggest there are users who prefer to use their heads only when gaze targets were not reachable by the eyes alone, and in contrast use a wider eye motion range. Evidence of individual head movement differences is highly significant for interaction in VR and HMDs, where head movement is central to navigation and visual experience of the environment.

In the following chapters we will explore how we can leverage the insights on eye and head coordination found in this chapter for novel gaze interaction (RQ2), and to tackle our defined gaze-interaction problems.

Chapter 4

Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection

Our previous chapter showed that there is a deep connection between our eyes and head as we perform gaze shifts in VR environments. These coordinated behaviours can now be leveraged to design novel techniques to expand the capabilities of gaze interaction and tackle the defined gaze interaction challenges (RQ2). In this chapter, we first identify design principles for Eye&Head gaze interaction, grounded in eye-head coordination literature, and then apply these in three novel gaze interaction techniques that integrate head and eye movements to address our defined gaze interaction challenges. All three techniques have been implemented in a head-mounted VR environment, with application examples that demonstrate their advantages. We then evaluate them against gaze pointing and dwell selection baselines.

Head movement has been used to support gaze pointing in a variety of ways. Head gestures such as nodding have been proposed for confirmation of targets users look at [104, 161]. These methods exploit eye-head coordination implicitly as they track the compensatory eye movement during a head gesture, without need for separate head tracking. In extension, head turning has been proposed for scalar input to controls fixated by gaze [119] and 3D target disambiguation [105]. In EyeSeeThrough, head movement controls a toolglass that can be moved over gaze-fixated targets [106]. Other work has supplemented eye pointing with subsequent refinement of the selection by head movement [76, 77, 86, 160]. Recently, Pinpointing compared head versus eyes as primary pointing modes, and a variety of techniques for subsequent selection refinement [87]. The existing body of work has in common that eye and head movement are treated in separation, for use of one after the other. In contrast, this work proposes pointing and selection techniques that build on the *integral* relationship of eye and head movement in gaze interaction.

4.1 Eye&Head Design Principles

We build on the results from the previous chapter and on fundamental insights from neuroscience to derive design principles:

- Larger gaze shifts require head movement. The eyes have a physical range of 50° but rarely rotate beyond 30° relative to a central position in the head [88]. The head is therefore needed to explore further.
- Not all gaze shifts involve head movement. Gaze shifts up to ˜20° can be performed with only eye movement [45, 88].
- The decision and timing to support a gaze shift with head movement is influenced by multiple factors, such as expected duration of maintaining gaze in the vicinity of the new direction, position of the next target, and initial eye-in-head position [47, 120].
- The head will start or continue to move after a target is first reached by the eyes, and the eyes perform compensatory movement based on VOR to maintain focus on the target [17, 45].
- The head will not typically move fully toward a target, as head movement requires more energy while a comfortable eye-in-head position is reached sooner [45].

A key design implication is that we can distinguish two types of gaze shift: *head-supported* gaze shifts where eye movement is accompanied by head movement, and *eyes-only* gaze shifts that are performed without contribution by the head. We identify three design principles that build on this dichotomy:

Head-supported gaze is more stable than eyes-only gaze. The head does not contribute to every gaze shift; it only supports the eyes when the attention shift is more substantial such that it requires or warrants a recentering of the area the eyes can comfortably explore. Consequently, gaze points selected with head support are less volatile than gaze points that change with every eye fixation.

Eyes-only gaze explores around objects selected by head-supported gaze. Objects acquired with head support are significant in that they constitute a base from which other objects in the vicinity are explored eyes-only, with less effort.

Head-eye alignment can signal intent. The head does not normally rotate all the way with the eyes to acquire an object, and an offset remains between head and eye at the end of a gaze shift. Though it may seem counterintuitive, it follows

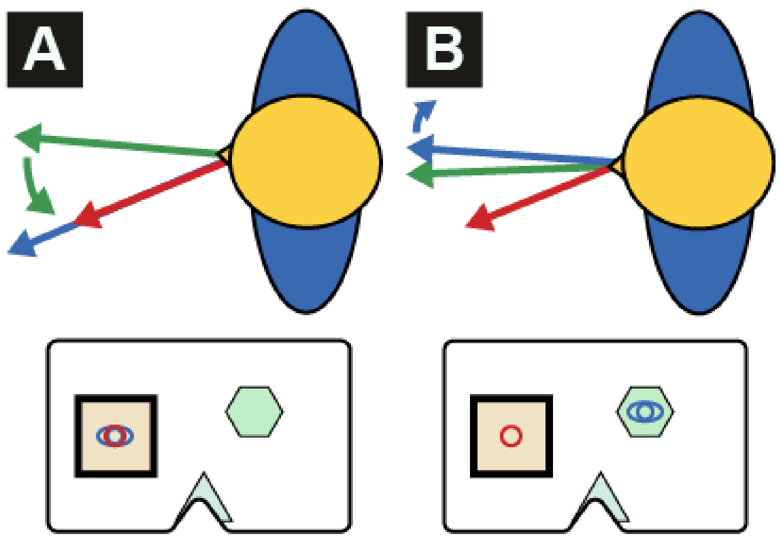

Figure 4.1: Eye&Head Pointing. The pointer moves to a new gaze position only when the gaze shift is head-supported. A: The pointer (red) follows the user's gaze (blue) to the square as the user is also moving their head (green). B: The user shifts their gaze to the hexagon, but as the shift is eyes-only without head movement the pointer remains on the square.

that we practically never look exactly straight ahead. Alignment of head and eye can therefore be available to signal intent.

Note the fundamental limitation of head pointing. As the head does not normally move the full distance to the target, head pointing does not accurately identify objects of interest, even when the gaze shift is supported by head motion.

4.2 Eye&Head Gaze Interaction Techniques

We designed three novel pointing and selection techniques based on the identified principles for Eye&Head gaze. The techniques are gaze-only, based on where the user looks, but combine information from both eye and head in the process.

4.2.1 Eye&Head Pointing

The first technique is for gaze pointing modulated by head movement. As shown in Figure 4.1, the *Eye&Head Pointer* moves to a new gaze position when the gaze shift is accompanied by head motion. When a user moves their gaze without head movement (eyes only), the pointer remains at the last position acquired with head support. Note that head movement only modulates the pointer. The points selected are gaze points,

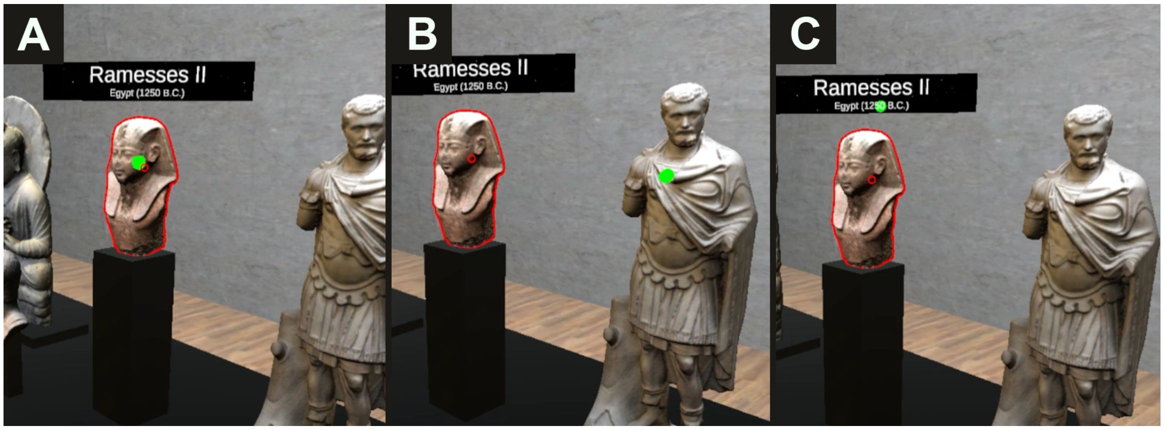

Figure 4.2: Eye&Head Pointing in a virtual museum. The green dot indicates the gaze position. A: An artefact is selected with head-supported gaze, triggering contextual information. B: The selection is maintained when the user looks at nearby artefacts, using eyes-only gaze. C: The context display can be viewed without needing to carefully maintain gaze on the artefact.

and it is not taken into account where the head points relative to a target. The technique can be implemented with a cursor or implicit feedback highlighting objects selected by the pointer. It has the effect that a cursor or object selection is more stable and less distracting than a conventional gaze cursor that follows eye movement continuously.

The Eye&Head Pointer enables users to fluidly couple gazing and pointing. They can move their eyes to look at objects without the pointer following. However as soon as they also move their head, the pointer will jump to where they are looking. The coupling can be implicit and entirely based on naturally eye-head coordination, but users can also choose to move their head to have the pointer follow them to a gaze target they might otherwise have attained eyes-only.

We developed a virtual museum application to illustrate Eye&Head pointing for exploration of artefacts by gaze (Figure 4.2). Head-supported gaze shifts trigger a contextual display over the artefact, in the same way as a mouse hover might in a desktop interface. Eyes-only gaze can be used to view other artefacts while the selection is maintained. The hover selection only changes once the user shifts their gaze with head support, turning to explore another artefact. The hover selection can also be extended by using a manual trigger (or other "click" method) to expand the contextual display for more detail. This demonstrates several advantages of the Eye&Head technique:

- Hover interaction can be driven by gaze while avoiding that the display changes

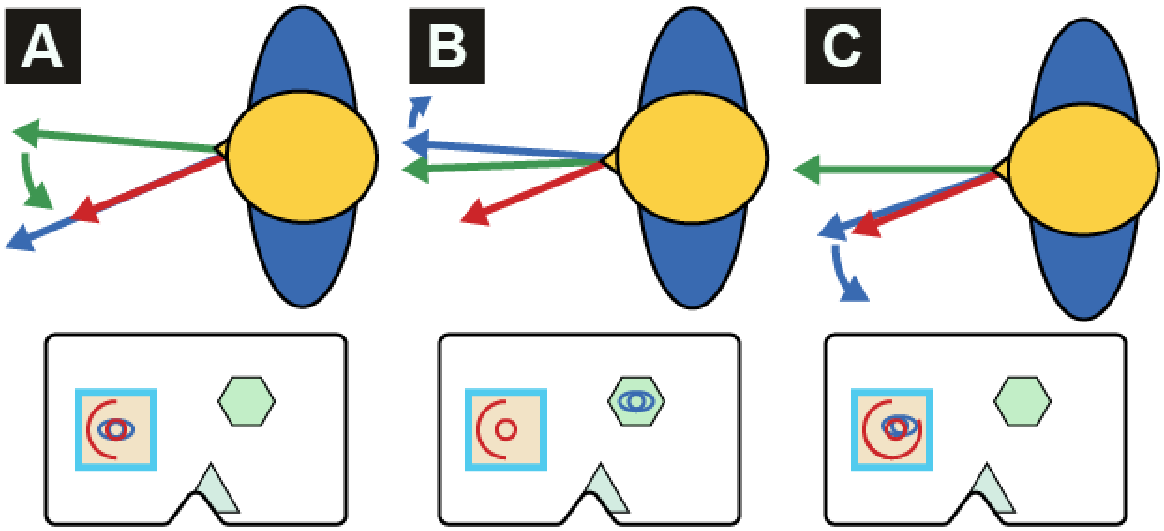

Figure 4.3: Eye&Head Dwell. A: Eye&Head pointing at an object triggers a dwell timer (red). B: The timer is paused when the user moves their gaze (blue) away from the object without also moving their head (green). C: The dwell timer resumes when the gaze returns to the selected object.

with every gaze shift.

- There is no need for users to carefully maintain their gaze on an object in order to maintain the selection.
- Users are free to visually explore the interface while a gaze selection is maintained.
- A gaze selection can be confirmed with a click method even when the user is no longer looking at the selected object.

4.2.2 Eye&Head Dwell

This technique complements Eye&Head pointing with a novel dwell method for confirmation of selected targets. As shown in Figure 4.3, a dwell timer is triggered only for targets that have been selected with head-supported gaze. If the user looks away from the target with eyes-only gaze, the timer is paused, and it resumes when they return their gaze to the target. If the user performs a head-supported gaze shift before dwell time has completed, selection is aborted and the timer reset.

Figure 4.4 shows Eye&Head Dwell with *Euler's Constellations*, a puzzle game we developed for illustration. Users are tasked to draw a star constellation by successive

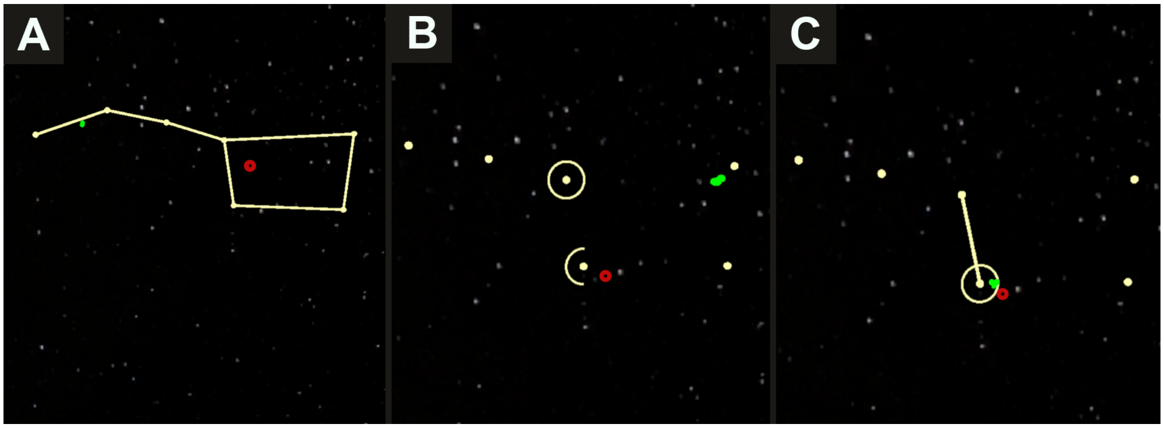

Figure 4.4: Eye&Head Dwell in a puzzle game (green dot: gaze; red dot: Eye&Head pointer). A: Display of a constellation to be reproduced by gaze. B: The user started selecting a star (near red dot) but moves their gaze to other stars to re-evaluate the selection, causing the dwell timer to pause. C: The user returns their gaze and completes the selection.

gaze selection of stars, with the challenge to draw in one line without traversing any path more than once. As such, users have to plan their selections ahead and revisit past selections to solve the puzzle. The Eye&Head Pointer combined with Eye&Head Dwell allows users to gaze on past and future selections without any time pressure and without risking that a selection is committed accidentally. This demonstrates key advantages the Eye&Head technique has over conventional gaze point-and-dwell:

- Users are free to dwell on potential targets without risk of unintended selection. This is useful for cognitively demanding tasks where thorough consideration of a choice can induce prolonged fixation.
- Selection can be paused while other options are inspected, and users save selection time when they decide to return to their first choice.

4.2.3 Eye&Head Convergence

Eye&Head Convergence is an alternative to dwell for confirmation, and can be combined with Eye&Head or conventional gaze pointing. The technique applies the principle of using head-eye alignment for confirmation. The underlying assumption is that the head does not fully align with the gaze vector when a new target is reached, for which we provide empirical support below. The additional head movement to "close the gap" can then be used to confirm the target selection. Figure 4.5 illustrates

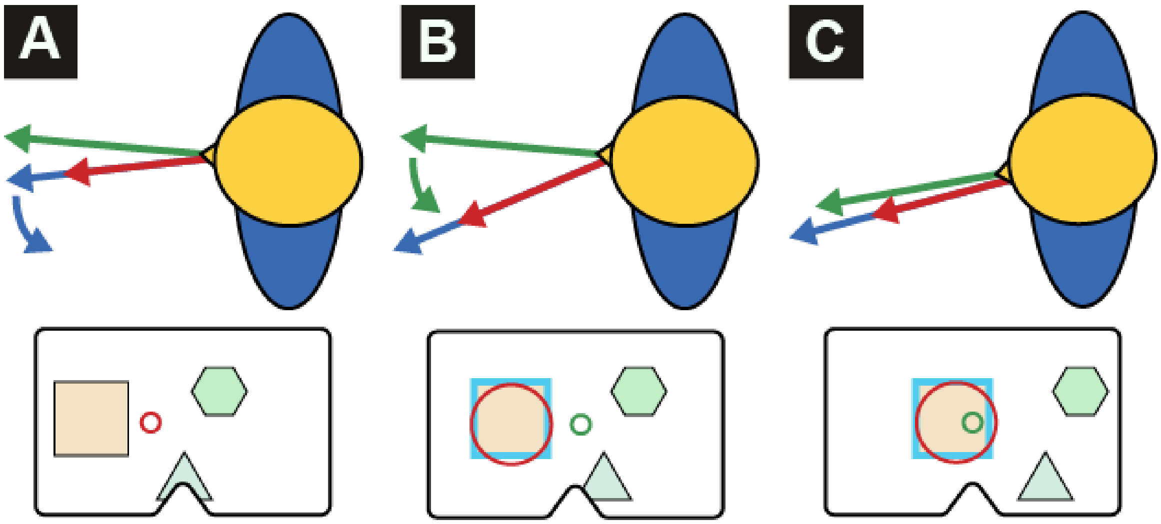

Figure 4.5: Eye&Head Convergence with conventional gaze. A: The pointer (red) follows the gaze (blue) toward the square. B: As gaze reaches the target, the cursor expands to define a convergence area and a head pointer (green) appears. C: The target selection is confirmed by moving the head pointer into the convergence area.

our implementation of the technique based on a cursor metaphor. When the user's gaze reaches a target, the gaze cursor expands to display a convergence area and additionally the head cursor is shown. The user can then confirm the selection by moving the head cursor to within the convergence area. If the head cursor is already within the convergence area, a timer is started during which the eyes and head have to remain within the threshold to confirm the selection.

The size of the convergence area is defined by an angular threshold around the gaze point, and impacts the behaviour of the technique. With larger angular thresholds, less head movement is needed and selections are faster. However a larger threshold also increases the likelihood that a head cursor is already within the convergence area, and the risk of accidental selection. A lower threshold reduces this risk, but requires more head movement. The required time to confirm a target (Equation 4.1), t, is defined as the angular distance from the gaze cursor to the head cursor, d, subtracting the angular threshold, a, divided by the angular head velocity, v. If the gaze and head distance are equal to or smaller than the chosen radius, the confirmation time is equal to a chosen dwell time, t_d.

$$t = \begin{cases} (d-a)/v & \text{if } d > a \\ t_d & \text{else} \end{cases} \tag{4.1}$$

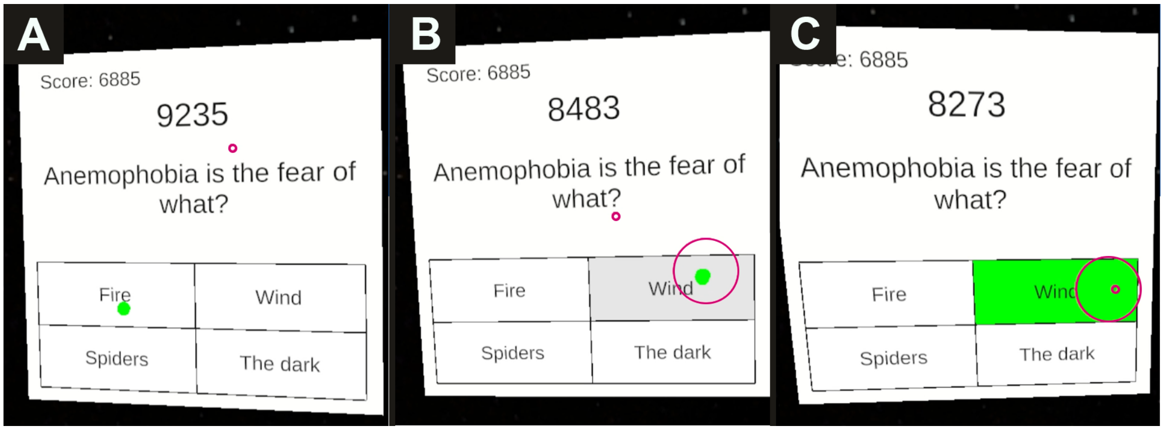

Figure 4.6: Eye&Head Convergence with Eye&Head Pointing in a quiz application. The green dot indicates gaze, and the red dot the head point. A: The user inspects answers without triggering interaction. B: When the user move their head toward an answer at which they look, the cursor jumps to their gaze point and a head pointer appears as the cursor expands to the convergence area. C: The user commits the selection by moving the head pointer to within the convergence area.

Figure 4.6 illustrates the convergence technique in a quiz game, combined with Eye&Head Pointing. Users are shown quiz questions and tasked to select the correct answer as quickly as possible. An erroneous answer awards no points and as such, users have to select both quickly and accurately. The application highlights a number of advantages of both Eye&Head Pointing and Eye&Head Convergence:

- Users are free to visually inspect choices without any distraction by cursor movement. This is useful when both speed and accuracy are important.
- Potential targets can be inspected for as long as users need to, without risking unintended selection.
- Users can traverse their gaze or head across other options when reaching for a target, without deselecting a currently highlighted option. This affords more freedom in the layout of choices on the interface.
- Selection by convergence can be faster than a conservative dwell time, as the required head motion can be performed in shorter time.
- Convergence is less error-prone for selection than short dwell times, as users have better control over their head movement than over the duration of gaze fixations.

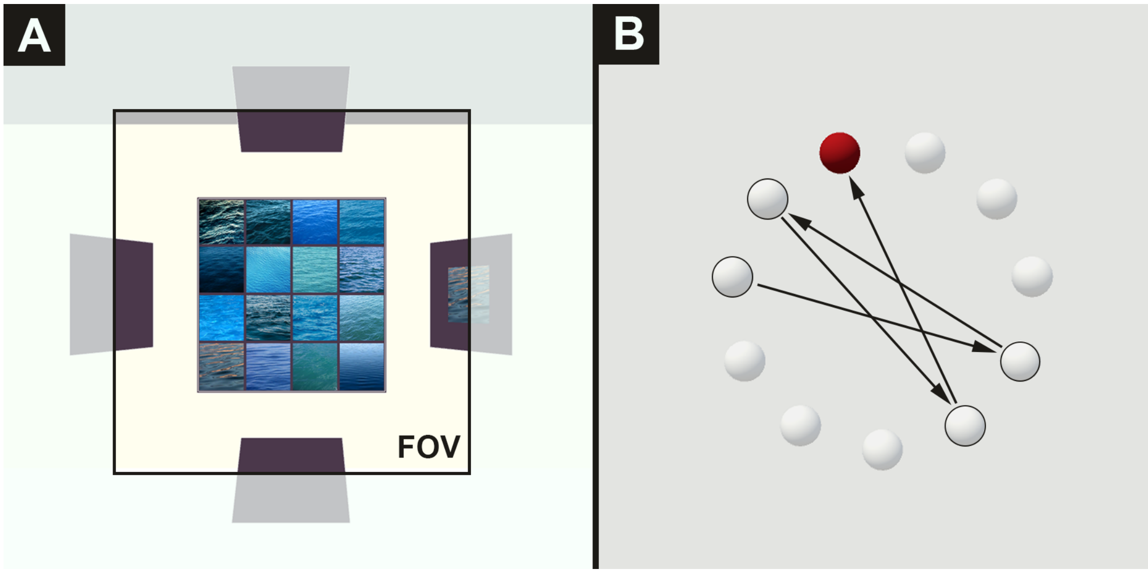

Figure 4.7: User study tasks. A: Participants were shown an image on one of the side panels and had to locate and select the matching image on the grid. Note, only one panel at a time was visible to the user. B: Participants were tasked to select the highlighted targets (red) in succession.

4.3 Performance Evaluation

The applications we developed demonstrate the potential qualitative advantages of Eye&Head over conventional eye pointing and selection. In addition, we conducted two user studies for evaluation, one on pointing and the other on confirmation. The objective of these studies was to compare user performance with Eye&Head against eye gaze pointing and dwell selection. The pointing study had the additional objective to quantify the offset between head and eye during naturalistic gaze, to test our assumption about head-eye alignment and inform the choice of a threshold for Eye&Head Convergence.

We designed two tasks for the purposes of our study (Figure 4.7). The first one is a search task, designed to require consideration of potential targets (and thus more gaze shifts) prior to selection. The second task, in contrast, highlighted individual targets in a predictable sequence, so that it could be performed with single gaze shifts. We used these tasks instead of our applications for evaluation, as they are more general and better suited for fair comparison against baselines.

Search Task. Participants had to find and select a matching picture within a 4x4 grid (Figure 4.7A). The original picture was shown on one of four panels surrounding the grid. 8 of the 16 pictures were shown at each panel, each picture being shown twice in total (32 trials per participant and condition). Each panel had a corresponding grid shuffle. The panel order and their corresponding stimuli order were randomised. All trials were performed sequentially without pause or head realignment to mimic a typical scenario where users perform multiple selections. The grid was placed at 4 metres distance from the participants and had a width and height of 50°. The centre of the adjacent panels was placed at 50° eccentricity from the grid centre to encourage movement outside the typical eye rotation range. We measured completion time of the whole task, number of errors (incorrect selections), amount of head and gaze movement, and offset between head and eye.

Circular Task. Here, participants had to select targets across a circular layout in a predictable sequence based on the ISO 9241-9 standard [71] (Figure 4.7B). The interface displayed eleven targets at 4 metre distance from the viewer. When all eleven targets had been selected, a new circle would appear. The target size was 4° in diameter, chosen to be large enough to minimise the effect of eye tracker error while avoiding target overlap. The diameter of the circle of targets was varied in four conditions with different pointing range (10°, 20°, 30°, 40°). Diameter sizes were chosen to have a mix of eccentricities where head movement would be unlikely (<20°) and likely (>20°), while avoiding that targets would move out of view and confound pointing with search. Participants performed five circles per condition (55 trials per condition), in a randomised order. We measured selection time, error rate, and the amount of head and gaze movement. Note this task was not included in analysis of head-eye alignment, as back-and-forth pointing discourages head following and would bias results.

4.3.1 Apparatus

We developed both tasks using Unity version 2017.4.3d1 and used an HTC Vive with the Tobii Pro VR Integration eye tracker (120Hz) for both studies.

4.4 Pointing User Study

This study compared two pointing techniques, Eye&Head and Gaze, for pointing. In the baseline technique (Gaze), the pointer followed gaze continuously. For both techniques, a cursor indicated the participant's pointing position and the HTC Vive hand-controller trigger was used to confirm selection. Eye&Head Pointing had a head rotational threshold of 15°/s and a translational threshold of 0.1m/s to activate the cursor, originating from informal testing.

Table 4.1: Search task performance and paired-samples t-test results. Results only showed differences in head movement and no other performance gain or loss when comparing gaze and Eye&Head pointing.

	Trial time (s)	Error (#)	Head motion (deg)
Gaze	196.79 (±52.37)	5.80 (±5.18)	1508 (±566)
Eye&Head	188.99 (±27.37)	5.30 (±4.05)	2651 (±678)
t-test	$t(11) = .601$, $p > .05$	$t(11) = .447$, $p > .05$	$t(11) = 4.401$, $p < .05$

4.4.1 Procedure

12 participants (5 female, 26.25 ±3.65 years) recruited from the local university participated in the study. Eleven participants had occasional VR experience, and one used VR daily. Ten participants had occasional or no eye tracking experience and two participants had daily experience. Participants first signed a consent form and answered a demographic questionnaire. Participants were then seated and put on the HMD. Participants started with the search task using both techniques before performing the circular task. Participants performed a five-point eye tracking calibration before each session. After calibration, participants had a training session before the test session. The pointing technique order was counterbalanced with a Latin square. After completing a task with a technique, participants removed the HMD and filled out a questionnaire consisting of eight 7-point Likert items based on common usability factors adopted from previous work [134]. A semi-structured interview was conducted after each completed task to extract preferences. The study took 30 minutes to complete.

4.4.2 Results

Unless otherwise stated, the analysis was performed with paired samples t-tests with the pointing technique ($Eye\&Head, Gaze$) as an independent variable. QQ-plots were used to validate the assumption of normality. Bonferroni-corrected post-hoc tests were used when applicable. Likert-scale usability data was analysed using Wilcoxon signed-rank tests.

4.4.2.1 Search Task

Paired samples t-tests showed no significant time or error difference (Table 4.1). However, Eye&Head Pointing showed significantly higher head movement. Figure 4.8

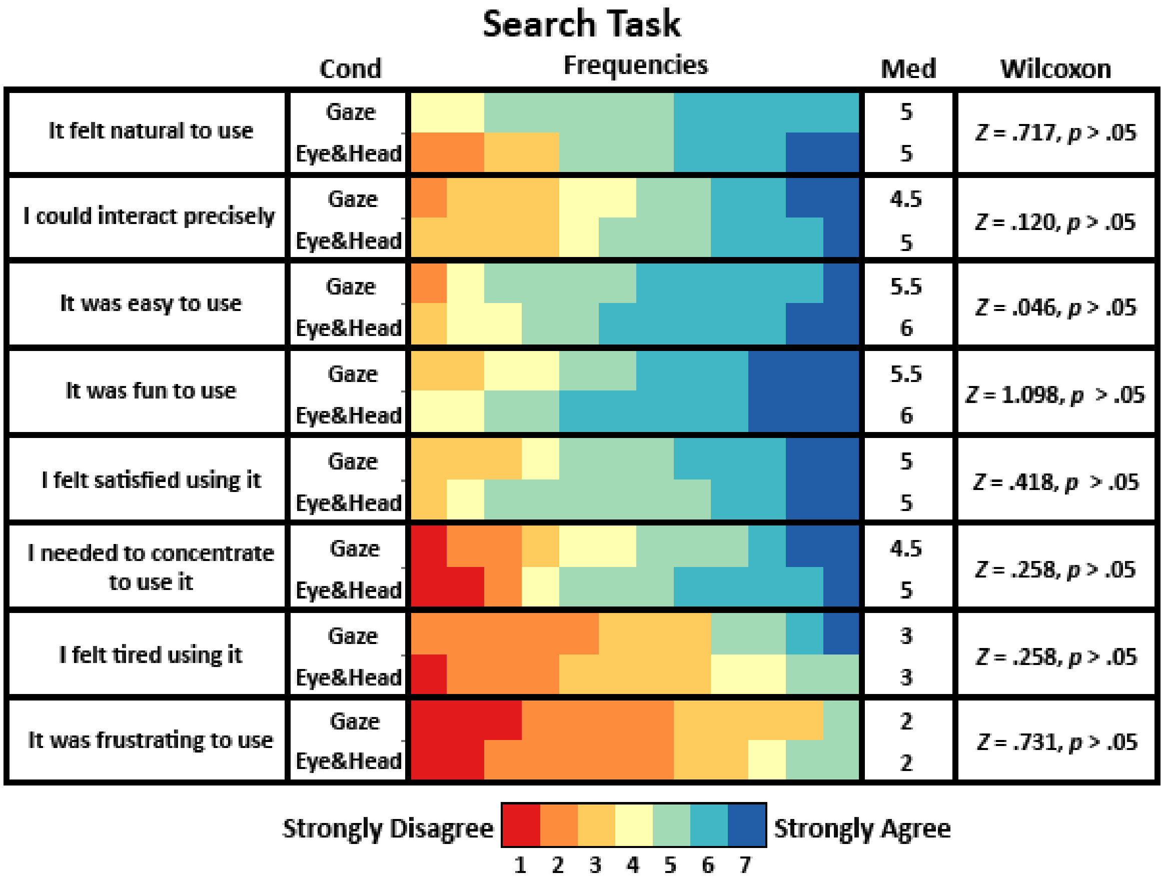

Figure 4.8: Search task questionnaire and Wilcoxon signed-rank test results. Results showed little differences between gaze and Eye&Head, indicating that adding head movements to pointing felt as natural as gaze while adding little strain.

shows questionnaire responses. Wilcoxon signed-rank tests showed no significant differences between the pointers.

Interview results showed that participants' opinions were split. Eight expressed favourable opinions towards Eye&Head Pointing. P5 claimed *"Eye&Head Pointing was much easier to use, and you did not have to focus as much since the pointer did not constantly follow your gaze"*. P2 commented *"Eye&Head Pointing is a nice way to get rid of all extra distractions and movements in the background"*. P2 also commented on the pointer's naturalness *"The head movement was already there, so for the majority of the selection I did not realise I was using the technique and that I needed an extra head movement"*. Finally, P9 noted *"Gaze pointing is more dependent on the eyes, and I felt like it was too responsive. I felt like I needed to concentrate*

Table 4.2: Circular task performance and paired-samples t-test results. Results only showed differences in head movement. The results show that adding head movement to gaze pointing does not add any interaction penalty.

		Time (s)	Error (%)	Throughput (bit/s)	Head motion (deg)
10°	Gaze	.43 (±.07)	7.4 (±4.9)	3.55 (±.55)	1.28 (±1.07)
	Eye&Head	.48 (±.14)	8.7 (±7.5)	3.41 (±1.33)	10.35 (±1.89)
	t-test	$t(11) = 1.509$, $p > .05$	$t(11) = .417$, $p > .05$	$t(11) = .403$, $p > .05$	$t(11) = 11.961$, $p < .001$
20°	Gaze	.50 (±.12)	10.0 (±5.8)	4.70 (±1.79)	2.39 (±2.31)
	Eye&Head	.49 (±.12)	13.4 (±6.7)	4.30 (±1.38)	12.88 (±2.46)
	t-test	$t(11) = .197$, $p > .05$	$t(11) = 1.277$, $p > .05$	$t(11) = .981$, $p > .05$	$t(11) = 9.861$, $p < .001$
30°	Gaze	.55 (±.09)	17.3 (±7.2)	4.34 (±1.15)	5.55 (±4.67)
	Eye&Head	.53 (±.12)	13.3 (±7.4)	4.72 (±1.60)	16.50 (±3.11)
	t-test	$t(11) = .758$, $p > .05$	$t(11) = 1.312$, $p > .05$	$t(11) = .993$, $p > .05$	$t(11) = 4.988$, $p < .05$
40°	Gaze	.62 (±.10)	19.4 (±7.1)	4.24 (± 1.21)	10.85 (± 2.54)
	Eye&Head	.59 (±.12)	16.4 (± 7.3)	4.51 (± 1.35)	21.85 (±3.88)
	t-test	$t(11) = .451$, $p > .05$	$t(11) = 1.264$, $p > .05$	$t(11) = .701$, $p > .05$	$t(11) = 3.546$, $p < .05$

on controlling it and it was distracting me". However, four participants preferred regular gaze pointing. Their main reason was the annoyance of being forced to use their head. P6 commented *"The Eye&Head Pointer was more of a challenge. It did not feel as natural as gaze pointing".* The participants of this group were among the five participants with the lowest head movement.

4.4.2.2 Circular Task

Analysis showed no significant differences in time, error rate or throughput. See Table 4.2 for comprehensive results. Eye&Head Pointing had significantly higher head movement at all distances but there was no significant difference in pointing performance. This was surprising for distances of 10° and 20° where targets are reachable eyes-only, and where we expected head motion to slow down pointing. However, efficiency of Eye&Head Pointing is explained by enabling users to still confirm a target while their gaze is already moving on.

Figure 4.9 shows the questionnaire results. The Eye&Head Pointer required

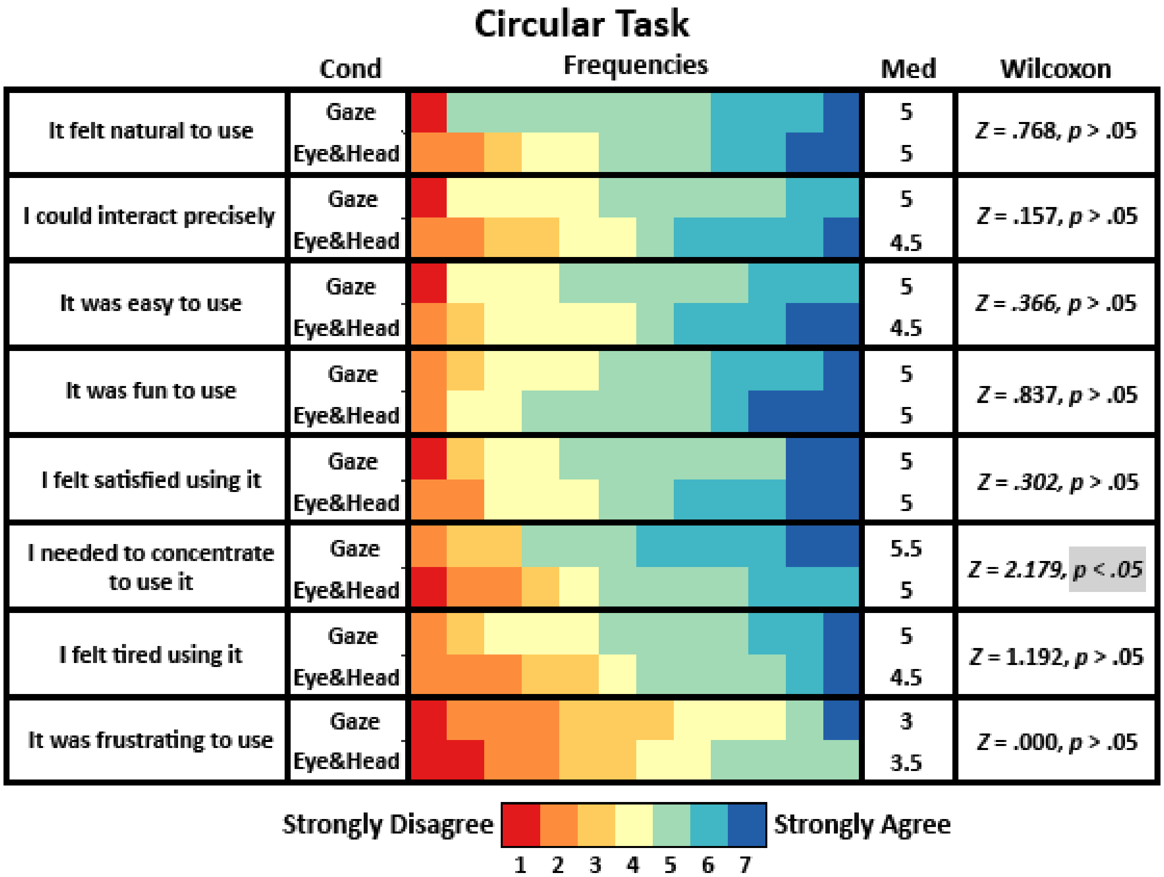

Figure 4.9: Circular task questionnaire and Wilcoxon signed-rank tests results. Results showed that users needed significantly less concentration with Eye&Head pointing.

significantly less perceived concentration. As in the search task, eight participants preferred Eye&Head Pointing while four preferred gaze pointing. Participants often performed mistakes with gaze pointing where the gaze would move faster between targets than they could press the button, whereas the Eye&Head Pointer provided more control over the selection. P1 expressed *"It was good that the pointer moved more discretely, it made it clearer to me what I was currently selecting"*. Participants also complained that the cursor was continuously following their gaze. P10 commented *"The cursor was distracting, and I felt I needed to concentrate more to select a target so that my gaze would not go somewhere else"*. However, four participants disliked the Eye&Head Pointer's required head motion, especially for shorter distances. P4 added *"I preferred gaze pointing for the shorter distances because the required head*

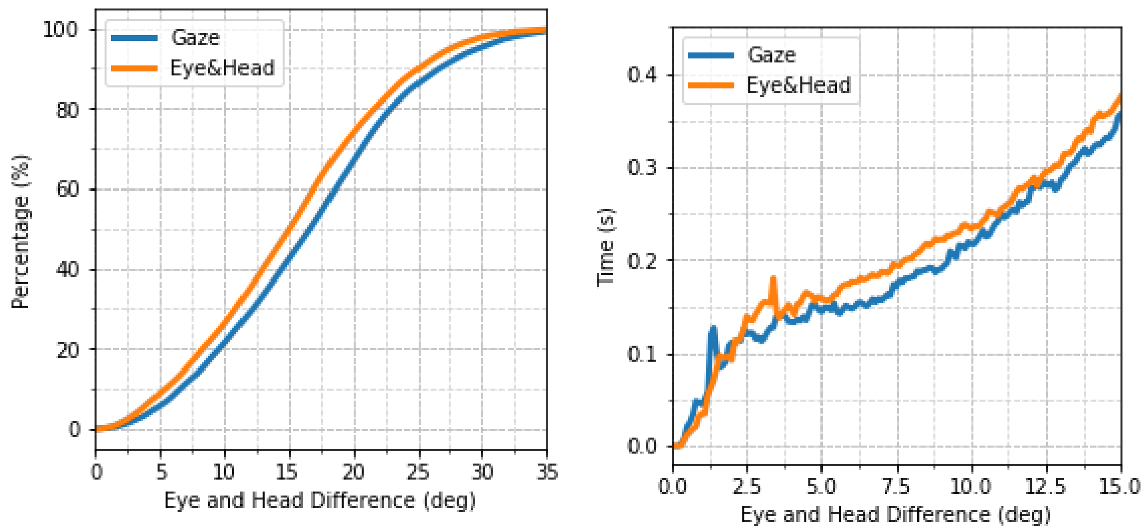

Figure 4.10: Left: Average percentage of search task spent within the angular difference between the eyes and head. We calculated the percentage of gaze samples during the search task where the eye-in-head angle was within a specified range (x-axis). Right: Average time spent within the angular difference between the eyes and head. We calculated the length (time) of segments were gaze samples were within a specified eye-in-head angle. Results show that the eye and head are rarely aligned, and would only occur when the eyes move across the head.

movement was annoying when I did not need it. However, it did not matter for the longer distances as I moved my head anyway".

4.4.2.3 Summary across Tasks

Performance results were consistent across tasks. Head modulation of the gaze pointer increases effort in terms of required head movement but this did not affect pointing speed and accuracy. Additionally, we found no significant differences between VR or eye tracking experiences. A participant majority (8 of 12) expressed favourable opinions towards Eye&Head Pointing as it gave them more control and fewer distractions. Results also indicated that Eye&Head Pointing required less concentration from participants. However, a participant subgroup that showed tendencies to rely less on the head did not favour Eye&Head Pointing due to annoyance or effort caused by the extra head motion.

4.4.2.4 Head-Eye Alignment

Head-eye alignment was analysed based on the search task (Figure 4.10). Unlike the circular task, the search task required users to inspect and compare images. This induced gaze shifts over different ranges which we deemed representative of natural gaze behaviour. We found that the offset between the head pointer and the gaze pointer was considerable for most of the time. On average participants would only spend 7.5% of the whole trial within 5° eye and head difference (Figure 4.10, left). Closer inspection showed that a closer alignment within this angle generally only occurred when the eyes would move across the head. Instances where head and eye were within 3° angular proximity were of short duration, with average time at 0.11-0.15s, which is significantly shorter than the time required for a gaze fixation.

We found no significant difference between regular gaze and Eye&Head Pointing in regards to head-eye alignment. However, Eye&Head Pointing led to 75% more head motion compared with unmodulated gaze, but this did not have any significant effect on head-eye alignment. The results support the proposed utility of head-eye alignment for signalling intent. This confirms the conceptual basis for the Eye&Head Convergence technique, and suggests a practical choice of angular threshold at 3° between head and eye.

4.5 Confirmation User Study

Our second study had the objective to evaluate the Eye&Head Dwell and Convergence techniques in comparison to regular gaze dwell. We used the same equipment, tasks and conditions as in the pointing study (Figure 4.7). The participants performed the tasks with four techniques: Gaze + Dwell (G + D), Eye&Head Pointing + Eye&Head Dwell (EH + D), Gaze + Eye&Head Convergence (G + C), and Eye&Head Pointing + Eye&Head Convergence (EH + C).

We chose a dwell time of 700ms for both dwell techniques and Convergence (t_d), designed to be proficiently usable by novice users and comparable to dwell times in prior similar work [18, 134, 141]. Other work has used dwell times as short as 300ms [56, 102] but such dwell times are for highly practised users and specific tasks [102]. Eye&Head Dwell had an angular threshold of 2° between the gaze point and cursor chosen via informal testing. Eye&Head Convergence parameters were decided via data collected from the pointing study (Figure 4.10). We set the angular threshold to 3° as participants had spent less than 2-3% of the search task time within this close range of head and gaze alignment. Occurences of alignment within this range had only lasted 110-150ms on average, well below the dwell time.

Table 4.3: Search task performance and repeated measures ANOVA. The results only showed a significant difference for head motion.

	Trial time (s)	Error (#)	Head motion (deg)
G + D	173.77 (± 36.76)	9.25 (± 8.51)	1559 (± 705)
EH + D	182.61 (± 39.55)	5.92 (± 2.84)	2391 (± 846)
G + C	181.38 (± 29.93)	5.42 (± 2.27)	3213 (± 711)
EH + C	176.30 (± 32.37)	5.83 (± 3.90)	3382 (± 678)
ANOVA	$F(3, 33) = .454$, $p > .05$	$F(3, 33) = 1.863$, $p > .05$	$F(3, 33) = 28.300$, $p < .001$

4.5.1 Procedure

12 participants (3 female, 28.08 ±3.55 years) participated in the study. Eleven reported occasional previous experience with VR, and one reported daily to weekly VR use. Ten had occasional experience with eye tracking and two participants reported daily to weekly experience. Six had participated in the pointing study. The same procedure was used as in the first study. The study took 45 minutes to complete.

4.5.2 Results

Unless otherwise stated, the analysis was performed with one-way repeated measures ANOVA with the selection technique $(G + D, EH + D, G + C, EH + C)$ as an independent variable. When the assumption of sphericity was violated, as tested with Mauchly's test, Greenhouse-Geisser corrected values were used in the analysis. QQ-plots were used to validate the assumption of normality. Bonferroni-corrected post-hoc tests were used when applicable. Likert-scale usability data was analysed using Friedman tests and Bonferroni-corrected Wilcoxon signed-rank tests were used for the post-hoc analysis.

4.5.2.1 Search Task

Repeated measures ANOVA showed no significant time or error rate differences (Table 4.3). However, G + D had a larger error count variance compared to the rest. Significant differences were found in head motion. Further Bonferroni corrected post-analysis showed that G + D also had significantly lower head motion compared EH + D ($p < .05$) and both Convergence combinations (both $p < .001$). Additionally, EH + D had a significantly lower head motion compared to both G + C ($p < .001$) and EH + C ($p < .05$). The head motion differences are not surprising as Eye&Head Convergence

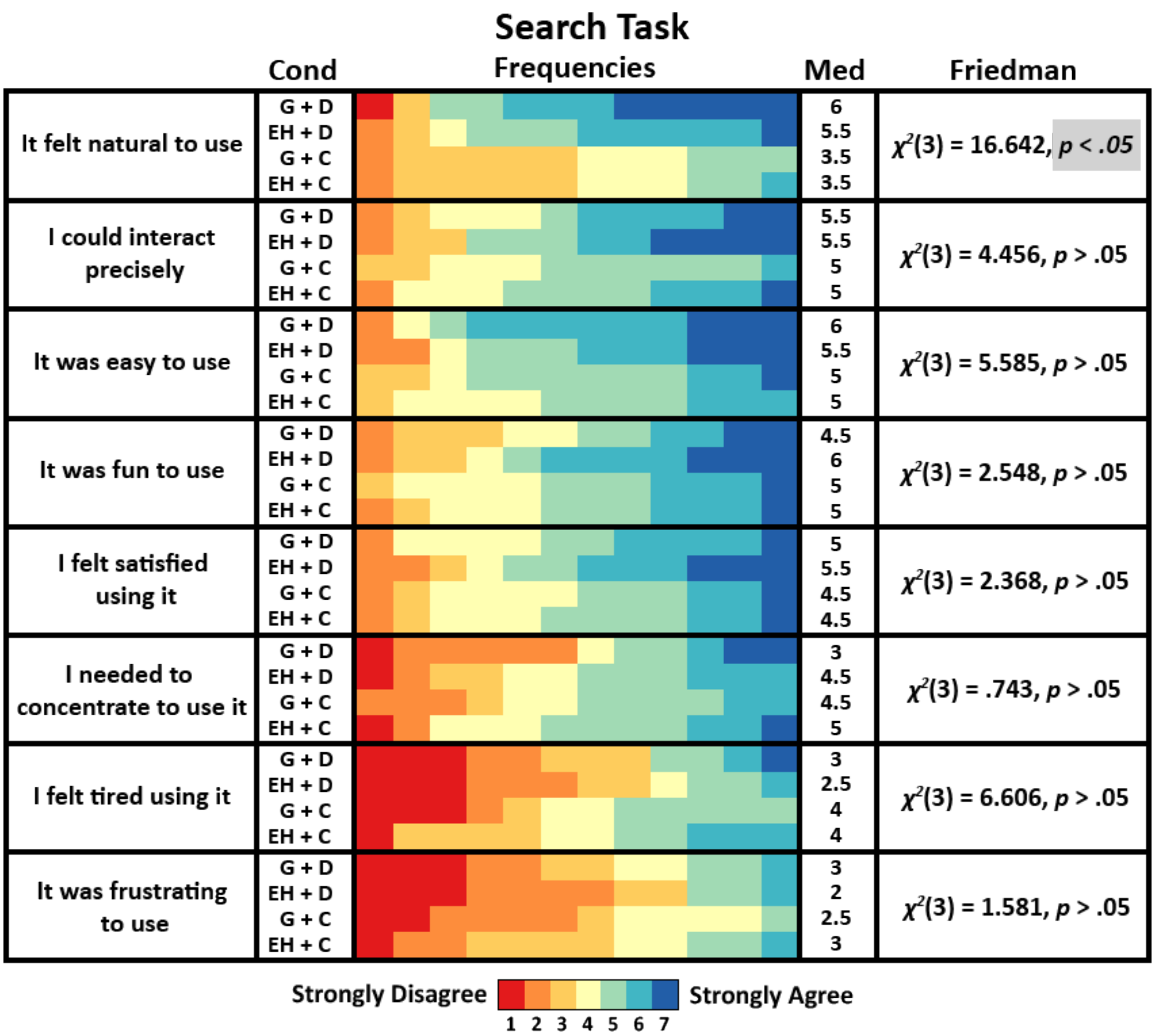

Figure 4.11: Questionnaire and Friedman test results for the search task. Results showed a significant difference in naturalness.

to some extent require head pointing, and Eye&Head Pointing also requires head movement to update the cursor position.

Figure 4.11 shows the usability ratings from the search task. Friedman tests showed significant differences in naturalness. Further Bonferroni adjusted Wilcoxon analysis showed that G + D was significantly more natural than G + C ($z = 2.434$, $p < .05$) and EH + C ($z = 2.297$, $p < .05$). EH + D was also significantly more natural than G + C ($z = 2.383$, $p < .05$) and EH + C ($z = 2.683$, $p < .05$).

Participants' opinions about the techniques were again varied. Six participants preferred EH + D. P9 expressed *"EH + D was good because it did not feel as stressful as G + D and not as tiring as G + C and EH + C"*. P10 added *"EH + D was very useful as it allowed me to move more freely without making a selection compared to the other techniques"*. P11 also stated *EH + D gave me more control over my selections and it suited very well with the search task*. Three participants preferred EH + C. P4 stated *"EH + C felt more natural, and I think I am inclined to be more precise when also using my head. The dwell techniques were tiring because you had to stare at a target which felt unnatural"*. Three participants mentioned that the expanded cursor used in Eye&Head Convergence was distracting and added that the EH + C was less distracting than G + C due to its discrete nature. However, the remaining participants expressed no major difference between them. Just as in the first study, three participants who tended to use less head movement preferred G + D which required the least head movement. P12 stated *"Moving my gaze felt more natural and effortless compared to moving my head"*.

4.5.2.2 Circular Task

We found significant circular task performance differences between the techniques (Table 4.4). Bonferroni corrected pairwise comparisons showed that G + C and EH + C compared to G + D and EH + D at all distances had significantly faster selection times (all $p < .001$, except EH + D and G + C at 40° ($p < .05$)), confirm times (all $p < .001$), and higher throughput (all $p < .001$ except EH + D and G + C at 30° ($p > .05$) and 40° ($p > .05$)). Head motion was significantly lower for G + D compared to all other techniques at all distances (all $p < .001$ except G + D and EH + D at 30° ($p > .05$), and 40° ($p > .05$)). Additionally, EH + D had a significantly lower head motion compared to both Convergence techniques at all distances (all $p < .001$).

In a post hoc analysis, we simulated shorter dwell times to investigate whether significant differences in selection and confirmation times were only due to the more conservative choice of dwell time. Results showed that both EH + C and G + C were significantly faster for all four distances also with a lower dwell time of 500ms; with dwell time chosen as low as 300ms, EH + C and G + C were still significantly faster for 10° distance, but not the larger distances. Note, that a larger Eye&Head Convergence angular threshold would lead to shorter confirm times and thus shorter times and higher throughput as less head motion would be required.

Friedman tests on usability ratings showed significant differences in naturalness, precision, easiness and enjoyment (Figure 4.12). Bonferroni corrected Wilcoxon analysis showed that participants considered EH + C to be significantly more natural ($z = 2.919$, $p < .05$) and easier to use ($z = 2.972$, $p < .05$) than G + C . Participants also considered EH + C to be significantly more precise than G + C ($z = 2.714$, $p < .05$) and EH + D ($z = 2.200$, $p < .05$). Finally, EH + C was considered more fun

Table 4.4: Circular task performance and repeated measures ANOVA. The results showed a significant differences for all metrics.

		Time (s)	Confirm time (s)	Throughput (bit/s)	Head motion (deg)
10°	G + D	.94 (±.05)	.70 (±.03)	2.75 (±.14)	2.06 (±2.05)
	EH + D	1.08 (±.03)	.73 (±.03)	2.34 (±.11)	10.80 (±1.38)
	G + C	.54 (±.10)	.24 (±.08)	4.95 (±.89)	13.96 (±1.57)
	EH + C	.52 (±.12)	.20 (±.06)	5.12 (±.97)	14.40 (±1.91)
	ANOVA	$F(3, 33) = 162.6$, $p < .001$	$F(3, 33) = 504.9$, $p < .001$	$F(3, 33) = 68.6$, $p < .001$	$F(3, 33) = 155.7$, $p < .001$
20°	G + D	1.00 (±.06)	.70 (±.02)	3.52 (±.18)	4.35 (±4.03)
	EH + D	1.09 (±.07)	.73 (±.06)	3.20 (±.22)	12.29 (±2.07)
	G + C	.74 (±.17)	.35 (±.11)	5.00 (±.80)	23.73 (±1.89)
	EH + C	.70 (±.15)	.33 (±.08)	5.28 (±.78)	24.87 (±2.03)
	ANOVA	$F(3, 33) = 26.6$, $p < .001$	$F(11) = 105.1$, $p < .001$	$F(3, 33) = 40.8$, $p < .001$	$F(3, 33) = 176.4$, $p < .001$
30°	G + D	1.14 (±.13)	.76 (±.07)	3.46 (±.54)	7.85 (±6.91)
	EH + D	1.05 (±.09)	.75 (±.08)	3.69 (±.55)	15.03 (±2.82)
	G + C	.89 (±.20)	.42 (±.13)	4.75 (±1.00)	32.65 (±1.46)
	EH + C	.84 (±.17)	.41 (±.16)	4.96 (±.88)	34.08 (±2.39)
	ANOVA	$F(3, 33) = 15.4$, $p < .001$	$F(3, 33) = 33.5$, $p < .001$	$F(3, 33) = 19.0$, $p < .001$	$F(3, 33) = 128.7$, $p < .001$
40°	G + D	1.33 (±.26)	.79 (±.10)	3.29 (± .52)	12.40 (± 8.39)
	EH + D	1.20 (±.28)	.77 (± .08)	3.84 (± .65)	19.36 (±5.05)
	G + C	1.02 (±.19)	.47 (± .13)	4.49 (± .80)	42.19 (±2.12)
	EH + C	.93 (±.18)	.42 (± .13)	4.93 (± .91)	43.46 (±2.47)
	ANOVA	$F(3, 33) = 21.0$, $p < .001$	$F(3, 33) = 41.5$, $p < .001$	$F(3, 33) = 23.8$, $p < .001$	$F(3, 33) = 101.7$, $p < .001$

than G + D ($z = 2.139$, $p < .05$).

Eight participants favoured the EH + C technique. P9 stated *"Eye&Head Convergence is easy and quick to select something when you know its position"*. P5 added *"I preferred EH + C for the circular task. It was more responsive than EH + D and G + D and less distracting than G + C"*. The chosen dwell time had a clear effect on the participants' responses. However, participants that expressed favourable opinions on Eye&Head Convergence thought that the chosen dwell time did not matter as the Eye&Head Convergence selection was instantaneous when reaching the angular threshold. One participant preferred EH + D for the circular task. Similarly to the search task, three participants preferred G + D. P12 stated *"It was hard and annoying to use my head all the time. I preferred the techniques where I could rely more on the*

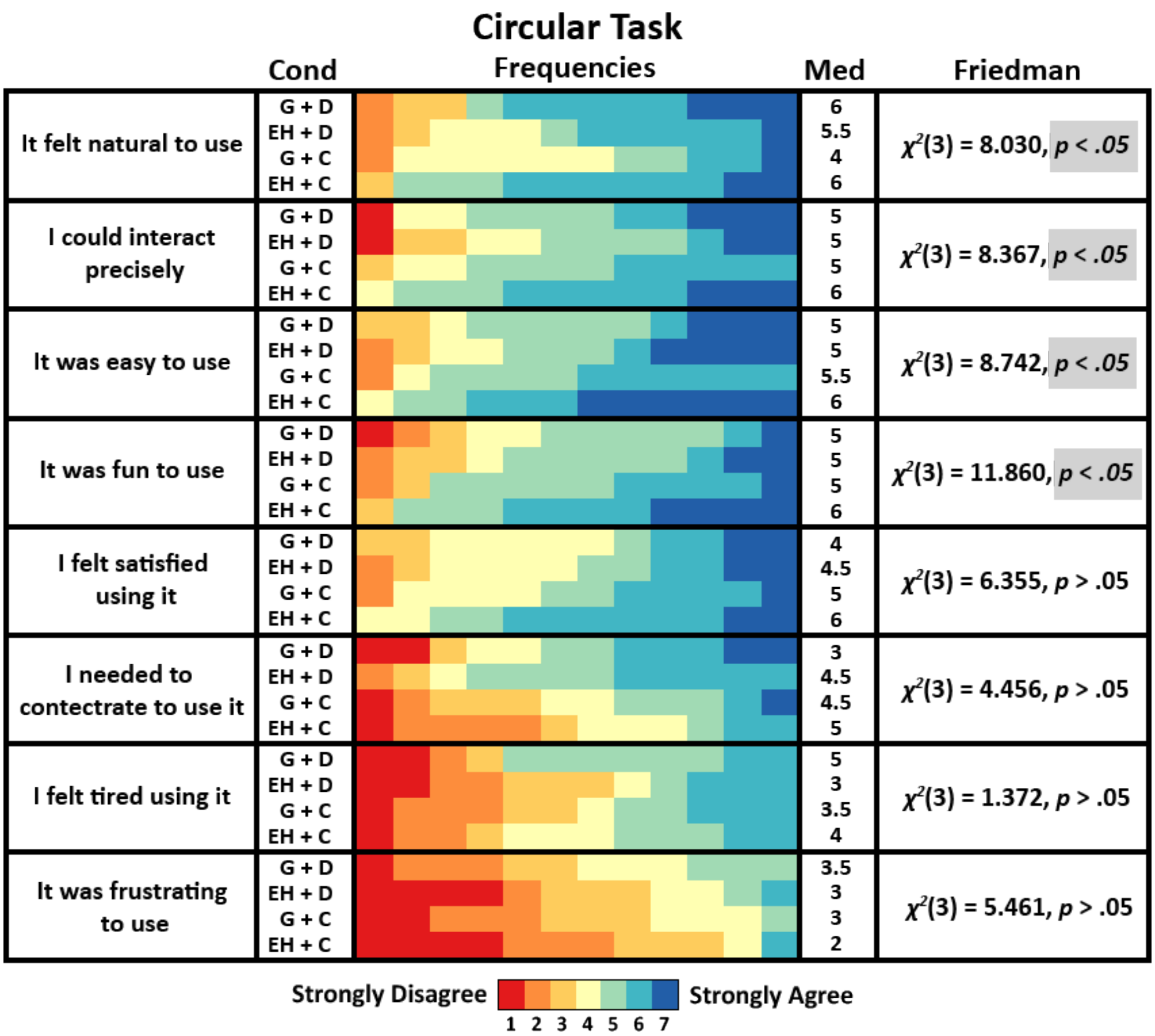

Figure 4.12: Questionnaire and Friedman test results for the circular task. The results showed a significant differences for naturalness, precision, ease and fun.

eyes". Finally, P6 commented *"Convergence was really easy for close targets. But not for big movements, then I preferred selection by gaze"*.

4.5.2.3 Summary across Tasks

The type of task affected both performance and preference. No significant performance differences were observed when participants had to search for targets to select. However, when participants knew the target in advance they were significantly faster using Eye&Head Convergence for confirmation than with a dwell technique.

Participants had differing opinions regarding their preferred technique, but mainly expressed favourable opinions for the combination of EH + D for the search task and the combination of EH + C for the circular task. Participants expressed preference for the Eye&Head Pointer as it provided more control and was found less distracting. As in the first study, a subgroup of the participants favoured regular gaze techniques due to annoyance or effort caused by the extra head movement needed with Eye&Head techniques.

4.6 Discussion

At the core of Eye&Head interaction is the distinction between head-supported gaze and eyes-only gaze. Head movement requires more effort and energy than eye movement, and an attention shift supported by the head can be considered to represent a higher level of investment and interest. Based on the distinction, different behaviours can be attached to objects, depending on whether they are turned to by both head and eyes, looked at without head turn, or not gazed at. In our application examples, we have attached automated gaze behaviour only to the higher level of interest, to allow for exploratory attention to objects without side effect. However, other mappings are possible. In visual search, for example, all objects looked at could be marked as viewed, and head-supported attention could additionally trigger selection.

The three principles we proposed for Eye&Head interaction are validated by the application examples and study results. The first principle refers to stability of head-support gaze and is directly reflected in the design of the Eye&Head Pointer. The museum and puzzle applications show how the pointer facilitates stable gaze selection and feedback decoupled from individual fixations. Also, in both studies the majority of users found the Eye&Head pointer to provide more control and less distraction. The second principle is that eyes-only gaze affords exploration around objects selected by head-supported gaze. All our applications illustrate this, for example, with free exploration around an artefact of interest in the virtual museum, and examination of alternative choices after initial selection in the puzzle. The third principle is that head-eye alignment can be used as explicit input. In our first study, we showed that head and eye do not normally become completely aligned, a premise for using alignment as deliberate signal. We applied the principle in the Eye&Head Convergence technique, and the results of our second study show that the technique is robust and effective for fast confirmation of gaze targets.

A principal advantage of Eye&Head pointing is that objects can be pre-selected by gaze but that gaze is free to wander before the selection is finalised. This can be useful for many scenarios, for instance double-checking other conditions before finalising selection, or completing selection in sync with other events. It also avoids that the selection focus is lost prior to completion, for example caused by jitter in the eye movement, eye-tracking inaccuracy, or a visual distraction. This could have a positive impact on the cognitive demand on users over time, as they do not need constant control of gaze-pointing. Future longitudinal studies would be of interest. Longitudinal experiments would also be of interest in investigating the impact of dwell time on technique performances. As has been shown in work on gaze typing [102], more proficient users could use less dwell time for greater performance.

Questionnaire results from the pointing study showed no significance difference between pointers (Figure 4.8 and Figure 4.9). A possible explanation for these results could be that our study tasks were too simple for the participants to notice the interaction benefits from Eye&Head Pointing. The results could also mean that head modulation does not affect usability factors in a significant manner. However, a key insight from our studies is that gaze target acquisition is not slowed down by head modulation for the tasks that we used. This is significant as the demonstrated advantages of moving and updating the pointer only with head-supported gaze are gained without comprising performance. Eye&Head Pointing was designed with exploratory gaze applications in mind but our results show that it is also efficient for pointing at known targets in fast succession.

A specific benefit for fast pointing is that targets can still be confirmed when the eyes are already moving to the next object. This is compelling for applications where gaze is combined with a separate "click" modality such as a button or other manual trigger, as it requires less concentration from the user for timing of gaze shifts. The user's eyes can move on when they are ready and do not need to wait until their hands have caught up. This matches natural eye-hand coordination, where the eyes lead manual action [202].

Eye&Head interaction can also be completely hands-free for which we introduced two novel "click" alternatives. Eye&Head Dwell extends the advantages of Eye&Head Pointing to dwell selection. A dwell-timer is only triggered when gaze is accompanied by head movement, providing more control over selection, and leaving the user free to pause selection to look at other objects. User can gaze at alternatives choices and consider them without time pressure, as the dwell-timer is only activated when both eyes and head move. This also addresses problems of gaze interaction with large objects where users require more time for visual inspection, a problem with regular dwell as it can result in unintended selection.

Eye&Head Convergence presents an alternative to dwell and employs alignment of the head pointer with the line of sight as the "click" alternative. The two techniques bring different strength to different applications. The convergence technique is faster and preferred when speed matters or targets are known, for example when selecting a tool from a menu and quickly returning to where the action is. Eye&Head Dwell, in contrast, is perceived as more natural when the tasks involves search and consideration of targets for gaze selection. However both techniques have in common that they support more stable selection, for example in crowded environments with selectable objects overlapping or in close proximity.

Eye&Head techniques can support users limited to eye and head movement for interface operation, as well as users leveraging gaze in conjunction with other modalities. We observed that a majority but not all users preferred Eye&Head over unmodulated gaze. Eye-head coordination literature suggests that there are "head-movers" versus "non-head-movers" [47], which our studies appear to confirm. Users preferring regular gaze disliked that our techniques required head movement, and on average moved their heads less than other participants. These results indicate the possibility of the Eye&Head techniques becoming burdensome over time due to the additional head movements. The techniques may also be less appropriate for users with physical disabilities or injuries that restrict head movement. Eye&Head pointing accommodates a reluctance to move the head as it requires only little head movement and does not depend on the extent to which the head supports a gaze shift. The sensitivity of our techniques could also be adapted for non-head-movers, for example lowering the rotational threshold at which head motion is detected, and increasing the selection radius in Eye&Head Convergence. Further longitudinal studies of the Eye&Head techniques' and dynamic thresholds based on head movement tendency would thus be of interest.

Eye&Head pointing is unique compared to other multimodal techniques in that only one modality (gaze) is used to point while the second modality (head) is used as a clutch to trigger pointing. Generally, multimodal pointing techniques, such as MAGIC pointing [202], Pinpointing [87], and other work [126, 168]. use multiple modalities in succession for pointing. This increases accuracy as a second, more accurate modality to refine from a course position. However, this sequential interaction procedure requires users to go through multiple steps for simple actions and can be straining and time-consuming over a long time. Our approach results in no penalty caused by extra movements while allowing more freedom in the pointing procedure compared to a normal gaze pointer. However, in contrast to most existing work, the accuracy of our technique depends on gaze accuracy, as the head is not used for pointing refinement. Similarly, the multimodal Eye&Head Convergence technique relies on alignment between two modalities where it is expected that users will align one modality with the other (i.e. head-to-gaze or gaze-to-head). Other work on bimanual interaction has investigated similar concepts in which two rays controlled by each hand are aligned so that the target is selected at the ray intersection [194]. A key distinction between our work and other work mentioned is that we allow hands-free interaction of objects while most other work rely on the hands in different capacities.

In contrast to other multi-modal pointings techniques such as MAGIC pointing Pinpointing , or other refinement-based methods is that

All our results were obtained in VR. Studies have observed users moving their head more in VR than in comparable real-world tasks, caused by peripheral FOV limitations of HMDs [84, 125]. However, we do not expect this to limit applicability of our techniques, as they build on basic eye-head coordination behaviours that are consistent with observations in real-world tasks [88]. Synergetic eye and head movement is more prevalent when interactions span a wider FOV, for instance on large displays or across devices in smart rooms, but our techniques are also applicable with narrower FOV displays and deliberate head movement.

4.7 Conclusion

This chapter introduced Eye&Head gaze interaction with design principles and techniques that we validated in application prototypes and user studies. Our main conclusions for gaze interaction design are: (1) It proves useful to distinguish between head-supported and eyes-only gaze; (2) Modulation of a gaze pointer and/or dwell timer by head motion provides users with more stable feedback, better control, and freedom to roam with eyes-only gaze, without compromising pointing efficiency; (3) Eye-head convergence is viable as signal of intent, and enables fast hands-free target confirmation.

The work in this chapter proved the usefulness of two of our design principles: *Differentiate between head-supported and eyes-only gaze shifts* and *Use head-eye alignment to signal intent*, and how they can be used to tackle the Midas touch issue and lack of confirmation techniques for expressive input in novel ways. In the following chapter we will expand the work on eye-head alignment. The chapter presents Radi-Eye for expressive and hands-free interaction through a holistic radial interface design. Radi-Eye incorporates goal-crossing as a new selection mechanic for gaze- and head-only interaction to expand the repertoire of techniques for selection and manipulation.

Chapter 5

Radi-Eye: Hands-free Radial Interfaces for 3D interaction using Gaze-activated Head-crossing

In the previous chapter, we have shown that input from the eyes and head can be combined for fast and robust pointing and selection, based on the natural coordination of the eye and head in directing gaze. In this chapter, we extend the work on the eye-head alignment design principle and introduce a holistic interface design, *Radi-Eye*, for user control and expressive interaction with only eye and head movements to showcase how coordination can be used beyond simple selection (RQ2). The work employs our proposed design principle of eye-head alignment combined with design considerations from Chapter 3. Through these insights, we develop a confirmation mechanism that provides advanced widgets such as sliders and nested menus while avoiding Midas Touch issues.

Radi-Eye is designed as a pop-up radial interface that provides widgets for discrete and continuous input, contextual interaction, nested control, and toggling of content. When invoked, Radi-Eye pops up in a head-centred position from where radially arranged widgets can be accessed comfortably and efficiently with gaze and head movement. While presented as a radial menu, the interface is scalable to large sets of features through additional rings, toggling of components displayed on the rings, and adaptation to objects over which Radi-Eye is opened. Figure 5.1 illustrates some of the affordances.

Look & Cross is a gaze-activated head-crossing technique that employs gaze for hover interaction and pre-selection of widgets, and head-crossing to complete selection of a gaze-activated widget. This is natural and efficient for widget selection, as head movement naturally follows eye gaze. It avoids Midas Touch, as head orientation normally remains offset from gaze direction unless users explicitly choose to fully align

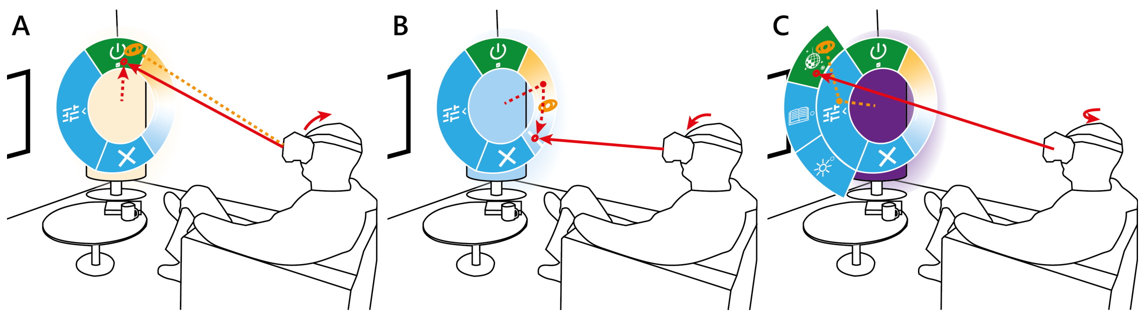

Figure 5.1: Radi-Eye in a smart home environment for control of appliances. A: The user turns on the lamp via a toggle selection with minimal effort using only gaze (orange) and head (red) movements. B: Selection can be expanded to subsequent head-controlled continuous interaction. The user moves the head cursor within the widget to adjust the light colour via slider interaction. The slider handle position is determined by the head position. C: Gaze-triggered nested levels support a large number of widgets and easy selection of one of the multiple preset lighting modes. The widgets enabled via Radi-Eye allow a high-level of hands-free and at-a-distance control of objects from any position.

head and gaze [151, 152]. As a result, Look & Cross enables fluid selection of multiple objects across the interface, without risk of unintended activation of objects crossed by head or gaze alone. The design of the technique reflects the relative strengths of gaze for visual exploration and fast pointing and head movement for more deliberate and precise control, which can also seamlessly extend from crossing-based selection to manipulation of continuous inputs.

Radi-Eye is designed to maximise user control and expressiveness with only eye and head movement for interaction. The radial interface structure combined with Look & Cross for fluid eye and head control enables a novel style of hands-free HMD interface that we explore through the implementation of three applications in VR and AR. The applications provide insight into the design space of Radi-Eye and design considerations in the interplay of interface layout, eye-head interaction, and visual feedback strategies. In a user study, we then evaluate the effect of radial interface parameters on Look & Cross performance, to gain insight into design choices for ring and button sizes.

5.1 Radial Interfaces and Crossing

For the design of Radi-Eye we are building on insight on eye-head coordination and the combined use of gaze and head motion from previous chapters, as well as prior work on radial interfaces and crossing.

5.1.1 Radial User Interfaces

Radial user interfaces place items along the circumference of a circle or ring. They were initially introduced as an alternative desktop interface in the 1980s [22]. Researchers have since established multiple advantages over traditional linear menus. Radial interfaces afford equivalent distance to all items, while also exploiting users' spatial memory by placing items in separate directions, allowing fast selection while reducing the need for precise pointing [11, 22]. Interaction starts from the centre, which makes radial interfaces attractive for modalities that have a natural "central state" from which they can move in any direction [25, 66, 116, 195, 196]. However, radial interfaces are limited to a number of items before there is an increase in erroneous selections due to decreased item size [182]. Also, as radial interfaces assume that the cursor is placed in the centre, they cannot be invoked at the edge of a screen without disrupting the interface structure [63]. In Radi-Eye, the pop-up interface combined with the HMD-based interaction ensures that interaction can always start from the centre.

In 3D environments, research has shown that radial interfaces perform better than their linear counterparts [30, 137, 146]. Gebhardt et al. also extended a hand-controlled radial interface to include more advanced widgets such as check-boxes, radio buttons, and sliders [50]. Although a hand controller using ray-casting combined with a button for confirmation is the dominant modality for radial interface input in 3D environments [32, 50, 51, 94, 137, 146], researchers have also explored a wide range of hands-free modalities. Previous work has shown that head [137, 196], feet [116], or body movements [195] can be effectively used for radial interface control. We extend the work on hands-free interface input beyond basic selection by combining gaze and head pointing for a variety of widgets that support discrete and continuous user input, interface scaling, and command composition for expressive user control.

5.1.2 Crossing

Crossing is a selection technique that selects a target by crossing its boundary with a cursor [1]. The technique allows fast and accurate selection of targets [1], and has been shown to be as expressive as the common pointing metaphor [6]. Crossing relaxes the constraint on fine-grained pointing within a closed area, as the user only has to cross the target boundary. Researchers have proposed crossing for users with limited motor capabilities [192], and for a wide array of modalities with limited fine-grained pointing or where the modality lacks an explicit confirmation mechanism [16, 100, 182, 198]. More recently, researchers have shown that crossing can be used effectively in 3D environments by ray-casting with a controller [174], or for hands-free selection

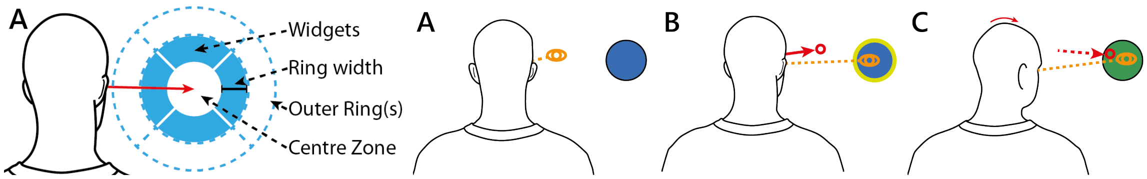

Figure 5.2: Radi-Eye layout (A) and Look & Cross object selection (B-D). A: Radi-Eye layout and components. B: The object is idle. C: Gaze (orange) moves to the object, triggering hover interaction, enabling the object for selection, and displays a head cursor (red). D: The head moves across the gazed-on object's boundary, selecting the object.

via the head [196, 198]. Finally, crossing also allows selection of multiple targets in a single user input [6, 169], allowing fast and expressive user interfaces.

Similarly to the Midas Touch problem, a common issue with crossing interaction is distractor targets that are in-between the cursor and the intended target, forcing the user to interrupt an input [7]. Modalities such as a stylus or mouse solve this issue by lifting the stylus or releasing a mouse button [7, 192]. However, the issue is more problematic for the head and gaze that are "always on" and have no inherent confirmation mechanism. Placing items in different directions makes radial interfaces an ideal interface for crossing interaction as it reduces the risk of distractor targets [1]. Alternatively, researchers have proposed techniques that add additional steps to the interaction, by for example, forcing the user to exit the target in the same direction as they entered [198], or moving the cursor to a secondary target after crossing [112, 147]. Finally, work has proposed to use gaze to "enable" targets as selectable by a second modality [192]. We build on this notion by using eye gaze to activate targets for crossing selection with the head, which allows users to freely move their head and gaze over distractor targets without triggering accidental selections.

5.2 Radi-Eye

The core concept of *Radi-Eye* is to use an eye and head-controlled interface for expressive hands-free control of objects in any 3D environment. To achieve this, Radi-Eye consists of three parts:

1. A pop-up radial interface for on-demand interaction;
2. Look & Cross, a gaze-activated head-crossing selection technique;
3. Widgets that enable discrete and continuous interaction, and interface scaling.

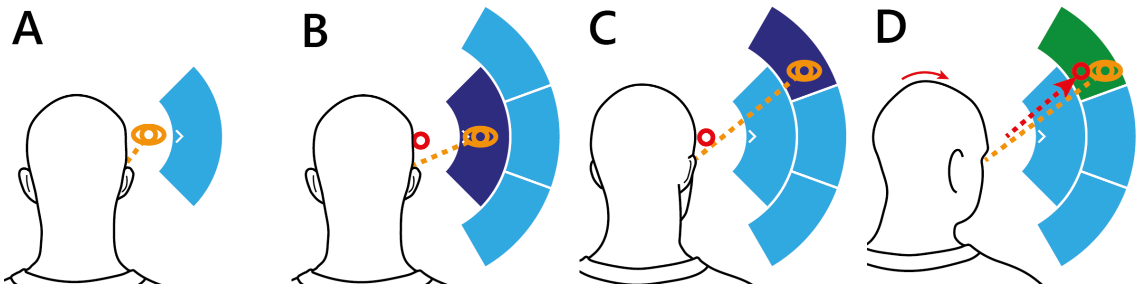

Figure 5.3: Radi-Eye nested interaction. A: The parent widget is idle. B: Gaze on the parent widget displays the nested widgets on an outer ring. C: Gaze moves to a nested widget, enabling selection. D: The head moves in a direct path to select the nested widget.

5.2.1 Pop-up Radial Interface

Radi-Eye has a radial structure composed of different widgets that are placed along the circumference of one or multiple rings (Figure 5.2a). The content of the interface (i.e. the widgets displayed) can be fixed or changeable to make additional functionality available. Also, ring widths can be alternated to account for eye tracking accuracy, and information to be displayed on the widgets. The radial structure of Radi-Eye is based on the eyes and head's ability to move in all directions when performing gaze shifts. Radi-Eye interaction is started from the inactive *centre zone* of the radial interface, from which the user can move their eyes and head in any direction for the interaction. This allows Radi-Eye to exploit proven advantages of the radial layout [22, 63].

The interface supports scaling in three ways to support large interfaces and increased functionality. First, the nested interaction is supported by expanding the interface on the outer rings (Figure 5.3). Hidden nested widgets are displayed when gazing on the parent widget, allowing users to search and traverse through a nested interface without committing to a selection. Users can then move their head in a direct path to a nested gazed-on widget for selection. Nested widgets are placed in a fan-like structure to avoid the cumbersome "criss-cross" gaze behaviour caused by widgets being at opposite ends of the interface [145]. Second, the interface supports replacing widgets on a ring by toggling a widget. Replacing widgets allows a single ring to store a large number of widgets without relying on multiple layers. Third, Radi-Eye supports the fluid composition of multiple commands (Figure 5.4), achieved by placing commands and their options on separate rings. The user can then select a command at each ring when traversing the rings. The user may even pass over a ring without performing a selection if a command is undesirable. Command composition allows efficient input of multiple parameters and the execution of advanced commands that require heterogeneous input.

The pop-up nature of Radi-Eye allows on-demand control of objects in the environment without causing clutter when not in use. Accounting for the user's position and head direction during invocation ensures that the interface can always be positioned in or close to the middle of the screen at the start of an interaction to ensure possible interaction in all directions [63]. Also, HMD-based interaction allows the space outside the field of view that is revealed during head movement for interface control.

The structure, position, and invocation of a Radi-Eye interface depend on the context of use. In its simplest form, the interface is used to directly manipulate a specific object in the environment. The characteristics of these *Object-dependent* Radi-Eye interfaces are dependent on the interacted object. The invocation method should single out a specific object in the environment (e.g. ray casting), the displayed widgets depend on the object's functionality, and the position and structure of the interface should be linked to the interacted object to allow an efficient feedback-control loop.

In certain situations, a user may want to interact with multiple objects simultaneously, or with objects that are not visible or at an uncomfortable head position. In such cases, an *Object-independent* Radi-Eye interface which is not linked to an object in the environment can be used for interaction. Object-independent invocation is performed independently of any specific object via a generic gesture or command. The interface position is not linked to a specific object but primarily considers user convenience, i.e. centred around the head. The available commands are not dependent on specific objects, and in cases of multiple interacted objects, new commands can arise from their common elements or any additional functionally arising from combining them.

As gaze is used for interaction, feedback that triggers visual attention should not disturb the flow of interaction. Feedback can thus be displayed in the centre zone, widgets, or outer rings depending on the context of use to guide the user toward the next steps of the interaction. Also, feedback should not be displayed too far from the head position (>30°) during interaction to ensure comfortable eye-in-head positions.

5.2.2 Look & Cross

Look & Cross is designed as a generic interaction technique for Radi-Eye, combining gaze for pre-selection of interface components with head-crossing for their invocation. The technique is based on the natural misalignment between the eyes and the head, and that gaze naturally precedes the head during a gaze shift [45, 151, 152]. It uses fast and effortless gaze to explore the interface, trigger hover interaction, and enable widgets for selection. The user then selects a widget by moving the stable head across the gaze-activated widget's boundary. Look & Cross thus supports a controlled three-step interaction process: from idle, to hover, and finally to selection (Figure 5.2b-d).

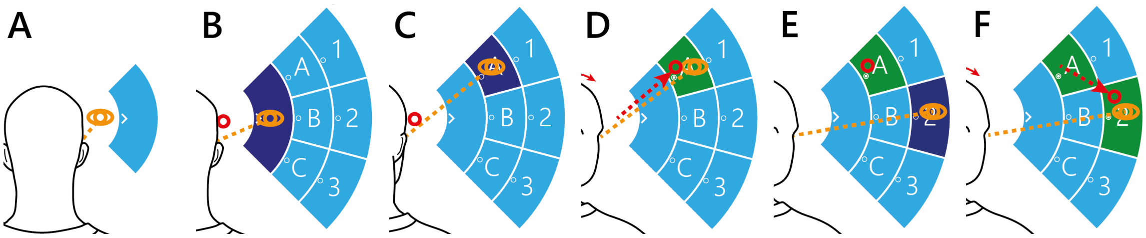

Figure 5.4: Radi-Eye command composition on multiple rings. A: The parent widget is idle, hiding the available commands. B: Gaze triggers the parent widget showing the nested commands displayed on separate rings. C-D: The user enables (C) and selects (D) the first command. E-F: The user moves on to enable (E) and select (F) the second command.

Users can dwell with their gaze on widgets without causing unintended input, which may be useful for cognitively demanding tasks where thorough consideration of choices can induce prolonged fixation.

Building on the natural offset of the eyes and head ensures robust selection as both modalities have to point on the same widget at the same time. The natural offset between the eyes and head allows users to "skip over" idle distractor targets with their head without performing selections. This is useful when performing a series of selections, such as selecting multiple widgets in a list. Also, users can shift their gaze and head outside the interface, allowing free exploration of the interface and the surrounding environment without triggering unwanted interactions.

Look & Cross is inspired by Eye&Head Convergence, a selection technique that uses eye and head alignment for selection [152]. However, the techniques differ in selection condition and handling of the eyes and head accidentally aligning over a target. Firstly, in Eye&Head Convergence, users have to move the head within an angular distance to the eyes, requiring a gaze cursor to display the selectable area, cluttering the selection space. Look & Cross defines the selection area as the border of the gazed-on widget and therefore requires no gaze cursor. Using the widget boundary is also beneficial for defining an area for subsequent continuous head interaction that is not dependent on the gaze position, or when multiple small targets are nearby (e.g. nested interface), as an angular area may overlap multiple targets, causing selection ambiguity. Secondly, Eye&Head Convergence starts a timer during which the eyes and head must remain within the angular threshold to confirm the selection if the head cursor is already within the convergence area when gaze first points at the target. In Look & Cross, we enforce the order of the interaction sequence; gaze has to precede the head on a target. Forcing the movement order ensures the hover interaction and the same interaction sequence for all selections. Also, the user does not have to worry about accidental selections caused by dwelling on head-pointed targets for too long.

Table 5.1: Widgets available with Radi-Eye.

Type	Discrete					Continuous		Scaling	
Widget	Simple selection	Toggle	Checkboxes	Radio button	Step increment	Slider	Hold down	Nested ring	Replace ring
Icon	None	☐☑	☐☑	○◉	− +	━●━	◷	›	≡‹

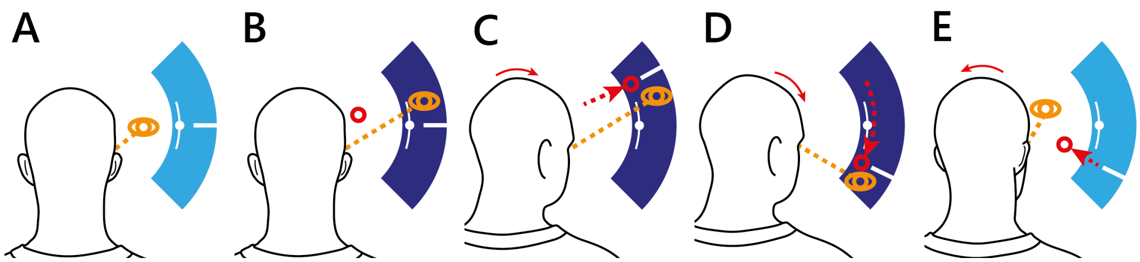

Figure 5.5: Radi-Eye slider interaction. A: The slider is idle. B: Gaze enables the slider for interaction. C: The head moves to trigger interaction. The slider handle jumps to the cursor position. D: The head moves along the slider's arc, changing the slider value. E: The head exits the slider, the slider handle stays at the last cursor position.

5.2.3 Radi-Eye Widgets

Combining the radial pop-up interface and Look & Cross provides the final Radi-Eye component, a broad set of widgets available for object control (Table 5.1).

5.2.3.1 Discrete selection

The Look & Cross interaction sequence supports a multitude of widgets using discrete selection, from basic selections to toggle widgets (see Table 5.1). Logically connected widgets, such as checkboxes and radio buttons, can be placed adjacent to each other along the ring to indicate their association with each other. Icons are placed on the widgets to indicate the type of widget and the required interaction before performing a head movement.

5.2.3.2 Continuous Interaction

Extending the interaction sequence of Look & Cross provides continuous interaction. To trigger a continuous interaction, the head cursor must remain within the widget boundary after the initial selection. The continuous interaction is then active until the head cursor leaves the widget boundary, similar to holding down a button. We can also transform the widget into a 1-dimensional slider along the ring arc by adding meaning to the cursor position (Figure 5.5). During continuous interaction, the user's gaze can move freely outside the widget to ensure that no strain caused by gaze being "trapped" within the widget.

5.3 Applications

We developed three applications in VR and AR environments to showcase the different faucets and highlight the design considerations of Radi-Eye: a VR media player, an AR smart home manager, and a VR city builder. We implemented all applications in Unity. We used the HTC Vive with an integrated Tobii eye tracker for the VR applications. For the AR application, we used a Zed mini see-through camera combined with an HTC Vive Pro Eye. Both HTC Vives have a vertical FOV of 110° and a horizontal FOV of 100°. However, the Zed mini only has a vertical FOV of 54° and a horizontal FOV of 85°. Note that the gaze position is not visible to the user in our applications. This decision was made so that any gaze cursor would not occlude widgets and the user's region of interest which may have a negative impact on user performance. However, all figures show the gaze position for illustrative purposes.

5.3.1 Media Player

Our first application is a VR media player designed to demonstrate some of the fundamental types of interaction. The user invokes the media player by dwelling their gaze on a button at the bottom of the television (TV). Once invoked, the interface is positioned centred around the TV and sized to encompass the TV inside the centre zone and support a large number of widgets (40°). Ring width was set to 5-7.5° to make the effect of eye tracking error negligible. Users can play/pause videos using a toggle interaction, toggle subtitles (Figure 5.6), and seek through the video via the timeline slider for coarse timestamp selection or reversing/forwarding the video using the "hold-down" reverse and forward buttons for further refinement (Figure 5.7). The menu is closed via a dedicated "close menu" button. To address space limitations, we use the *replace ring* widget to switch from playback control to video browsing by toggling. The users can then select videos to view alphabetically, displaying the corresponding videos in an outer layer (Figure 5.6c-f).

The media player exemplifies an *Object-dependent* interface as it is directly related to the controlled TV. We placed any continuous input in the innermost ring, the shortest distance to the central feedback (20°), to ensure a comfortable eye-in-head position while users look back at the TV to, for example, inspect the timestamp (Figure 5.7). On the contrary, the nested outer ring is used for its increased item real estate when selecting different media to display (Figure 5.6e). The application showcases a number of advantages of Radi-Eye:

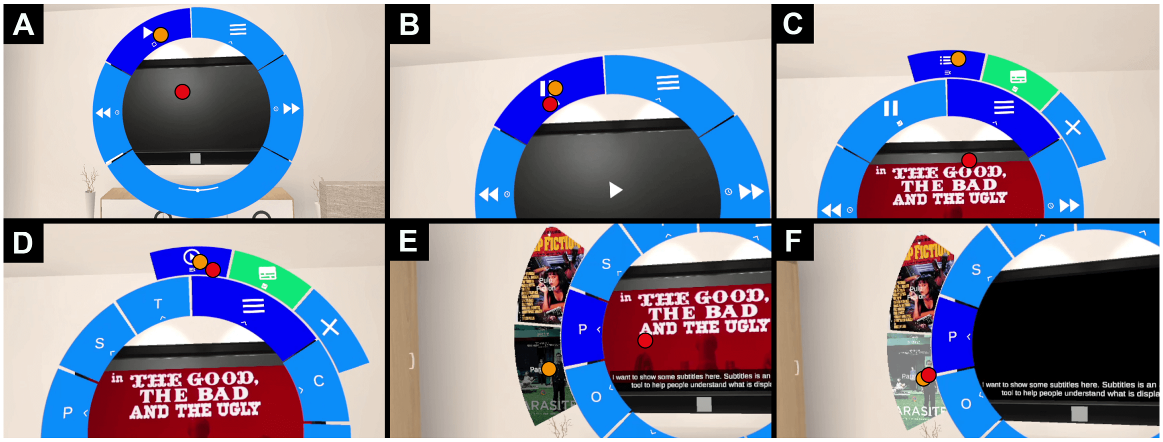

Figure 5.6: The user plays and changes a video. Gaze (orange) moves to the play button to enable selection (A) while the head (red) crosses the gazed-on button's boundary to start the video (B). The user gaze on the nested select media button (C) and selects it to replace the inner ring's playback controls with media selection controls (E). The user searches through available media (E) and selects a new video (F).

- Users can inspect widgets for as long as users need without risking unintended selection.
- Users can traverse and inspect nested items with gaze without committing a head movement until selection.
- Widgets can complement each other to support different levels of granularity (slider and reverse/forward buttons).
- The interface can be extended to support additional functionality (media selection) via ring replacement and a large number of options via nested interaction.
- Users are free to explore the interface or feedback while a continuous interaction is maintained. This is useful when observing feedback that is external to the widget, for example, seeking through a video.

5.3.2 Smart-Home

Our second application is an AR smart home manager where users control lighting and home appliances via Radi-Eye. The application supports both contextual and

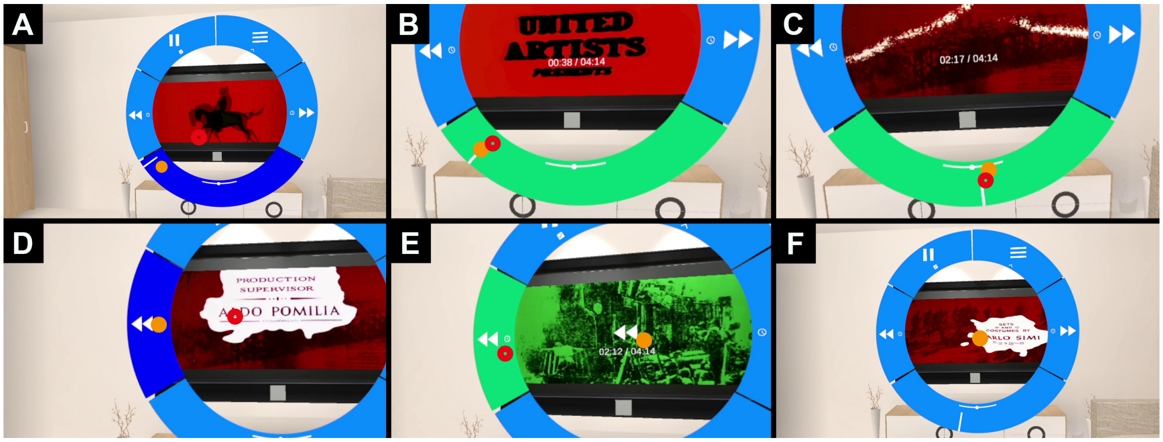

Figure 5.7: The user changes the video timestamp. Gaze enables the timeline-slider (A) and the head crosses the slider selecting the timestamp at the head position (B). The head moves along the timeline to fast-forward the video (C). The reverse button is then enabled the with gaze (D) and activated with the head (E) for further timestamp refinement. The head exits the reverse button to resume the video (F).

non-contextual interaction. Radi-Eye interfaces for individual appliance control are *Object-dependent*. Similar to the media player, the user performs an invocation via gaze dwell on the physical appliance. The interface controls depend on the selected appliance: selecting a kettle shows simple toggle controls for power (20° centre zone size, Figure 5.8c), while a lamp has more advanced controls for adjusting brightness and colour via sliders, thus requiring a slightly larger interface for increased continuous precision (30° centre zone size, Figure 5.8b). Individual appliance control requires line-of-sight for invocation, which can be problematic depending on a user's position in the room.

We developed an *Object-independent* "master-menu", which can be invoked from any position and enables users to control appliances en masse from one interface. The master-menu has a slightly larger centre zone (40°) to accommodate easy selection of all appliances in the outer ring. In this application, the user invokes the master-menu by performing a simple hand gesture in front of the HMD AR-camera (Figure 5.9a-b). Voice commands or head gestures can be used as hands-free alternatives. From the master-menu, users can toggle all available appliances, toggle lamps individually and en masse, and toggle lighting presets such as reading, day-light, or disco (Figure 5.9b). The application highlights multiple Radi-Eye advantages:

- Users can invoke an *Object-dependent* interface or an *Object-independent* interface depending on their needs.

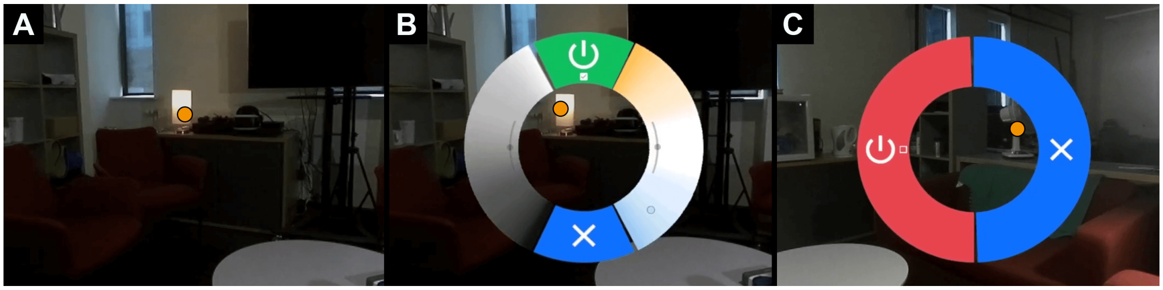

Figure 5.8: A: The user invokes an Radi-Eye interface for appliance control by gaze dwelling on an appliance. B-C: The object-dependent interface adapts to the selected appliance (B: lamp, C: fan) and is centred around the user's head.

- Interfaces can be adaptive to context and interacted objects.
- Through the master-menu users can control out-of-view objects without significant body movements.
- Users can traverse their gaze or head across other options when reaching a target, affording freedom in the choice of interface layout (Figure 5.9c-f).
- Radi-Eye can be used in both VR and AR and can be adapted to cluttered domestic or workplace settings.

5.3.3 City Builder

The final application is a VR city builder in which users can inspect, build, and edit a city using Radi-Eye. To inspect and edit a building, the user invokes the "Change-menu" by gaze dwelling on a plot with an existing building (Figure 5.10a-b). The selected building can then be demolished or adjusted by choosing a new rotation or colour.

The user places a new building by dwelling on an empty plot. This invokes a Radi-Eye interface where users can choose between commercial, industrial, residential, or public buildings that are available for the selected plot. The user can also select the rotation and colour of the building via command composition before confirming the building placement (Figure 5.10c-h). As command composition requires multiple layers depending on the number of performed actions, placing feedback in the centre zone (in this case the placed building) would cause it to be further away from the user the more commands they performed. This feedback placement can lead to issues when users want to observe feedback, having to gaze back to the centre zone, disrupting the command composition. Instead, we display feedback on the outermost

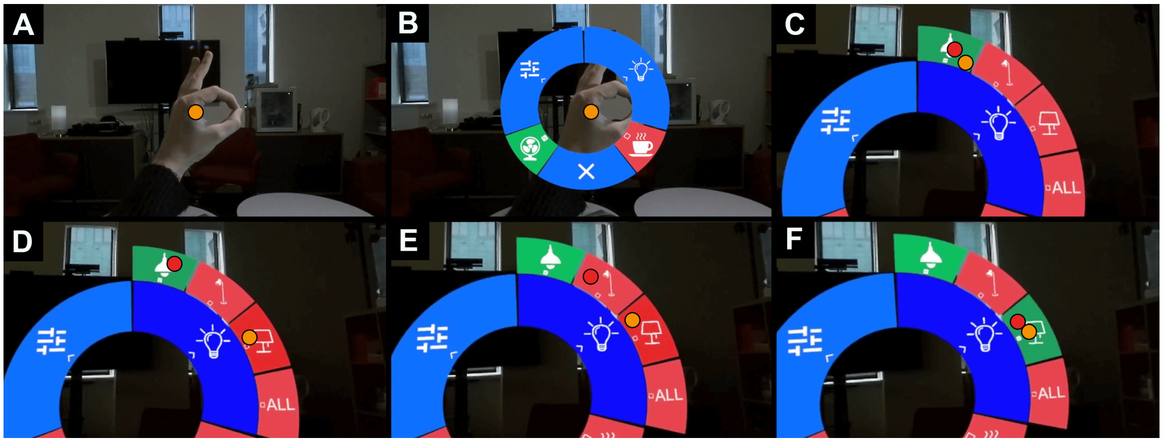

Figure 5.9: The user invokes and turns on multiple lamps via the master-menu. The user performs a hand gesture (A) to invoke the interface (B). The user then selects and turns on the first lamp (C). Gaze then moves to enable the second button (D). The head moves over a non-enabled button which does not trigger (E) to turn on the second lamp (F).

confirm button, the direction to which the head-gesture and gaze point are moving toward (Figure 5.10c-h). The centre zone is set to smaller (20°) to accommodate multiple layers, and ring widths are narrower (5°) to allow comfortable selection of the outermost layer (35° from the centre position). The application highlights a number of advantages of Radi-Eye:

- Radi-Eye supports complex object manipulation and rapid multiparameter input via command composition.
- Radi-Eye allows users to skip over commands during command composition to more rapidly perform an action.
- Radi-Eye supports users to change and omit selections during command composition before performing an action via the confirmation button.
- Radi-Eye supports a variety of feedback strategies that can be adopted on or off the interface, depending on the interaction.

5.4 User study

In a previous chapter, we have compared gaze-head alignment with dwell and found the technique to be faster for the selection of known targets and perceived as more

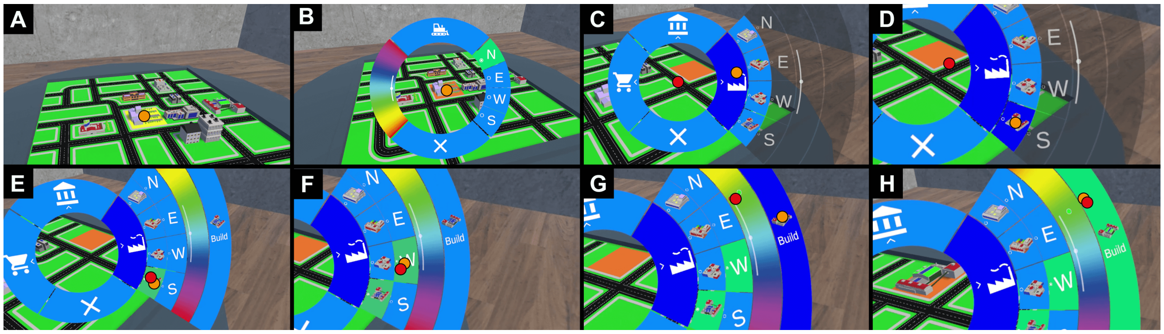

Figure 5.10: The user adjusts a building (A-B) and places a new building (C-H). Dwelling on a building invokes the change interface (A) which can adjust its rotation or colour, or demolish it (B). The user can invoke the build interface by dwelling on an empty plot. Gazing on a building type shows nested available buildings (C). The user then enables (D) and selects a building (E), enabling further commands. The selected building and its properties are shown on the outermost confirm button. The user selects a rotation (F) and colour (G) which updates the confirm button, and then selects the confirm button to place the building with the selected properties (H). Skipping steps F and G places a building without specifying rotation and colour.

natural and easier to use (Chapter 4). This established that users are effective and efficient with the technique. For Radi-Eye we therefore focussed not on comparison against other selection mechanisms, but the effect of radial interface parameters on Look & Cross performance. Some of the fundamental design choices relate to *ring size*, *widget amount*, and *widget direction*. We conducted a user study to investigate how these factors impact user performance and reception in a task in which participants had to select a particular widget out of many. For the sake of simplicity, we focused our study on basic selection.

In this study, we refer to the ring size as the centre zone size. We decided on a set ring width of 10° visual angle. Ring size impacts the required distance for selection, widget size, and also the occluded areas of the environment during interaction. We investigated ring sizes that cover a range from only requiring small head movements for selection to covering the whole HMD (S: 10, M: 35, L: 60, XL: 85° visual angle). The widget amount of a ring has an impact on widget size, interface complexity, and search time. We set the range of widget amount to 4, 8, and 16, limited by widget size. The widgets were of equal size. Finally, we vary the selection direction so that selections were performed in both the cardinal and diagonal directions. We used two layouts for the 4-widget condition to support both direction types.

Participants were tasked to select the correct widget from many as fast and

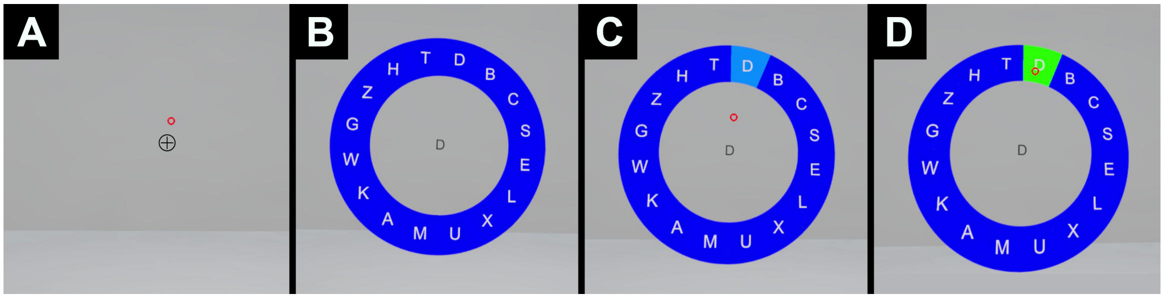

Figure 5.11: Example study trial. A: The participants start a trial by moving the head cursor and gaze into the central target. B-C: The participant has to find and select the letter "D". D: Colour feedback indicates selection success.

accurate as possible at varying ring sizes, widget amounts, and widget directions (Figure 5.11). To start a trial, participants aligned their gaze and head towards a central target. The trial started after a 300ms of alignment when a radial interface and a single letter appeared in the centre zone at 8 metres distance. Participants were tasked to find and select the widget containing the letter that matches with the centre letter. Gaze feedback was shown by widget colour. A head cursor appeared when the participants gazed at a widget. Participants moved the head cursor across the gazed-on widget to perform a selection. A correct or incorrect selection was shown via colour feedback. A trial was completed regardless of whether the selection was correct or incorrect. Participants would then realign their gaze and head to start the next trial.

5.4.1 Apparatus

We used an HTC Vive with an integrated Tobii eye tracker for the user study. The HMD has a FOV of 100° and a 90Hz framerate. Eye tracking data was recorded at 120Hz. The study environment was developed in Unity version 2017.4.3f1.

5.4.2 Procedure

We recruited 12 participants for the user study (Six male, six female 27.64 ±6.23). For previous VR experience, two participants reported no experience, nine reported occasional, and one reported daily. For previous eye tracker experience, three participants reported no experience, seven reported occasional, one reported weekly, and one reported daily. Participants first signed a consent form and answered a demographic questionnaire. Participants were then seated and put on the HMD. Participants performed an eye tracking calibration and a training session before each test session. Ring size order was counterbalanced with a Latin square. Widget amount and widget direction were randomised. After completing the task with a ring size, participants removed the HMD and answered a NASA TLX Workload questionnaire. A semi-structured interview was conducted at the end to extract preferences and opinions. In total each participant performed 4 ring sizes x 3 widget amounts x 16 repetitions = 192 selections. Note that for the 4-widget condition, half of the selections used the cardinal layout and the other half the diagonal layout. The study took 30-40 minutes to complete.

5.4.3 Results

Our five dependent variables were search, selection and total time, error rate, and workload. Unless stated otherwise, the analysis was performed via a Ring Size $\times$ Widget Amount two-way repeated-measures ANOVA (4×3), Greenhouse-Geiser-corrected in cases where Mauchly's test indicated a violation of sphericity, and with Bonferroni-corrected post hoc tests where applicable. Effect sizes are reported as partial eta squared (η_p^2). Time Shapiro-Wilks tests and Q-Q plots validated the assumption of normality. Only successful trials were used for analysis.

5.4.3.1 Search time

We defined search time as the time from the interface appears to the participant gaze on the matching widget. No interaction was found between ring size and widget amount ($F_{6,66}$=1.33, p>.05, η_p^2=.11). However, results showed a significant ($F_{2,22}$=279.63, p<.001, η_p^2=.96) increase in search time with increasing widget amount (Figure 5.12a). Post hoc analysis showed significant differences between all conditions (all p<.001). We also found a significant main effect on ring size ($F_{3,33}$=17.82, p<.001, η_p^2=.62). Further tests showed that XL had significantly higher search time than other ring sizes (all $p < .008$), indicating that ring size increases search time if users have to rely on head-movement for search. Both ring size and widget amount should be considered to reduce search time. Widget layout showed no significance for the 4-widget condition ($F_{1,11}$=1.22, p>.05, η_p^2=.10).

5.4.3.2 Selection Time

We defined selection time as the time from that the participant gaze on the matching widget until a selection was made. Participants were able to select widgets at 2 seconds or less for all conditions (Figure 5.12b). We found a significant two-way interaction between ring size and widget amount ($F_{1.70,18.72}$=15.01, p<.001, η_p^2=.58). Further analysis showed that Ring size had a simple main effect at 4 ($F_{3,33}$=15.68, p<.001, η_p^2=.59), 8 ($F_{1.70,18.64}$=17.22, p<.001, η_p^2=.61), and 16 widget amounts

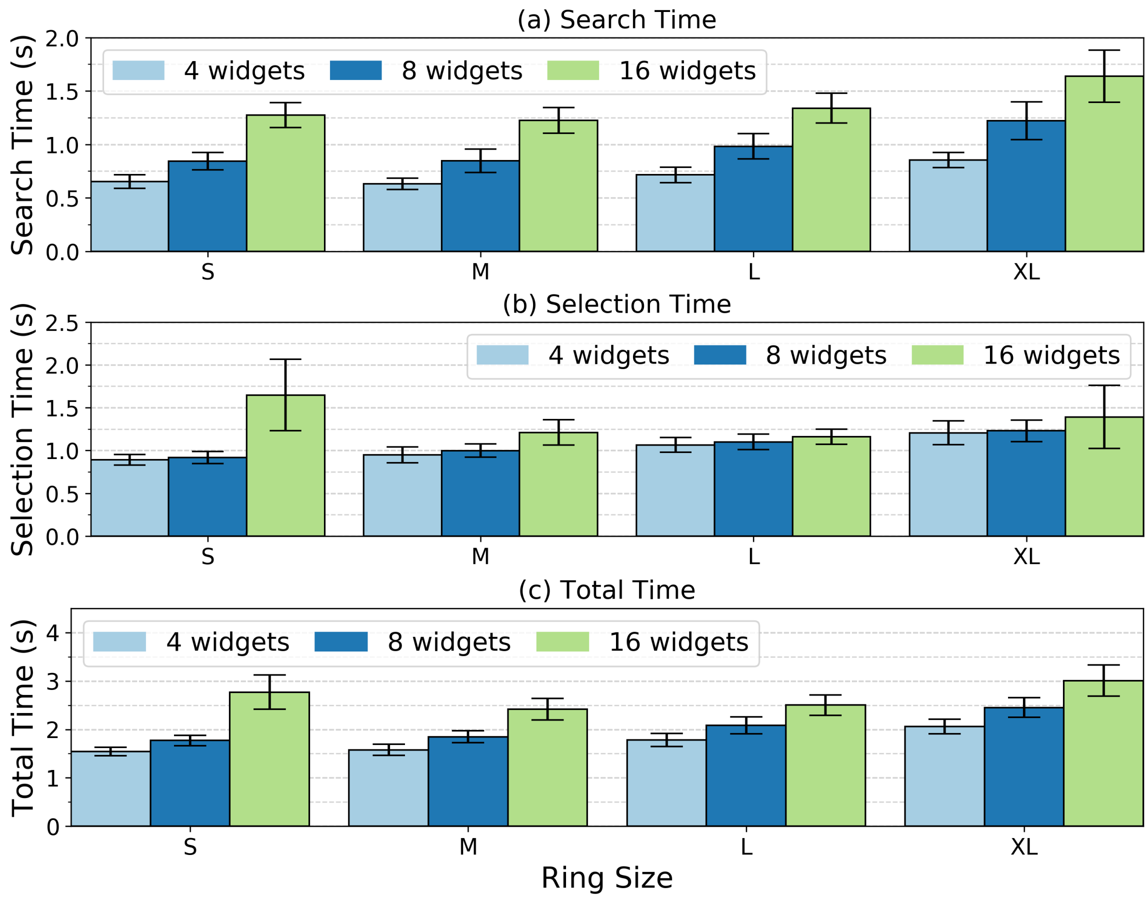

Figure 5.12: Mean search, selection and total time. Error bars represents mean 95% confidence interval.

($F_{1.67,18.33}$=5.21, p<.05, η_p^2=.32). Similarly, widget amount had a simple main effect at ring sizes S ($F_{1.09,11.95}$=30.79, p<.001, η_p^2=.74), M ($F_{2,22}$=22.07, p<.001, η_p^2=.67), and L ($F_{2,22}$=8.83, p<.05, η_p^2=.45). However, no significant main effect was found for ring size XL ($F_{2,22}$=1.60, p>.05, η_p^2=.13). A larger ring size led to a higher selection time as larger head-movements are required for selection. Also, as the combination of ring size and widget amount decides the widgets' arc size, a large widget amount combined with a small ring size can thus lead to slower selections as head movements have to be more precise. But the effect of widget amount becomes negligible if the ring size is large enough. Widget layout at the 4-widget condition also showed a significant effect on selection time ($F_{1,11}$=16.50, p<.05, , η_p^2=.60) where the cardinal layout was slightly but significantly faster than the diagonal layout (0.05s).

5.4.3.3 Total time

We defined total time as the time between the moment a ring is presented until the moment a selection was made which includes both search and selection time. We found a significant interaction between ring size and widget amount (Figure 5.12c, $F_{1.82,20.26}$=6.32, p<.05, η_p^2=.37). Further investigation into simple main effects showed that ring size had an effect at 4 ($F_{3,33}$=68.85, p<.001, η_p^2=.86), 8 ($F_{1.76,19.40}$=34.00, p<.001, η_p^2=.76), and 16 widget amounts ($F_{1.61,17.70}$=4.61, p<.05, η_p^2=.30). We also found simple main effects for widget amount on S ($F_{1.04,11.42}$=35.57, p<.001, η_p^2=.76), M ($F_{1.38,15.14}$=77.79, p<.001, η_p^2=.88), L ($F_{2,22}$=89.15, p<.001, η_p^2=.89), and XL ($F_{2,22}$=80.77, p<.001, η_p^2=.88). Similar to the selection time results, an increase in widget amount and ring size lead to an increase in total time. A larger ring size lead to larger distances for the head to travel, while a high widget amount combined with a small ring size can lead to slower selections due to an increase in required precision. We also found that widget layout had a small but significant influence on total time ($F_{1,11}$=12.34, p<.05, η_p^2=.53) where the cardinal layout was slightly but significantly faster than the diagonal layout (0.05s).

5.4.3.4 Error Rate

We define an error as the percentage of erroneous selections among all selections. Participants performed a low amount of erroneous selections. In total, only 39 out of 2304 (1.7%) selections were erroneous. Figure 5.13a highlights how the error rate is affected by a combination of ring size and widget amount, which decides the widgets' visual angle arc size. The results showed an increase in error rate for widgets with an arc size smaller than 2°.

5.4.3.5 Workload

Friedman test on the overall workload from the NASA TLX questionnaire (Figure 5.13b) showed that ring size had a significant effect ($\chi^2(3)$=10.66, p<.05). Further Bonferroni corrected Wilcoxon comparisons showed that the ring size XL had significantly more workload than M (p<.05) and L (p<.05). Friedman tests on the weighted averages of each sub-scale showed significant differences in Physical Demand ($\chi^2(3)$=12.69, p<.05). Bonferroni corrected Wilcoxon analysis showed that ring size S had significantly lower Physical Demand than ring size XL (p<.05).

5.4.3.6 Preference

Look & Cross was considered to be *"easy to use"* (P7) and *"intuitive"* (P8). Participants also mentioned that it was *"natural to use"* (P2) and the use of both modalities *"helps preventing false selections"* (P9). When asked about their preferred

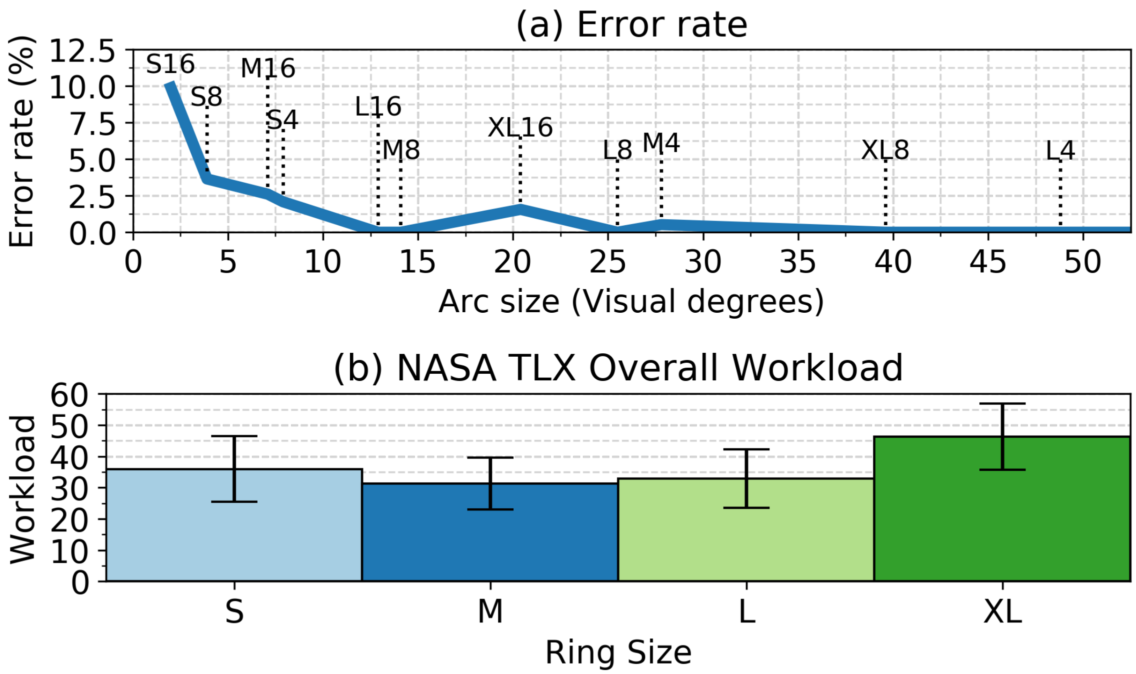

Figure 5.13: A: Error rate as a function of arc size. Labels indicate study conditions. B: Average overall workload. Error bars represents mean 95% confidence interval.

ring size, 8 participants stated the ring size M as their most preferred. Three participants stated that sizes M and L were equally preferred. Lastly, one participant preferred the smallest size S due to its *"quickness"* (P9). Ring sizes S and XL were disfavoured for various reasons. Participants had trouble with size S due to *"being unable to control the eyes enough to stay on the buttons"* (P12) or because it *"restricted head movement"* (P7). The ring size XL was disliked due to *"too much head movement"* (P5) and *"long search time"* (P4). Finally, all participants mentioned either horizontal or vertical directions as their most preferred selection direction.

5.5 Discussion

Radi-Eye is effective for hands-free selection and control of objects and affords a wide variety of widgets for discrete and continuous input in any 3d environment. The pop-up nature of Radi-Eye allows convenient and on-demand control of the surrounding environment (both visible and occluded) from any user position. The ability to control objects outside the FOV and the reliance on only eye and head movements is significant and highly relevant for contexts with constrained user motion input (e.g. seated usage).

Radi-Eye is versatile and lends itself to implementation in different configurations of invocation method, widgets, and interface structure as showcased by our three applications. Radi-Eye can control single or multiple objects to the users' discretion as demonstrated in the smart home application. Widgets can complement each other for verbose and refined interactions as highlighted in the media player which uses a slider for initial selection and the hold-down widget for further refinement of the timestamp (Figure 5.7). We can also combine widgets into a fluid series of commands via command composition for efficient and expressive interaction, as shown when placing a building (Figure 5.10). Both the applications and user study points toward design aspects to consider when designing a Radi-Eye interface.

The key Radi-Eye design consideration that encompasses all design aspects of the interface are the natural behaviours of the eyes, head, and their coordination. Feedback supports and guides the user through the interaction rather than disturbing the interaction flow. For example, feedback is always shown on the outer rings guiding the user towards the next steps of the interaction when placing a building in the city builder (Figure 5.10). Also, the interface design should keep users within a comfortable eye-in-head position. In the media player, feedback can be safely displayed in the centre zone as the distance between the centre and the widget positions are small (Figure 5.7), while feedback displayed in the centre zone for the build menu could potentially force users to move outside their comfortable eye-in-head range (Figure 5.10). All our application interfaces were positioned centrally or approximately central in front of the user's head when invoked to allow gaze exploration in all directions.

In addition to the eye-in-head position, neck ergonomics is an important factor in the Radi-Eye interface design. For example, the number of layers in the interface should be considered to avoid neck strain. A small centre zone may be appropriate for large interfaces (city builder) to limit the use of head movement while a shallow menu can use a larger centre zone without causing discomfort. Furthermore, object-independent interfaces can be used if the interacted object is at an uncomfortable neck position (e.g. a roof lamp). Future work could investigate interfaces that adapt to the current neck position and adjust to ensure neck comfort regardless of the performed interaction. Examples include oval interfaces that are closer to the starting position in an uncomfortable direction, or interfaces that adapt the widgets so that multi-layered widgets are placed along the direction with the widest range.

The study results showed that users can quickly and accurately select Radi-Eye widgets. However, the combination of ring size and the number of widgets on a ring has to be taken into account as they define widgets' arc sizes. The combination of a small ring size and a high number of widgets thus increases the reliance on refined movement and the effect of eye tracking error. Fewer widgets on a single ring allow small ring sizes which reduces selection time and extraneous head movement, while also reducing search time and the need for precise head movements for selection. However, if a large number of widgets is necessary for object control, offloading widgets to outer rings via nested interaction or via the replace ring feature as shown in the media player allows easy selection (Figure 5.6). Finally, placing frequently used widgets along the cardinal axes allows faster and more comfortable selection, as shown in the study results.

Look & Cross builds on fundamental eye-head coordination insights by utilising a head-crossing metaphor that caters to a user's natural sequence of gaze shifts. Our results show that the technique is intuitive and easy to learn as users only have to add a small extra head movement for a natural gaze shift to turn into a selection. However, the offset between the eyes and head is large enough during exploration so that users can safely explore the interface or environment without triggering accidental selections. This capability allows users to easily avoid selection of unwanted targets which is useful for the composition of commands (Figure 5.10) or selection of multiple widgets in a list (Figure 5.9c-f).

At the core of Look & Cross lies the distribution of different parts of the interaction between the eyes and the head. This allows the user to utilise the relative strengths of the modalities. The eyes enable quick and effortless search and hover interaction without risking accidental selections. The stable head can then be used for precise selection confirmation. In our work, Look & Cross was combined with a radial interface to extend the available interactions as highlighted by the diverse set of widgets used in our applications. However, Look & Cross is not limited to the usage in radial interfaces and can be extended as a general technique for object selection and manipulation in 3D or 2D environments.

While eye-head alignment was accurate for selection in our user study and previous work [152], it has not been evaluated for continuous or nested interactions. Similarly, our user study primarily focussed on fundamental menu properties and their impact on selection performance. Future work could evaluate the use of gaze and head for continuous or nested interaction to discover limitations. In the media player, we mitigated the effect of limited head precision by using widgets to complement each other (Figure 5.7). Alternatively, adjusting the control-display ratio has been effective for head-based selection refinement and could increase precision during slider interaction [87]. In nested interaction and command composition, we introduced a delay (0.4s) before hiding nested items to avoid interaction interruption caused by eye tracker jitter or unexpected saccades moving outside widgets. Also, the head was used as an "anchor" to keep nested items open if it hovered over an item in the hierarchy while the eyes moved away. Finally, we carefully placed feedback to ensure that the gaze would not wander outside the widget hierarchy and hide the nested widgets.

We can also extend the Look & Cross concept to other modalities. In this work, we use the synergetic relationship of the eyes and head for interaction, and similar relationships exist between other modalities. For example, the eyes and hands are highly connected as we use our eyes to guide our hands, and previous research has leveraged this coordinated relationship for interaction in 3D environments [129, 131, 154]. Therefore, we could imagine a crossing-based technique that uses gaze for object activation and the hands for crossing. The combination of different modalities offers an exciting future research direction, where the choice of modalities will have a significant effect on the technique's capabilities. Finally, future work can also further extend Radi-Eye by developing new widgets to increase interface expressiveness or by extending existing widgets via, for example, having nested widgets expand inwards to minimise head movement or adjusting the widget functionality to be dependent on the crossing direction.

5.6 Conclusion

In this chapter, we introduced Radi-Eye, a novel radial interface for on-demand and hands-free object control via gaze and head input that we validated in application prototypes and a user study. The reliance on only gaze- and head-movements for input is useful in situations where the hands are unavailable and highly relevant for contexts where the hands and body are limited in movement. Furthermore, combining gaze and head for interaction extends their capabilities to allow effortless interface exploration and hover interaction, and provides users with more stable selection, feedback, and alleviates Midas Touch issues to support freedom to roam the interface and the surrounding environment with gaze and head movements without compromising interaction efficiency.

The radial interface is efficient in supporting gaze- and head-based interaction, and together with Look & Cross provide users with a wide variety of widgets for discrete and continuous interaction, and heterogeneous input via command composition for expressive hands-free control. However, the choice of interface layout, dimensions, and feedback can have a significant impact on user performance and user experience of the interface. Careful consideration of these factors have to be made to support easy interaction that does not disturb natural gaze or head behaviours. Furthermore, the Radi-Eye pop-up design supports control of both visible and occluded objects, and supporting users with object-dependent and object-independent interfaces is important to allow comfortable object control in any interaction context without having to perform significant body movements.

Radi-Eye shows how our introduced design principles can be used not only for simple selection, but also to provide more expressiveness for continuous and nested interaction. The design principles allow designers to develop new techniques based on eye-head coordination to tackle our defined gaze interaction challenges such as Midas Touch and lack of expressive input in novel ways. In the next chapter, we introduce our final design principle of *differentiating between natural and gestural head movements* and show how differentiation can be used to address eye tracking accuracy difficulties.

Chapter 6

BimodalGaze: Seamlessly Refined Pointing with Gaze and Filtered Gestural Head Movement

In previous chapters, we showed how eye-head coordination can be leveraged to address Midas Touch issues by classifying gaze shifts as eyes-only or head-supported and through head-eye alignment. We further show how head-eye alignment can be leveraged to enable expressive interaction in a radial interface. In this chapter, we address our final gaze interaction challenge of gaze accuracy and calibration challenges with eye and head coordination. We introduce BimodalGaze, a novel technique for seamless head-based refinement of a gaze cursor. The technique leverages eye-head coordination insights to separate *natural* head movement related to the eye-head coordination of gaze shifts, and head movement that we adopt as *gestural* as it is independent from gaze. As illustrated in Figure 6.1, users point primarily with gaze (A) but can seamlessly transition to refine the cursor position (B). The *Head Mode* for refinement is automatically invoked when gestural head movement is detected, while users are free to use natural head movement in *Gaze Mode* (C). This allows users to quickly shift their gaze to targets over larger fields of view with naturally combined eye-head movement, and to refine the cursor position with gestural head movement. In contrast to an existing baseline, head refinement is invoked automatically and only if a target is not already acquired by the initial gaze shift. We compared our technique with *Eye+Head Pinpointing* as a recent baseline for head-assisted gaze pointing [87]. While users were able to reliably achieve fine-grained selection we observed a comparatively higher rate of initial input error that affected selection time. We provide a detailed analysis of participant behaviour with our technique that gives insight into different causes for error and how they relate to design choices.

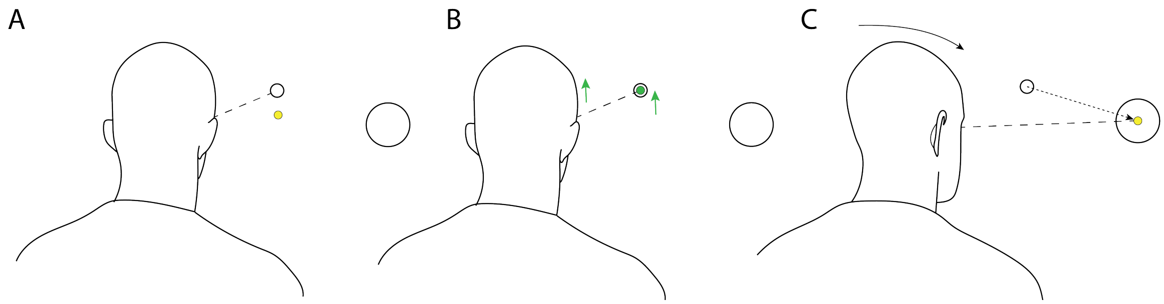

Figure 6.1: BimodalGaze enables users to point by gaze and to seamlessly refine the cursor position with head movement. A: In *Gaze Mode*, the cursor (yellow) follows where the user looks but may not be sufficiently accurate. B: The pointer automatically switches into *Head Mode* (green) when *gestural head movement* is detected. C: The pointer automatically switches back into Gaze Mode when the user redirects their attention. Note that the Head Mode is only invoked when needed for adjustment of the cursor. Any *natural head movement* associated with a gaze shift is filtered and does not cause a mode switch.

6.1 Gaze Refinement

A variety of work has addressed gaze inaccuracy by resorting to another modality for the refinement of gaze input. Head pointing, in particular, is interesting for complementing gaze, as head movement affords stable and precise input while retaining the advantage of hands-free pointing [15]. The switch from gaze to the second modality is usually performed by the user. In the AR technique *Pinpointing* by Kytö et al. [87], the user presses and holds down a button to switch modality and selection is then performed when the button is released. Similarly, the work on the touch-refinement of gaze treated every touch-down event as a trigger for switching modalities and release as a selection [130, 166]. These techniques are challenging for head-based refinement as there is no inherent confirmation mechanism (i.e. click) that can be leveraged for modality switching. Alternately, MAGIC pointing used mouse movements to trigger gaze refinement via the mouse [202]. Switching based on movement is possible since gaze and hand movements can be performed independently of each other. However, this is challenging for head-based refinement as head movements are performed during gaze shifts (Chapter 3).

Head correction of gaze has been demonstrated on displays with narrow-field-of-view (FOV), where gaze shifts were assumed to be performed by the eyes alone, thus allowing head movement to be treated as independent input for relative cursor displacement [76, 86, 87, 160]. However, eye movement research has shown that only small gaze shifts are performed solely with eye movement, whereas more significant shifts naturally feature head movement to reach targets and maintain a comfortable eye-in-head position [46, 89]. It is therefore not straightforward to use head movement for gaze correction, in particular when gaze is considered for pointing across a larger FOV, such as on large displays, across devices, or in XR. BimodalGaze overcomes the need for a manual switch and the risk of accidental modalitiy switching caused by gaze shift by classifying head movements as gesture (independent of gaze) and natural (part of a gaze shift).

Algorithm 1 BimodaGaze algorithm to decide the current pointing mode (*mode*). The algorithm compares the time between gaze and head movement (t), the angular difference between the trajectories of the latest head and gaze movements (d_{GH}) and the angular difference between gaze and pointing directions (d_{GP}) against set thresholds t_l, θ_{GH}, θ_{GP}, respectively. G_t and H_t are set thresholds to define gaze movements and head movements, respectively.

$v_{gaze} \leftarrow$ current gaze velocity
$v_{head} \leftarrow$ current head velocity
$traj_{gaze} \leftarrow$ trajectory of last gaze movement ($v_{gaze} \geq G_t$)
$traj_{head} \leftarrow$ trajectory of last head movement ($v_{head} \geq H_t$)
$d_{GP} \leftarrow Angle(dir_{gaze}, dir_{pointer})$
if $mode = Gaze$ **and** $v_{head} \geq H_t$ **then**
  $t = time_{now} - time_{gaze}$
  $d_{GH} = Angle(traj_{gaze}, traj_{head})$
  if $t \leq t_l$ **and** $d_{GH} < \theta_{GH}$ **then**
    $mode \leftarrow Head$
  end if
end if
if $v_{gaze} \geq G_t$ **or** $d_{GP} > \theta_{GP}$ **then**
  $mode \leftarrow Gaze$
  if $v_{gaze} \geq G_t$ **then**
    $time_{gaze} \leftarrow time_{now}$
  end if
end if

6.2 BimodalGaze

BimodalGaze uses gaze as the primary modality for quick, effortless and hands-free pointer control. When gaze is not accurate or stable enough for selection,

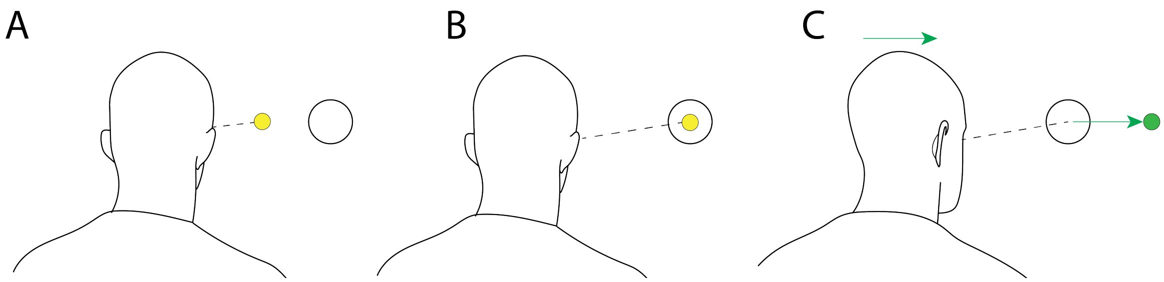

Figure 6.2: Typical situation without head movement filtering. A: The user use a gaze cursor (yellow) for pointing. B: The user moves their gaze and pointer onto the target. C: The user's gaze stays on target while performing a head movement as a natural part of the gaze shift. The pointer thus switches to head pointing (green) which drags the cursor away from the target during the natural head movement.

BimodalGaze allows automatic switching to head pointing when users perform deliberate head movements for further pointer refinement.

BimodalGaze leverages eye-head coordination insights to determine if a head movement is *gestural* for refinement; or a *natural* movement during a gaze shift. When a head movement is classified as natural, it is ignored and the user's gaze controls the pointer. When the movement is classified as gestural, BimodalGaze switches to head pointing for further pointer refinement. We define two criteria that both must be satisfied to differentiate natural from gestural head movements (Algorithm 1):

1. A natural head movement starts t_l ms after an eye movement.

2. A natural head movement will move in a similar direction to the prior eye movement (θ_{GH}).

The first criterion relates to the timing of head movements. Natural head movements are used to further the range of the eyes or to move the eyes into a comfortable position [181]. Research has shown natural head movement occurs at the same time as, or a shortly after, the initial eye movement to maintain a comfortable eye-in-head position [151]. If the head does not move within a certain time (t_l) after an eye movement we can assume that the head movement is gestural. Secondly, as the purpose of natural head movements is to increase the eyes' reach, or to move the eyes closer to their central position, it is reasonable to believe that a natural head movement will move in a similar direction as the eyes. If the angular difference between the trajectory of the eyes and head (d_{GH}) are within a certain range (θ_{GH}), we assume the head movement to be natural.

Switching to Head Mode is suppressed during a deliberate gaze movement, which we define as any gaze movement over a velocity of G_t. Alternatively, the system will switch back to Gaze Mode if the distance between the gaze and pointer position, d_{GP}, is greater than θ_{GP} (measured in visual angle). The latter condition is to prevent the cursor becoming too detached from the gaze due to misclassification of head movement (e.g. a natural movement detected as gestural Figure 6.2), or due to eye movements which are not detected as a deliberate gaze movement.

6.2.1 Implementation

We implemented BimodalGaze in VR. The selection of thresholds has a large impact on BimodalGaze's behaviour and requires careful consideration. For saccade detection, G_t was set to $160°/sec$ which is a relatively high value to avoid unintentional switches to Gaze Mode during refinement caused by corrective saccades, vestibulo-ocular reflex, or smooth pursuit eye movements. The sensitivity of head movement detection, H_t, was set to $15°/sec$, to be low enough to detect smaller head movements but high enough to ignore minor unintentional head shifts caused by the user. This value was originally based on the definition of head movement in Chapter 3 ($20°/sec$, however, we lowered the value as we noticed in informal tests that small head movements were not detected for during small pointing refinements. We based t_l and θ_{GH} on prior work on eye-head coordination in VR. Sidenmark and Gellersen found that head movements generally start moving 150ms after the eye movement (Chapter 3), therefore we set t_l to 150 ms. We used the data from Sidenmark and Gellersen's work to calculate the angular difference between the eye and head trajectory (Figure 6.3). We found that for 90% of all gaze shifts with accompanying head movement, the eye and head trajectory were within 20° of each other. As such, we set θ_{GH} to 20°. Finally, we set θ_{GP} to 10°, so that users can freely adjust the pointer position, while not being able to move the pointer too far out in the periphery.

6.3 Evaluation

We conducted a user study in VR to evaluate BimodalGaze and gather insights regarding its performance and user feedback. The flexibility of BimodalGaze raises the question of how often users transition to Head Mode, and how often selections can be made with gaze only under different conditions. We also want to assess how effective the automatic switching of BimodalGaze is, by comparing it with the Eye+Head Pinpointing technique, where the switch from gaze pointing to refinement via head movement is done manually by the user via a button click [87].

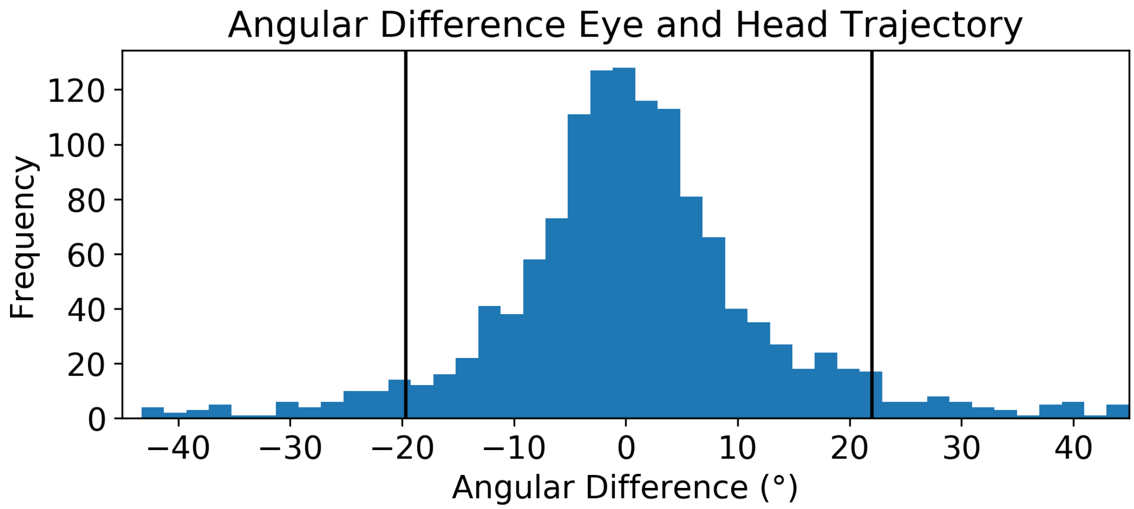

Figure 6.3: Angular difference between eye and head trajectory during a gaze shift. The zone within the lines represent 90% of all gaze shifts with accompanied head movement.

6.3.1 Task

Participants were required to select spherical targets with a diameter of 3° at 2m distance. From a central starting position, targets were found in one of eight directions (cardinal and intercardinal) at three target distances (10, 25, 40°). Smaller target distances are within reach of gaze and do not necessitate head movement, whereas we would expect users to move their heads naturally towards the larger targets. The accuracy of the eye tracker could dictate how much the techniques rely on Head Mode. We investigated this effect by artificially inducing an offset into the eye tracker's gaze estimation to simulate different levels of eye tracker accuracy. For each trial, we sample a random value from one of three normal distributions that represents gaze error amplitude of the trial. The accuracy distributions were varied to cover three scenarios. First, minor errors which offset the gaze pointing position but the pointing position will likely remain inside the target (Acc. 1). Refinements may be necessary for minor adjustments. Second, medium errors where refinement is likely needed but not for every trial (Acc. 2). Third, significant errors where the gaze position is guaranteed to lie outside the target and large refinements are necessary (Acc. 3). The error was set to max 5° as a larger error would mean difficulties in seeing the cursor far out in the periphery and would represent a case where an eye tracker would be considered unusable. The direction of the induced eye tracking accuracy error was randomly selected for each trial. The sampled amplitude and direction are translated

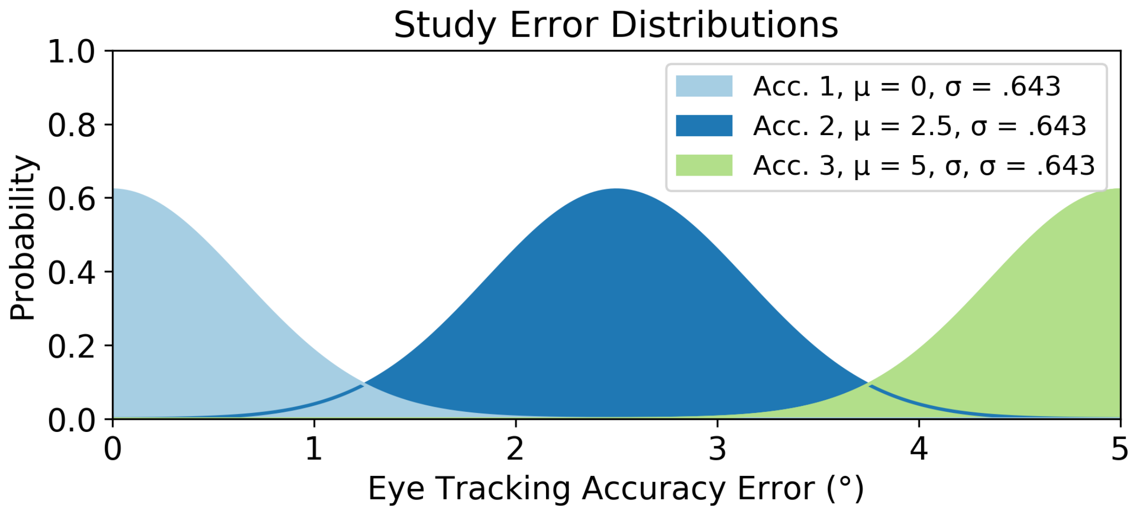

Figure 6.4: The different accuracy error distributions used for each accuracy condition. Note that the minimum value of Acc. 1 and maximum value of Acc. 3 was restricted to 0° and 5° respectively.

into a rotation which is then added to the gaze direction as a rotational offset to represent a constant error for the whole trial. A new offset direction and amplitude are sampled at the start of the next trial.

To begin a trial, the participant had to align a visible cross-shaped head pointer and their gaze with a central starting target (2° diameter). After 500ms, a spherical target appeared at one of the 24 predefined positions, chosen in random order. Participants were instructed to select the target as precisely and quickly as possible.

The user study employed a within-subjects design, with independent variables and levels as follows:

- *Technique:* BimodalGaze, Eye+Head Pinpointing
- *Gaze estimation accuracy:* Acc. 1, Acc. 2, Acc. 3
- *Target direction:* Up, Down, Right, Left, Up-right, Up-left, Down-right, Down-left
- *Target distance:* 10, 25, 40°

For each technique, each participant completed 3 blocks (one for each gaze estimation accuracy) of 72 trials (8 directions x 3 distances x 3 repetitions). Half of participants performed Pinpointing first followed by BimodalGaze, and the other half performed the reverse. The total number of trials per participant was 2 Techniques x 3 Blocks x 72 Trials = 432.

6.3.2 Apparatus

The techniques and task were developed in Unity version 2017.4.3. An HTC Vive with an integrated Tobii Pro Eye Tracker and data output frequency of 120Hz was used to record eye and head movement. We used the directional vectors of the eyes and head to calculate movements. We were able to record data at a mean gaze accuracy of $0.981 \pm 0.232°$ and a mean gaze precision of $0.427 \pm 0.152°$ calculated through a calibration procedure at the start of each block where participants fixated on 9 calibration points in a square arrangement. The standard hand-held controller of the HTC Vive was used for manual input. In BimodalGaze, a cursor was always visible to show the current pointing position. The cursor colour was used to show if the pointer was currently in Gaze Mode (yellow) or Head Mode (green). A selection was made by pressing the hand-controller trackpad. As described by Kytö et al., the Head Mode of the Pinpointing technique was triggered by pressing and holding down the hand-controller trackpad and a selection was made by releasing the trackpad. The cursor was only visible during the refinement stage for Pinpointing [87].

6.3.3 Procedure

We recruited 12 participants (8 male, 4 female, age: 25.42 ± 2.91) for our user study. Ten participants reported none or occasional VR experience, while two participants reported weekly VR experience. Eleven participants reported none or occasional eye tracking experience, while one reported daily eye tracking experience. Participants first signed a consent form and answered a demographic questionnaire. The participants were then seated and put on the HMD and handed the controller. The user study consisted of six test sessions, where the participants performed the task with one technique and one accuracy condition per session. The participants performed a five-point eye tracking calibration before each test session. The order of technique and accuracy conditions was counterbalanced with a Latin square. After each test session, participants removed the HMD and filled out a post-task questionnaire consisting of seven 5-point Likert items based on common usability factors (Precision, Ease, Learnability, Concentration, Physical effort, Frustration, Accurate switching), and were offered the opportunity to rest. A semi-structured interview was conducted after each completed task. The study took 45 minutes to complete.

Table 6.1: Prevalence of BimodalGaze selections and errors made in Gaze Mode based on accuracy error and target distance.

(a) Gaze Mode Selections

	10°	25°	40°	Total
Acc. 1	53.1%	33.7%	17.5%	34.7%
Acc. 2	19.1%	14.9%	11.8%	15.3%
Acc. 3	1.7%	2.4%	3.5%	2.5%
Total	24.7%	17.0%	10.9%	17.5%

(b) Gaze Mode Errors

	10°	25°	40°	Total
Acc. 1	65.2%	58.5%	50.0%	58.5%
Acc. 2	61.8%	51.4%	68.8%	60.4%
Acc. 3	26.1%	41.9%	50.0%	40.2%
Total	55.3%	52.1%	56.3%	54.4%

6.3.4 Analysis

For each trial we measured the completion time, incorrect selections, time spent in Head Mode and total head movement. Unless stated otherwise, we conducted a four-way repeated-measures ANOVA ($\alpha = .05$) for performance metrics with *Technique*, *Gaze estimation accuracy*, *Target direction* and *Target distance* ($2 \times 3 \times 8 \times 3$) as independent variables. When the assumption of sphericity was violated (tested with Mauchly's test), we used Greenhouse-Geisser corrected values in the analysis. The post-hoc tests were conducted using pairwise t-tests with Bonferroni corrections. Usability Likert-scale data was analysed with Friedman tests with Bonferroni-corrected Wilcoxon tests for post-hoc analysis.

6.4 Results

In this section we analyse performance metrics for BimodalGaze and Pinpointing. For BimodalGaze we investigate how often refinement is required for selection, and reflect on this in our analysis of error rates and selection time.

6.4.1 Refinement

BimodalGaze provides the flexibility to use gaze pointing when it is sufficient, or to enter Head Mode when necessary. Out of all 2594 selections made with BimodalGaze,

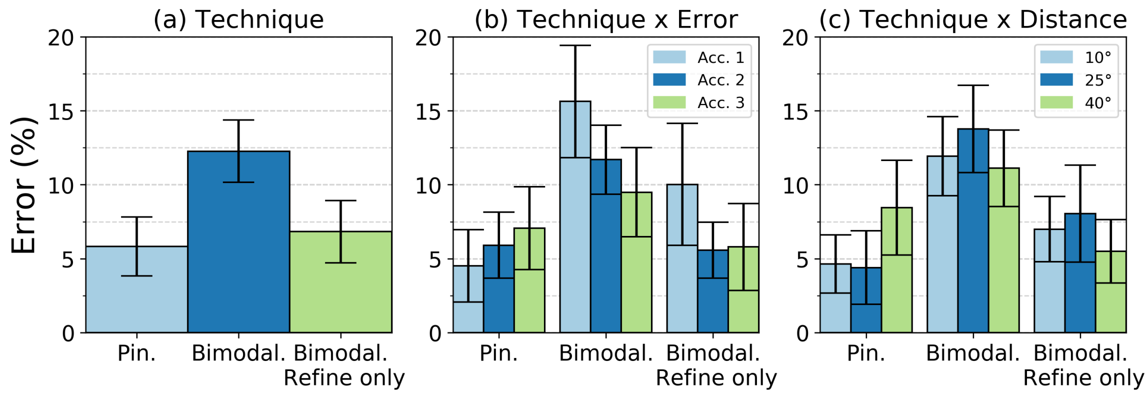

Figure 6.5: Mean error rate. Error bars represents mean 95% confidence interval.

454 (17.5%) were made in Gaze Mode and 2138 (82.5%) were made in Head Mode. Further insights (Table 6.1a) showed that both eye tracker accuracy and target distance affects the prevalence of Gaze Mode selections for BimodalGaze. Lower eye tracking accuracy and further target distances led to more selections made in Head Mode. Direction had no effect on the prevalence of selections made in Gaze or Head Mode.

6.4.2 Error Rate

All participants completed all trials with both Pinpointing and BimodalGaze. We define an error as when the participant missed the target prior to a correct selection. We report the number of errors as the error rate, i.e. the number of trials resulting in an error divided by the total number of trials (Figure 6.5).

We found no significant four-way or three way-interactions. However, we found a significant Technique × Accuracy two-way interaction ($F_{1.33,14.66}$=7.17, p<.05). Investigation of simple main effects revealed the error rate for Pinpointing was unaffected by eye tracking accuracy ($F_{2,22}$=2.26, p>.05). However, the error rate significantly decreased for BimodalGaze as eye tracking accuracy decreased ($F_{2,22}$=5.94, p<.05). Looking across techniques at each level of accuracy showed that Pinpointing had fewer errors than BimodalGaze at Acc. 1 ($F_{1,11}$=52.15, p<.001), and Acc. 2 ($F_{1,11}$=28.06, p<.001), but not at Acc. 3 ($F_{1,11}$=1.47, p>.05).

We also found a significant Technique × Distance interaction ($F_{2,22}$=8.46, p<.05). Distance had a significant simple main effect on Pinpointing ($F_{2,22}$=5.89, p<.05), revealing that participants made more errors as the distance increased. This could have been caused by the increase in head motion at larger distances, and thus the increased risk of timing issues between clicking and head movement, see Figure 6.2.

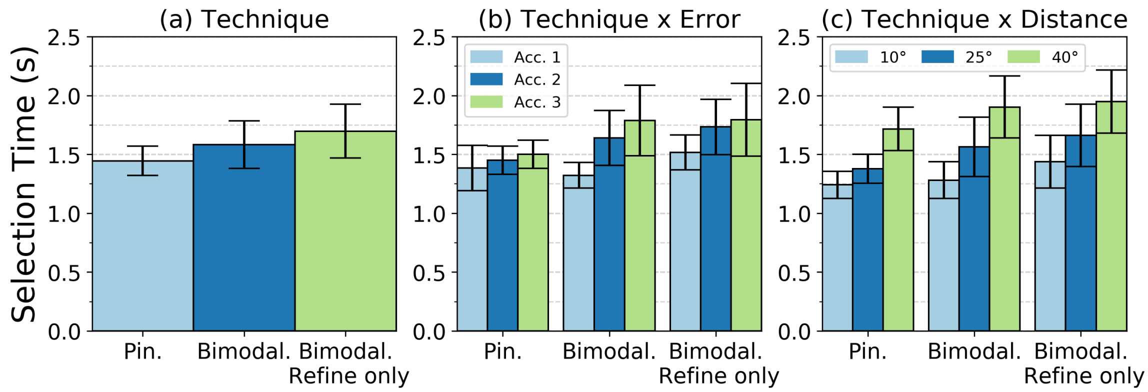

Figure 6.6: Mean selection time. Error bars represents mean 95% confidence interval.

For BimodalGaze, distance had no effect on error rate ($F_{2,22}$=1.94, p>.05). Looking at individual distances, we found Pinpointing resulted in significantly fewer error than BimodalGaze at 10° ($F_{1,11}$=44.23, p<.001) and 25° distance ($F_{1,11}$=59.99, p<.001), but not at 40° distance ($F_{1,11}$=3.01, p>.05).

Further investigation of errors made with BimodalGaze revealed over half were made in Gaze Mode, see Table 6.1b. Based on this finding, we investigated the number of errors made in Head Mode for BimodalGaze by calculating the error rates based on trials in which Head Mode was used, and in which an error was made during Head Mode. A three-way repeated measures ANOVA of Technique × Accuracy × Distance, revealed no significant difference between techniques for error rate (Pinpointing: 5.8% ± 0.9%, BimodalGaze: 7.2% ± 0.9%). We also found significant two-way interactions for Technique × Distance ($F_{2,22}$=10.21, p<.05) and Technique × Accuracy ($F_{1.4,15.1}$=4.13, p<.05). Further investigation of the simple main effects revealed that Pinpointing had significantly lower error rate for the highest accuracy condition compared with BimodalGaze (Pinpointing: 4.5%, BimodalGaze: 10%, p<.05).

6.4.3 Selection Time

We were interested in how automating the switch between Gaze and Head Mode affects selection time. We define selection time as the time between the start of a trial to a successful selection, irrespective to the amount of prior incorrect selections.

We found no significant interactions for selection time. Technique had a significant main effect ($F_{1,11}$=5.17, p<.05) where Pinpointing was significantly faster overall than BimodalGaze (Figure 6.6a). However, the mean overall difference was only 140ms. Accuracy also had a significant main effect ($F_{2,22}$=46.84, p<.001) where a lower

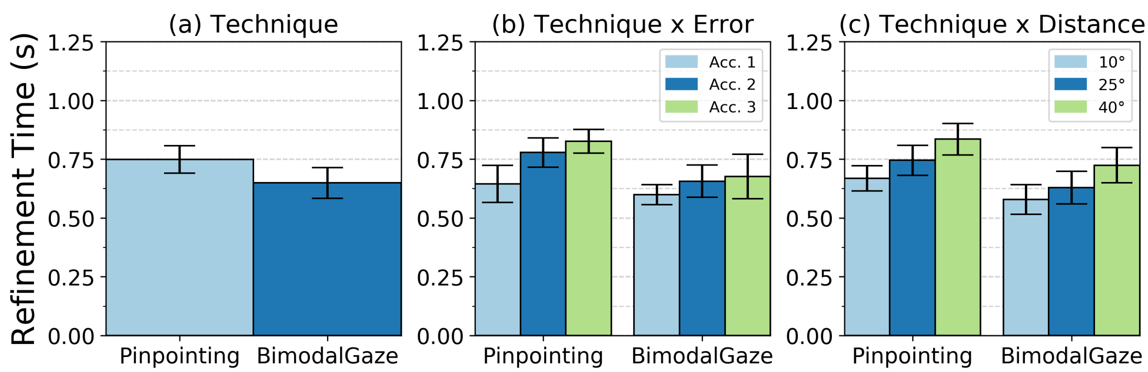

Figure 6.7: Mean refinement time for trials where refinement was used. Error bars represents 95% confidence interval.

accuracy, and therefore a higher reliance on Head Mode, lead to higher selection times. Post hoc-tests showed significant differences at all levels (all $p<0.05$). Finally, Distance had a main effect ($F_{2,22}$=61.52, $p<.001$). Unsurprisingly, higher distance lead to significantly higher selection time. Post-hoc tests showed significant differences between all levels (all $p<.001$).

Similarly to error rate, we analysed the average selection time for all trials in which Head Mode was used to select the target with BimodalGaze (and thus discard trials when selection was made only with gaze) using a three-way repeated measures ANOVA of Technique × Accuracy × Distance. We found BimodalGaze's selection time was significantly slower than Pinpointing by 220ms when we discounted selections made using only gaze (Pinpointing: 1.44±.56s, BimodalGaze: 1.66±.95s, $F_{1,11}$=11.57, $p<.05$).

6.4.4 Refinement Time

We excluded the trials where no refinement was used to investigate the average time spent in Head Mode for each technique (Figure 6.7). We found a main effect of Technique which showed users spend significantly more time in Head Mode with Pinpointing (0.75s) compared to BimodalGaze (0.65s) ($F_{1,11}$=15.78, $p<.05$). In Pinpointing the pointer is not visible before the user clicks the button to enter Head Mode, which could lead to participants spending more time finding and processing the pointer position at the start of refinement stage. BimodalGaze's lower refinement time combined with its higher selection time compared to Pinpointing, suggests that the switching could be further optimised. Eye tracking accuracy ($F_{2,22}$=44.25, $p<.001$) and Distance ($F_{2,22}$=71.20, $p<.001$) also had significant main effects. Post-hoc tests

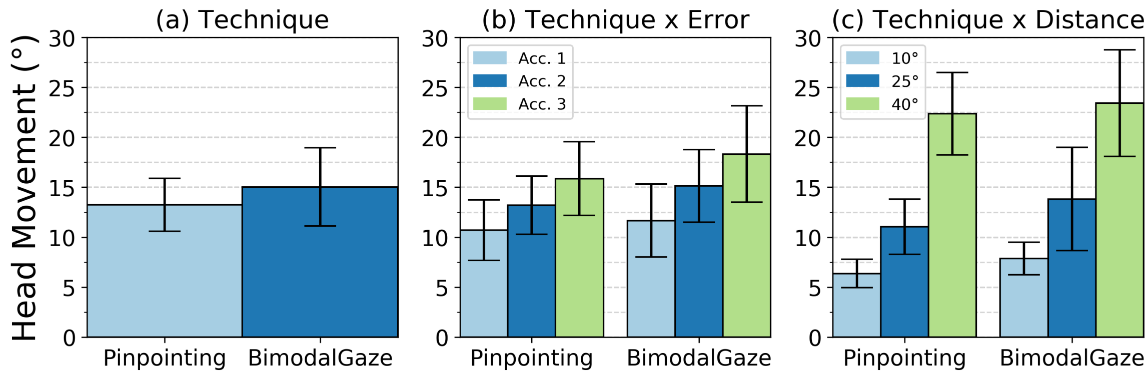

Figure 6.8: Mean head movement. Error bars represents mean 95% confidence interval.

showed that decreasing accuracy or increasing distance leads to more refinement time for both techniques (all p<.001).

6.4.5 Head Movement

Investigation into head movement during trials showed showed no significant interactions (Figure 6.8). Significant main effects of Accuracy ($F_{2,22}$=31.55, p<.001) and Distance ($F_{1.26,13.87}$=101.39, p<.001), indicate that lower eye tracking accuracy, or larger distances led to more head movement for both techniques. Direction seemed to have no effect on the prevalence of selections in Gaze Mode. This demonstrates that users did not need to perform significantly larger head movements for BimodalGaze compared with Pinpointing.

6.4.6 Qualitative results

Friedman tests on usability ratings showed significant differences on all metrics except learnability, however Bonferroni corrected Wilcoxon post-hoc tests showed no significant differences between conditions. In general, participant preferences were split between the two techniques (BimodalGaze: 5, Pinpointing: 7), with each offering unique advantages.

Participants praised the seamless switching of BimodalGaze which was *"effortless"* (P1), and because *"it worked all the time and I did not have to do anything"* (P11). The automatic switching between modes did not appear to work as well for some participants, which appeared *"a bit random"* (P4). Participants also noted that the technique worked better for conditions with low eye tracker accuracy where *"head*

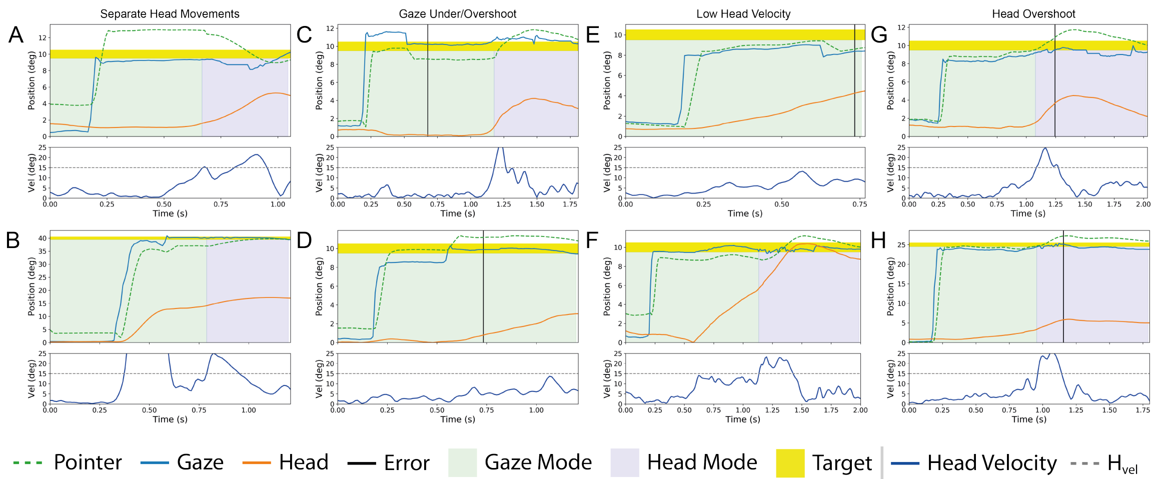

Figure 6.9: Trial examples of selections made with BimodalGaze.

movement was more pronounced which made the clutch control feel more natural" (P11), because for conditions with high eye tracking accuracy *"I only had to do a small head movement, which was not always identified by the system"* (P11). Participants also positively mentioned BimodalGaze's continuous feedback which provided opportunities to make informed decisions of whether to enter Head Mode - *"I knew immediately if and how I should adjust the pointer"* (P5). However, participants also mentioned that the continuous feedback could be distracting *"especially when the eyes were controlling the cursor"* (P2) or *"when the eye tracker was inaccurate"* (P12).

For Pinpointing, the main advantage was that participants felt *"more in control [because] I decided when to show the feedback"* (P2) because of the manual mode switching, which in turn meant for some *"it was easier to select the target"* (P10). However, some participants expressed difficulties with timing their button presses. P5 stated *"I would move my gaze towards the object and then start aligning my head. During this time I would press the button which resulted in me overshooting the target and then having to readjust. Sometimes, I would also press the button preemptively and then realise that my gaze is still at the centre"*, a point reiterated by P11 *"I found that I pressed the button too early. I would press the button too early so that the cursor was in the wrong place or the head would drag the cursor away from the target."*

6.4.7 Participant Behaviour

Based on our analysis and feedback, we further investigated individual trials to unveil participants' selection and gaze behaviour during pointing and selection using BimodalGaze (Figure 6.9). From this in-depth investigation, we found interesting characteristics of how eye tracker accuracy and system thresholds affected selection.

We observed users pausing to assess whether or not to enter Head Mode before making a gestural head movement. If a natural head shift was not needed to reach the target comfortably, participants would perform a gaze shift, assess the pointer position and then perform a gestural head shift (e.g. Figure 6.9a). Likewise, in the event of a natural head movement, users would stop or significantly slow down their natural head movement before performing a gestural head movement – they would very rarely transition from natural to gestural during a single head movement (e.g. Figure 6.9b).

Participants performed more selection errors in conditions where Gaze Mode was initially used for selection, see Table 6.1b. A common situation that leads to errors in Gaze Mode was the natural overshooting (e.g. Figure 6.9c) or undershooting (e.g. Figure 6.9d) of the primary saccade which may last up to 500 ms [10]. This, combined with the eye tracking accuracy error, may cause the pointer to appear on the target when the participant's gaze was not. This situation may cause the user to press the button for selection - which according to the Keystroke-Level Model proposed by Card, Moran, and Newell [24] could take anywhere between 80-280ms. However, by the time the button press was registered by the system, the users' gaze had moved onto the target, which in turn moves the pointer outside the target. This situation is more common for conditions with higher eye tracking accuracy, and not as prevalent for low accuracy conditions because the eye tracking accuracy error is larger than the error induced by over/undershooting, and as such the pointer rarely appears on the target.

We also found that our choice of a higher value for H_t led to an inability of BimodalGaze to identify small gestural head movements for small refinements. Users would rely entirely on Gaze Mode, which either caused incorrect selections (e.g. Figure 6.9e) or resulted in exaggerated head movements (e.g. Figure 6.9f). The latter led to overshooting the target when in Head Mode, which in turn led to unsuccessful selections as participants attempted to select the target whilst in the process of overshooting (Figure 6.9g, h). This phenomenon is more common when the eye tracking accuracy was higher, as only very small head movements would be needed for accurate refinement, and may explain the significant difference in error rate we see between the techniques for the highest eye tracker accuracy condition.

6.5 Discussion

The study results validate the principal approach of distinguishing between natural and gestural head movement for head-refinement of a gaze cursor. The comparative evaluation against a manually switched technique points to performance limitations that we discuss below. However, the headline results are:

- Users are effective with BimodalGaze. All participants successfully completed all selection tasks. When initial selections were off-target (i.e. counted as error in the study), users had no problem correcting their input.
- BimodalGaze demonstrates that gestural head movement can be reliably differentiated from natural head movements that are implicit with gaze, and harnessed as explicit input. Our work shows that the switching between gaze and head modes can be automated using the underlying knowledge of the way in which our head movement supports gaze.
- BimodalGaze enables selection of targets that users are not able to successfully acquire with gaze alone. Despite only accounting for 17.5% of selections, gaze-only errors accounted for over 50% of errors. This highlights that with gaze alone, users may not be able to select a target at all if the eye tracking accuracy is too poor, while our results show that participants are able to select all targets successfully with BimodalGaze, irrespective of scale of eye tracking error.

Compared to Pinpointing, BimodalGaze automates the mode switch between gaze and head input. The advantage is that our technique does not require explicit input for mode-switching, resulting in a more seamless transition. This also avoids the need for an additional input modality such as manual input in Pinpointing. Note that in our study, BimodalGaze was combined with a button click for selection confirmation. However, it could also be combined with alternate confirmation techniques such as dwelling to make the whole selection hands-free. This can be useful in situations where the hands are busy or unavailable.

A second feature by which BimodalGaze is different from Pinpointing is that head-refinement is optional rather than enforced. As our results show, over half of selections were made by gaze alone when eye tracker error was lower and target distance shorter. This validates the design choice in principle. However, 54.4% of selections made in gaze mode were off-target and required correction, which compromised the advantage and led to users spending longer time in gaze mode. Users made errors in gaze mode as a result of premature selection whilst under/overshooting the target. Clearer pointing feedback, such as target highlighting, or target acquisition techniques (e.g. BubbleCursor [53]) could be used to improve this aspect. In the latter case, Head Mode would be necessary in cases where the size of the bubble exceeds the density of the targets.

In performance comparison, BimodalGaze was on average slower for selection than Pinpointing, though not substantially ($\approx$ 140 ms). This was due to longer time spent in gaze mode, while BimodalGaze was faster in Head Mode ($\approx$100 ms). Overall, there was no significant difference in error rate in Head Mode. However, the error rate with BimodalGaze dropped when eye tracking error increased, and users made fewer errors than with Pinpointing when eye tracking error was highest. This indicates that our technique is particularly beneficial when eye tracking accuracy is poor.

The performance results suggest the mode switching could be further optimised. The time spend in gaze mode can be reduced by techniques that address premature selection, as discussed above. Another area of improvement are the criteria for entering Head Mode. The high H_t resulted in difficulties entering Head Mode when only small movements were required, and in turn causes the head to overshoot as a result of exaggerated head movement. We selected a value of H_t heuristically to minimise consistent mode switching. Optimisation of H_t, or use of more sophisticated techniques (e.g. accuracy-dependent thresholds), could alleviate this problem. Instead of a rule-based approach as used in this work, machine learning could be adopted for classifying head movements, or to optimise the system's parameters. Alternatively, dynamically adapting H_t based on users' individual movement tendencies could also be an interesting future direction.

The distinction of natural and gestural head movements makes it possible to attach different behaviours to objects that take gestural head movement as input while avoiding unwanted behaviours caused by natural head movements. In BimodalGaze, we used gestural head movements to refine a gaze cursor but other mappings are possible. For example, assuming that gaze pointing is accurate enough, gestural head movements could be used to manipulate (scaling, rotating, etc.) gazed on objects.

All our results were obtained in VR. However, we do not expect this to limit the applicability of BimodalGaze, as the technique builds on eye-head coordination behaviours that are consistent with observations in real-world tasks [88]. Head movement is more common when interactions span a wider FOV, for instance on large screens or across devices, but BimodalGaze is also applicable with narrower FOV displays where natural head movements is less prevalent.

6.6 Conclusion

In this chapter we introduced BimodalGaze, a novel technique for fine-grained control of a gaze cursor. The technique enables users to refine a gaze-cursor with head movement and is entirely hands-free, which is useful for situations where the hands are busy or unavailable. The transition from a gaze mode for initial cursor placement to refinement by head movement is implicit, based on detection of gestural head movement, which is significant as it removes the need for any manual or other explicit input thus making the process more seamless. Evaluation of the technique highlights advantages of the technique in particular when eye tracking accuracy is poor but also points to performance limitations in the present implementation, which this work addressed with in-depth analysis of user performance and errors observed. This not only provides insight for improvement of BimodalGaze, but also generally into the classification of gestural versus natural head movement. The notion of classifying natural and gestural head movements extends beyond refinement of a pointer, opening up new opportunities for mapping gestural head movement without affecting the head's natural ability to support gaze.

In this chapter we showed that leveraging classification of head gestures can be used to refine to address eye tracking accuracy and calibration challenges (RQ2). As such, we propose our final design principle for eye and head coordinated interaction: *Differentiate between natural and gestural head movements.* This distinction can be further employed to other aspects of gaze interaction and movement classification. BimodalGaze also show how advanced aspects of eye and head coordination, such as timings between movements and their directions can be leveraged for interaction to tackle gaze interaction challenges. In the following chapter we will reflect on the works presented in this thesis, and discuss lessons learned, limitations and future work.

Chapter 7

Discussion

The previous chapters have individually explored the behaviour and capabilities of using eye and head-coordination as input for 3D interfaces. In this chapter, we revisit our proposition of considering gaze as multi-modal, and reflect on the research questions and design principles posed in Chapter 1. We then look at lessons learned from across our studies. Finally, we discuss future work, including unexplored avenues of eye-head coordination for interaction, extending and combining the proposed design principles, and including additional modalities for coordinated interaction.

7.1 Gaze is Multi-modal

In the introduction of this thesis, we proposed that (1) gaze should be treated as multi-modal, i.e. the sum of the eye, head and body directions relative to the world; and (2) that considering the coordination of the eyes and head could support new forms of gaze-based interaction. The work presented in this thesis provides evidence about the significant potential the coordination of movements holds for interaction compared to how gaze was traditionally considered. The results demonstrate that gaze movements are worth exploring as a whole, including all movements of our body, not just the eyes. The behaviours we identified and leveraged for interaction allowed us to propose novel solutions to established problems within gaze and 3D interaction and extend the capabilities of hands-free gaze-based interaction in various settings. In addition, we presented several applications that provide qualitative benefits to users in ways that were usually considered awkward, inconvenient and straining via traditional gaze interaction. The lessons learned from this research lead us to propose several design principles for gaze-based interaction and avenues for future research.

Treating gaze as multi-modal goes beyond the scope of this thesis and may be helpful in multiple application areas. A significant amount of work has been spent on classifying eye movements for fields such as psychology, medicine and neuroscience.

Treating gaze as multi-modal allows such work to move beyond fixed environments and procedures where it can be expected that users only perform movements with their eyes to in-the-wild settings where the whole body is used. Treating gaze multi-modal also opens up a novel design space for interaction that includes the entire body, not just parts of the performed movements (e.g. eye or head movements).

7.2 Research Questions

Eye and head coordination has been extensively studied within psychology and neuroscience. We built on this work by investigating the common patterns of movements in VR and their applicability for 3D interaction by investigating the research questions introduced in Chapter 1.

7.2.1 RQ1: How does the eyes and head coordinate in VR?

In Chapter 3, we provided an extensive analysis of eye, head and body coordination during gaze shifts at different amplitudes, directions and postures. In general, we found that our results aligned well with findings established in real world settings of previous work. We established multiple common patterns during coordination including: gaze shifts can occur with or without head and body movement; the head helps the eyes reach the target and keeping them within a comfortable range; the eyes and head are in general not aligned to preserve energy; the amount of movement can depend on posture, direction, and user idiosyncrasies. These insights provide guidelines for ensuring that interaction does not interfere with the natural movements of our body, and as inspiration for novel interfaces.

7.2.2 RQ2: How can insights on eye-head coordination improve gaze-based interaction?

Based on the insights from Chapter 3, we proposed several interaction techniques aimed at tackling established gaze-interaction challenges and to provide novel interaction capabilities. In Chapter 4 we showed via our Eye&Head techniques how classifying gaze-shifts based on whether they are head-supported and gaze-head alignment can be used to tackle Midas touch issues while allowing free roaming of gaze. In Chapter 5 we expanded the work on gaze-head alignment to enable nested and continuous hands-free interaction in a wide array of contexts. Finally, in Chapter 6 we showed that utilising eye-head coordination is not only useful for classification of gaze, but can also be used to classify head movements - as part of a gaze shift or as part of a gesture for interaction. This distinction between head movements is especially relevant to HMD contexts where head movements are used to explore surroundings. Summarising, these techniques provide the vast design space of eye-head coordination in and as inspiration for future research.

7.3 Design Principles

We derived several general design principles for gaze-interaction in XR through the design, development and evaluation of the interaction techniques detailed in this thesis. These principles can serve as general rules of thumb for future gaze-interaction design.

7.3.1 Differentiate Between Head-supported and Eyes-only Gaze Shifts

The Eye&Head Pointer in Chapter 4 showed how the distinction between head-supported and eyes-only gaze shifts can effectively mitigate the Midas touch problem by letting head movement act as a clutch for cursor control. However, the difference can also be helpful as a general distinction beyond cursor control. The head requires a lot of energy to move. Previous research has shown that head movements are more likely when a user is interested in a general area or is expected to gaze at an area for an extended period. As such, every head movement can be treated as a change of user state and used as contextual cues in behavioural models or for changes in the UI or virtual environment. This aspect is especially relevant in XR environments since the content is all around the user, and head movements will therefore be prevalent in contrast to desktop settings where head movements are far between.

7.3.2 Use Head-eye Alignment to Signal Intent

Our second design principle leverages the tendency of users to not align their eyes and head during gaze shifts. This tendency is logical as moving the head the extra distance to 'close the gap' is a waste of energy as it has little impact on our vision. In Chapter 4, we leveraged this insight as a general selection technique where an alignment of gaze and head led to selection and found that it was fast and accurate for target selection. The natural separation of the eye and head direction avoids Midas touch issues. The technique serves as a general alternative to the typical dwell-based confirmation mechanism. In Chapter 5, we extended this principle to a holistic user interface. Eye-head alignment allowed us to expand gaze-based and hands-free interaction from simple selection to more advanced interactions such as nested and continuous interactions that are usually unavailable when using gaze.

7.3.3 Differentiate Between Natural and Gestural Head Movements

Our final design principle regards distinguishing between head movements in conjunction with a gaze shift or as part of a gesture independent from gaze. As we showed in Chapter 3, gaze shifts and their associated head movements follow specific patterns. We then showed that these patterns can be detected through simple heuristics in Chapter 6 with our BimodalGaze technique. The ability to distinguish between head movements is especially relevant in XR environments where both will be commonly performed by users. Trying to decide whether a head movement is part of a gaze shift as a gesture in a social context or as an interaction is nearly impossible when only using the head. In previous work where head gestures have been deployed, it is often assumed that natural head movements are few or nonexistent. However, using the combined information of the eyes and head allows for a more straightforward classification. Head gestures can be safely deployed in varying contexts without worrying about false-positive gesture activations.

7.4 Gaze Interaction Challenges

Through the use of eye-head coordination, we have shown in this thesis that established gaze interaction challenges can be tackled in novel ways.

7.4.1 Midas Touch

The Midas Touch issue is the most well-known and established interaction challenge for gaze. Typically, Midas Touch is addressed through dwelling where users dwell at a target longer than what they would naturally do to perform an interaction which is usually straining. In this thesis, we showed that adding the head as extra information can be used effectively to avoid Midas Touch by being able to act as a clutch for gaze interaction (Chapter 4) to allow free eye movements without interaction, or through alignment of the eyes and head instead of long and strenuous fixations (Chapter 4 and Chapter 5). Since head information is always available in head-mounted XR, designers should always consider its usage when developing gaze-based techniques where Midas Touch issues may be present.

7.4.2 Lack of Expressive Interaction

A significant disadvantage for gaze compared to other modalities such as the hands is the lack of established techniques for interaction beyond simple selection. Advanced interaction (sliders, nested menus etc.) that is used in daily interaction is missing for gaze-based interaction, limiting its possible use cases and applications. Furthermore, the eyes are unsuitable for high-granular interactions due to their jittery and saccadic nature. In this thesis, we showed how these limitations of gaze-based interaction can be countered by not treating gaze as eyes-only but as a coordination of the eyes and head. In Chapter 5, we combined coordination with a radial menu to expand the possibilities of gaze-based interaction to include continuous interaction, multi-choice widgets, and nested menus. The presented applications highlight the potential of turning gaze into an expressive and effective modality for hands-free interaction.

7.4.3 Accuracy and Calibration Issues

Finally, the accuracy and calibration issues inherent in gaze interactions are a significant roadblock to integrating gaze into everyday interaction. In this thesis, we showed how the more accurate head can be used as a complement to gaze-based pointing. This leverages the quickness of the eyes together with the accuracy of the head to enable fast and accurate hands-free pointing irrespective of the presence of accuracy error caused by calibration issues or slippage of the HMD. Furthermore, the techniques also allow the selection of smaller targets that would be difficult to select with gaze alone, allowing more freedom in the design of gaze-based interfaces. In conclusion, considering gaze as multimodal presents new ways of tackling these issues and expands the possibilities of gaze-based interaction.

7.5 Lessons Learned

In addition to the proposed design principles and the learnings from tackling gaze interaction challenges, we found insights that may be useful to future gaze-based interfaces. These insights originated while working on the interaction techniques and applications, but also from study participants' comments and informal discussions with other researchers during the PhD.

7.5.1 Separate the Responsibilities of the Eyes and Head

A key insight from our works lies in separating the responsibilities for interaction between gaze and head. Multiple previous results have shown that gaze is faster than the head for pointing and selection, while the head is more stable and controlled [18, 56, 77, 87, 138]. In our techniques, we showcased how we can leverage the modalities' strengths but avoid their weaknesses through the separation of interaction responsibilities. This allows distinction for fast yet refined hands-free interaction. For example in Radi-Eye, the fast eyes are used for pre-selection and highlighting, while refined head movements are used for selection confirmation and continuous input.

Using the head for continuous and extended input also allows gaze to wander freely, avoiding lengthy and straining dwell times. In Eye&Head Convergence, we find similar results and show that including the head for interaction does not mean that selection performance or speed is lost. In sum, the role of the eyes and head should be considered and weighed against their relative strengths and weaknesses.

7.5.2 Consider the Eye-in-head Angles

A second insight from the combined works is that the eye direction relative to the head has to be considered during the design of eye-head coordinated interaction techniques and interfaces. As we can see in Chapter 3, participants tended to keep their eyes within a comfortable eye-in-head angle (∼15-30 degrees). The eyes were waiting for the head to catch up to remain within that comfortable range. Interactions should not force participants to move beyond the eye-in-head angle too far, as it may cause strain, or even the interacted object is no longer visible. This aspect was most noticeable during the design of Radi-Eye, where large interfaces could induce significant gaze shifts and head movements if the menu consisted of multiple nested levels. In these situations, the menu size and individual widget sizes have to be considered to ensure that users can perform interactions without the eyes and head moving too far apart. This aspect is also considered in the design of Eye&Head Convergence and Look & Cross, where the eyes and head are aligned for interaction. These techniques "close the gap" between the eye and head so that they can move away from each other again for the following selection.

7.5.3 Gaze Shifts are Varied

In Chapter 3, we showed that gaze shifts and the movements of the eyes, head and torso follow specific patterns. However, we also showed that these patterns are directly linked to factors such as target amplitude, posture, direction, and user-specific factors. Furthermore, our experiment only considered gaze shifts that started from a central position. Other starting positions, e.g. the user glancing at the start of the shift, may also affect the amount of movement deployed. The takeaway from this insight is that we can't expect gaze shifts to always be the same when designing gaze-based interfaces. This insight was most apparent in Chapter 6. For the BimodalGaze technique, we used velocity thresholds to define head movements for pointer refinement. This approach was simple and easy to implement. However, it limited the head movements users could perform for refinement. Most notably, users could not perform minor refinements as the system did not recognise the small head movements due to not reaching the velocity threshold. This highlights that gaze shifts and movements can be performed in various ways. Interaction designers must consider what movements they disregard or ignore when designing interfaces and what impact they may have on user behaviour and interaction performance.

7.6 Future Work & Limitations

This research developed various interaction techniques and applications from which we derived design principles. While this shows the rich possibilities of eye- and head-coordinated interaction, our research uncovered many directions we could not cover within this thesis. This section points toward some areas of future research.

7.6.1 Generalised Design Principles

In this thesis, we proposed multiple general design principles for gaze. Still, we only demonstrated them in specific application examples or studies. It is therefore worth questioning their generalisability. We believe that the design principles' usefulness goes beyond the singular shown application. For example, differentiating between head-supported and eyes-only gaze shifts may be helpful in applications that want to track users' attention. A head-supported gaze shift can then be treated as a marker for a shift in attention. Such applications can also be implicit instead of the explicit control of the Eye&Head Pointer. Furthermore, differentiating between natural and gestural head movements goes beyond simply pointing refinement and can be used for a wide variety of interactive scenarios (i.e. head gestures [198, 199]) or for the classification of user movements.

7.6.2 Individual Gaze Differences

In Chapter 3, we found individual differences in the amount of head movement deployed during gaze shifts - most participants relied heavily on the head during shifts, while some only used the head when necessary. These findings are further supported by previous research in neuroscience and psychology, who coined the terms "head-movers" and "non-head mover" to describe people [47]. Our study contained too few participants to thoroughly investigate this aspect but is a future research direction of interest. Future studies should be conducted to fully uncover this phenomenon. Furthermore, calibrating users based on their head-movement tendency and then tailoring the virtual experiences may be beneficial for user experience in 3D interfaces. Interaction benefit examples may be to reduce head strain or for movement prediction.

7.6.3 Gaze Zones

Another insight from Chapter 3 left unexplored is the proposal of three different gaze zones: gaze shifts up to 25° that can be comfortably achieved eyes-only; gaze shifts in the range 25–50°, where eyes and head together reach targets comfortably; and gaze shifts over amplitudes larger 50° where eyes, head and also torso need to work together for comfortable viewing. These zones could be leveraged for interaction. For example, widgets could be attached to the head and placed outside the inner zone. This allows gaze-based interaction but avoids Midas touch issues as the widgets are placed outside the inner gaze zone. This concept was recently explored by other researchers who developed the "Kuiper Belt" interface where widgets were placed outside the "Out-of-natural Angle" region [27]. They showcased the concept's potential through a multitude of studies and example applications.

7.6.4 Leveraging VOR Movements

A key aspect of eye-head coordination as shown in Chapter 3, is the VOR eye movements that users perform during head motion to maintain gaze on an object. The interaction techniques and principles presented in this thesis only implicitly leverage these movements (i.e. eye-head alignment). Future work could explore how VOR could be explicitly used for interaction. For example, previous work has shown that this aspect can be used for target disambiguation in VR [103, 105]. In particular, VOR movements are also extra relevant in contexts where the body and, in extension, the head are constantly moving.

7.6.5 Beyond Pointing

This work is mainly limited to pointing interaction in 3D interfaces. All the developed techniques presented in this thesis to some extent use ray-casting for pointing or manipulation. However, the use of eye and head coordination goes beyond pointing and future research could investigate how the insights of eye and head coordination can be leveraged for other types of interaction. For example, extensive research has been done on both gaze [150] and head gestures [198, 199]. The coordinated relationship could also be further explored as features for machine learning based activity recognition and as contextual information in intelligent user interfaces.

7.6.6 Beyond Stationary Contexts

Another limitation of our work is the focus on stationary contexts for interaction. In all our studies and applications, it is assumed that users are stationary and thus only perform rotational movements during gaze shifts. However, a significant part of our coordination is performed while we are in motion, for example, during walking. In addition, many of our daily tasks are performed while moving between different areas, such as when making a cup of tea [90]. This aspect adds another level of complexity to the design of coordinated interactions that should be investigated in future work. We have already started seeing XR interfaces designed for "mobile-contexts" such as when walking [99] and adding coordination insights may increase their usability.

7.6.7 Head-strain

A typical comment from user study participants was that techniques that rely on head movement caused head strain. Head-based techniques don't only require users to move their heads more than what they would typically do. Still, users must also move their heads in particular ways for interaction. These aspects, in combination with the extra weight of modern HMDs on top of the head, can quickly lead to extensive neck strain. In the design of our techniques, we tried to minimise head movements by relying on velocity rather than amplitude and alignment instead of unalignment. However, these designs did not fully alleviate the issue. Similarly to eye-in-head angles previously discussed, the head positions and movement must be considered when creating eye-head coordination techniques. Extensive research has been conducted on arm fatigue during mid-air interaction which has spawned models to quantify arm strain [58, 78] and ergonomic mid-air interfaces that limit arm strain [41]. Similar research should be conducted on the head and neck strain to quantify its effects and design interfaces designed to mitigate them.

7.6.8 Wider Usage

In addition to head strain, fundamental questions remain about how coordinated eye-head techniques affect users during prolonged and widespread use. Gaze-based techniques have been shown to cause digital eye strain (DES) in users after use, which can negatively affect their well-being and productivity [59]. Whether our techniques cause DES is an area for future research. Similarly, cybersickness is a common issue in VR, with users showing symptoms similar to motion sickness [93]. Whether the techniques and principles presented in this thesis are likely causes of cybersickness has yet to be investigated. Finally, all the techniques presented in this thesis were designed to rely only on the common outputs of the 3D eye trackers (directional vectors), making them device independent. However, all depend on accurate and stable eye tracking output to function properly, including BimodalGaze, which was designed to address accuracy issues but rely on stable and precise eye tracking. The performance of the techniques during poor eye tracking and the signal quality required for proper function must be investigated.

7.6.9 Including the Body

Another limitation of this thesis is that it focusses only on eye-head coordination for interaction. As we stipulated at the start of this chapter (Section 7.1), *gaze is multi-modal* - the sum of eye, head and body directions relative to the world. We further showed in Chapter 3 that the torso can vary significantly during gaze shifts depending on direction and posture. Suppose that the torso is restricted, for example, in a seated position. In that case, the head has to compensate and move further relative to the torso. These aspects may have an impact on user behaviour during interaction. In this thesis, we decided to focus only on eye-head interaction. Thus, the role of the body and its effects on interaction were ignored.

Future research investigating the role of the body may be of significant interest. For topics such as head-strain (Section 7.6.7) where the head's relative position to the body (i.e. "head-in-body" position) is a significant cause, the inclusion of body sensing may be of utmost importance. Furthermore, tracking the body may also allow eye-head interaction to better consider the current context for interaction. For example, suppose the system knows that the torso is restricted due to the user being seated or in a crowded area. In that case, interactable content should only be limited to directions comfortably reachable with only the head and eyes. In summary, including the body as part of gaze would be a more accurate representation of "gaze in the world", open up new opportunities for gaze interaction, and be a natural continuation of the work presented in this thesis.

7.6.10 Privacy Implications

Our work finally exposes privacy implications that should be taken into account. All eye tracking applications expose multiple security and privacy threats, such as user identification, information extraction, and action prediction without the user's knowledge [31, 82]. Furthermore, previous VR work has shown that body movements and gaze behaviour that users perform during interaction can be used for identification [129], including the combination of eye and head movements [95]. We found evidence in this thesis (Chapter 3) that points to the possibility that users have idiosyncratic head movement tendencies. These idiosyncrasies could potentially be used for implicit user identification. The consequences of such systems and the potential privacy risks exposed by eye-head coordination should be investigated in future work.

Chapter 8

Conclusion

In this thesis, we explored the capabilities of gaze-based interactive techniques and applications that leverage the coordination of the eyes and head in 3D environments. We expanded the view of gaze by not considering gaze as a single input channel but rather multi-modal and continuously coordinating between the different systems and emphasising the natural movements performed during gaze shifts in XR. We used various data collection methods to ground our work empirically. We showed that the coordination of the eyes and head holds great potential for tackling gaze interaction problems and expanding hands-free expressiveness for interacting in XR.

This work proposed gaze as multi-modal and has demonstrated the potential of understanding users' coordinated movements during gaze shifts to create interfaces that can adapt to the users' natural movements in 3D environments. The work has shown that it is beneficial to account for the limitations of these movements and to be aware of their underlying processes, especially in 3D environments where gaze behaviour involves significant movement of both the eyes and head. These insights allow the design of interactive techniques that do not constrain users' gaze movements but instead take advantage of their common patterns to enable new forms of interaction.

We believe that this work and treating gaze as multi-modal has opened a new way to think about gaze in the design of gaze-based interfaces. Beyond traditional uses of gaze and eye tracking, we have shown a wide variety of coordinated movements to explore. Treating gaze as multi-modal opens up a new exciting design space for gaze-based interaction that is a more accurate representation of the movements performed by users. These learnings can be useful in developing novel behaviour for interaction, reducing movement strain and for movement classification or intent prediction. We have only scratched the surface of how interaction techniques and interfaces can be used in novel ways while respecting the natural roles of gaze. There are more aspects to gaze and their coordinated movements than we have investigated in this thesis. We hope future work will continue to discover new elements of coordination that we may have overlooked for interaction.

While we have shown that eye and head coordination can be effectively leveraged as an input channel in various XR settings, it is essential to highlight the modalities' respective limitations. The eyes are a sensory organ, and their erratic behaviour of constantly seeking new visual stimuli is unsuited for granular and precise interaction. On the other hand, head movements should be limited to not induce excessive neck strain on users due to the head's weight. Therefore it is crucial to understand what humans use both modalities for and respect their freedom of movement so that interaction and visual seeking can co-exist in harmony. Gaze should not be considered as a means to an end for interaction, but the context of use and the advantages that gaze bring should first be considered before being introduced for interaction to ensure that the natural use of gaze is not unnecessarily affected or hampered.

The interaction techniques and applications we have developed in this thesis raise questions to investigate next. While we have shown the advantages of gaze-based interaction in 3D environments, we have only given simple examples of their usefulness. Further work is needed to thoroughly investigate their broader potential, impact, and limitations for human-computer interaction in the coordination of the eyes and head. The full coordination of the eyes and head is still not fully established. All our techniques have only been investigated in lab contexts and single-user environments. Their use in social settings and beyond lab environments leads to much bigger questions about the role of gaze in the future than what we have studied in this thesis.

As XR become increasingly common in our daily lives, so will 3D interfaces and techniques for interaction. It is interesting to think about what role gaze-based interfaces will have as XR merge into people's daily lives. Designers will need to create natural interfaces that are easy to learn and master while not increasing cognitive load or physical strain. Specific attention should be given to the audience contexts and which movements the users can comfortably perform in their current context. By doing so and by leveraging the insights provided in this thesis, future gaze-based interfaces can lead to powerful and seamless interactions between users and the future digital world.

Appendix A

Modality contributions

A.1 Eye Contribution

Table A.1: Repeated measures ANOVA statistical analysis on the effect of amplitude on eye contribution.

Eye Contribution Amplitude						
All Amplitudes	3-way Interaction			$F(8.86, 168.39) = 1.20$, $p > .05$		
	2-way Interaction Amplitude, Direction	**Sitting** $F(7.86, 149.43) = 4.92$, $p < .001$			**Standing** $F(5.17, 98.18) = 5.64$, $p < .001$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	$F(2.91, 55.28) = 0.53$ $p > .05$	$F(3.83, 72.68) = 0.32$ $p > .05$	$F(3.41, 64.81) = 1.71$ $p > .05$	$F(3.02, 57.43) = 1.43$ $p > .05$	$F(2.45, 46.57) = 1.01$ $p > .05$
	Amplitude Main Effect, Sitting Posture	$F(2.92, 55.45) = 215.90$ $p < .001$	$F(2.77, 52.60) = 283.64$ $p < .001$	$F(2.78, 52.87) = 527.47$ $p < .001$	$F(3.00, 56.97) = 426.78$ $p < .001$	$F(2.81, 53.46) = 289.84$ $p < .001$
	Amplitude Main Effect, Standing Posture	$F(2.15, 40.77) = 207.53$ $p < .001$	$F(2.43, 46.16) = 267.35$ $p < .001$	$F(2.20, 41.70) = 594.60$ $p < .001$	$F(1.83, 34.82) = 492.58$ $p < .001$	$F(2.43, 46.10) = 259.62$ $p < .001$
Within-view	3-way Interaction			$F(6.85, 130.05) = 0.99$, $p > .05$		
	2-way Interaction Amplitude, Direction	**Sitting** $F(5.88, 111.77) = 3.64$, $p < .05$			**Standing** $F(3.82, 72.54) = 4.91$, $p < .05$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	$F(2.07, 39.35) = 0.76$ $p > .05$	$F(2.40, 45.64) = 0.14$ $p > .05$	$F(2.39, 45.38) = 1.18$ $p > .05$	$F(1.91, 36.37) = 1.33$ $p > .05$	$F(1.45, 27.58) = 0.98$ $p > .05$
	Amplitude Main Effect, Sitting Posture	$F(2.42, 46.02) = 36.76$ $p < .001$	$F(1.98, 37.67) = 24.27$ $p < .001$	$F(1.77, 33.64) = 40.42$ $p < .001$	$F(1.81, 34.47) = 25.00$ $p < .001$	$F(1.30, 24.64) = 24.14$ $p < .001$
	Amplitude Main Effect, Standing Posture	$F(1.60, 30.35) = 41.84$ $p < .001$	$F(1.59, 30.29) = 24.21$ $p < .001$	$F(1.42, 26.95) = 33.96$ $p < .001$	$F(1.33, 25.33) = 25.85$ $p < .001$	$F(1.58, 29.98) = 11.77$ $p < .001$
Beyond-view	3-way Interaction			$F(4.32, 82.02) = 1.45$, $p > .05$		
	2-way Interaction Amplitude, Direction	**Sitting** $F(3.78, 60.43) = 2.49$, $p > .05$			**Standing** $F(4.32, 82.02) = 1.45$, $p > .05$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	$F(1.21, 22.93) = 0.11$ $p > .05$	$F(1.52, 28.88) = 0.59$ $p > .05$	$F(1.42, 26.91) = 0.14$ $p > .05$	$F(2, 38) = 2.78$ $p > .05$	$F(2, 38) = 1.66$ $p > .05$
	Amplitude Main Effect, Sitting Posture	$F(1.22, 23.10) = 44.22$ $p < .001$	$F(1.25, 28.66) = 28.66$ $p < .001$	$F(2, 38) = 65.45$ $p < .001$	$F(1.34, 25.42) = 47.74$ $p < .001$	$F(2, 38) = 38.21$ $p < .001$
	Amplitude Main Effect, Standing Posture	$F(1.34, 25.51) = 57.43$ $p < .001$	$F(1.40, 26.59) = 34.23$ $p < .001$	$F(2, 38) = 133.01$ $p < .001$	$F(2, 38) = 158.15$ $p < .001$	$F(2, 38) = 41.04$ $p < .001$

Table A.2: Repeated measures ANOVA statistical analysis on the effect of posture and direction on eye contribution.

Eye Contribution Posture and Direction						
5°	2-way Interaction Posture, Direction	F(4, 76) = 0.89, $p > .05$				
	Direction Main Effect	**Sitting** F(4, 68) = 1.67, $p > .05$			**Standing** F(4, 76) = 1.34, $p > .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 2.35 $p > .05$	F(1, 19) = 0.47 $p > .05$	F(1, 19) = 4.11 $p > .05$	F(1, 19) = 3.82 $p > .05$	F(1, 19) = 3.89 $p > .05$
15°	2-way Interaction Posture, Direction	F(2.23, 42.29) = 0.67, $p > .05$				
	Direction Main Effect	**Sitting** F(1.96, 37.28) = 1.26, $p > .05$			**Standing** F(4, 76) = 2.46, $p > .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 0.01 $p > .05$	F(1, 19) = 2.04 $p > .05$	F(1, 19) = 3.63 $p > .05$	F(1, 19) = 0.19 $p > .05$	F(1, 19) = 4.02 $p > .05$
25°	2-way Interaction Posture, Direction	F(4, 76) = 0.86, $p > .05$				
	Direction Main Effect	**Sitting** F(2.08, 39.55) = 3.95, $p < .05$			**Standing** F(1.99, 34.44) = 4.70, $p < .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 0.281 $p > .05$	F(1, 19) = 1.37 $p > .05$	F(1, 19) = 1.90 $p > .05$	F(1, 19) = 4.26 $p > .05$	F(1, 19) = 0.02 $p > .05$
35°	2-way Interaction Posture, Direction	F(2.91, 55.36) = 0.98, $p > .05$				
	Direction Main Effect	**Sitting** F(4, 76) = 7.10, $p < .001$			**Standing** F(1.97, 37.42) = 5.70, $p < .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 0.233 $p > .05$	F(1, 19) = 0.10 $p > .05$	F(1, 19) = 0.45 $p > .05$	F(1, 19) = 0.87 $p > .05$	F(1, 19) = 3.99 $p > .05$
45°	2-way Interaction Posture, Direction	F(4, 76) = 0.96, $p > .05$				
	Direction Main Effect	**Sitting** F(2.82, 53.67) = 6.11, $p < .05$			**Standing** F(2.08, 39.49) = 7.18, $p < .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 1.54 $p > .05$	F(1, 19) = 0.95 $p > .05$	F(1, 19) = 0.17 $p > .05$	F(1, 19) = 0.48 $p > .05$	F(1, 19) = 0.06 $p > .05$
60°	2-way Interaction Posture, Direction	F(2.40, 45.61) = 1.02, $p > .05$				
	Direction Main Effect	**Sitting** F(2.58, 48.92) = 4.01, $p < .05$			**Standing** F(2.46,46.73) = 3.05, $p < .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 0.37 $p > .05$	F(1, 19) = 1.12 $p > .05$	F(1, 19) = 3.54 $p > .05$	F(1, 19) = 0.85 $p > .05$	F(1, 19) = 1.76 $p > .05$
80°	2-way Interaction Posture, Direction	F(2.41, 45.83) = 1.56, $p > .05$				
	Direction Main Effect	**Sitting** F(2.95, 56.02) = 8.01, $p < .001$			**Standing** F(2.72, 51.75) = 7.14, $p < .001$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 3.60 $p > .05$	F(1, 19) = 2.85 $p > .05$	F(1, 19) = 2.08 $p > .05$	F(1, 19) = 0.75 $p > .05$	F(1, 19) = 3.01 $p > .05$
100°	2-way Interaction Posture, Direction	F(2.28, 43.36) = 2.91, $p > .05$				
	Direction Main Effect	**Sitting** F(2.44, 43.88) = 6.61, $p < .05$			**Standing** F(4, 76) = 14.24, $p < .001$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 0.74 $p > .05$	F(1, 19) = 4.41 $p > .05$	F(1, 19) = 3.52 $p > .05$	F(1, 19) = 2.06 $p > .05$	F(1, 19) = 0.56 $p > .05$

A.2 Head Contribution

Table A.3: Repeated measures ANOVA statistical analysis on the effect of amplitude on head contribution.

Head Contribution Amplitude

All Amplitudes	3-way Interaction			F(9.329, 177.259) = 6.592, $p < .001$		
	2-way Interaction Amplitude, Direction	**Sitting** F(7.55, 143.53) = 4.21, $p < .001$			**Standing** F(6.18, 117.50) = 7.77, $p < .001$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(3.55, 67.42) = .434 $p > .05$	F(2.63, 49.88) = 7.80 $p < .001$	F(3.62, 68.84) = 22.79 $p < .001$	F(2.84, 53.94) = 25.44 $p < .001$	F(2.33, 44.35) = .94 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(3.03, 57.49) = 200.75 $p < .001$	F(2.44, 46.39) = 301.23 $p < .001$	F(2.52, 47.88) = 508.84 $p < .001$	F(2.81, 53.36) = 441.60 $p < .001$	F(2.52, 47.92) = 288.64 $p < .001$
	Amplitude Main Effect, Standing Posture	F(2.53, 48.01) = 227.77 $p < .001$	F(2.44, 46.26) = 165.12 $p < .001$	F(3.12, 59.28) = 306.83 $p < .001$	F(2.18, 41.32) = 267.12 $p < .001$	F(2.61, 49.54) = 268.02 $p < .001$
Within-view	3-way Interaction			F(7.01, 133.23) = 1.07, p = .389		
	2-way Interaction Amplitude, Direction	**Sitting** F(5.91, 112.34) = 3.08, $p < .05$			**Standing** F(4.38, 83.26) = 4.53, $p < .05$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(2.50, 47.46) = 0.30 $p > .05$	F(2.41, 45.81) = 0.31 $p > .05$	F(2.56, 48.61) = 1.51 $p > .05$	F(1.81, 24.42) = 0.86 $p > .05$	F(1.48, 28.07) = 1.00 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(2.43, 46.25) = 34.73 $p < .001$	F(1.88, 35.77) = 29.45 $p < .001$	F(1.74, 33.14) = 45.73 $p < .001$	F(1.76, 27.04) = 27.04 $p < .001$	F(1.35, 25.72) = 28.04 $p < .001$
	Amplitude Main Effect, Standing Posture	F(1.75, 33.16) = 46.47 $p < .001$	F(1.80, 34.23) = 31.98 $p < .001$	F(1.69, 32.18) = 42.02 $p < .001$	F(1.33, 25.32) = 28.51 $p < .001$	F(1.69, 32.14) = 13.11 $p < .001$
Beyond-view	3-way Interaction			F(4.69, 89.09) = 2.34, $p > .05$		
	2-way Interaction Amplitude, Direction	**Sitting** F(4.23, 80.45) = 2.67, p = .035			**Standing** F(4.11, 78.12) = 10.75, p ¡ .001	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(1.48, 28.04) = 0.99 $p > .05$	F(1.40, 26.58) = 14.98 $p < .001$	F(2, 38) = 19.69 $p < .001$	F(1.49, 28.24) = 8.70 $p < .05$	F(2, 38) = 1.27 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(2, 38) = 31.15 $p < .001$	F(1.31, 24.90) = 29.15 $p < .001$	F(2, 38) = 59.17 $p < .001$	F(1.23, 23.41) = 22.82 $p < .001$	F(1.25, 23.82) = 37.03 $p < .001$
	Amplitude Main Effect, Standing Posture	F(1.50, 28.44) = 36.65 $p < .001$	F(1.43, 27.14) = 4.12 $p < .05$	F(2, 38) = 14.56 $p < .001$	F(2, 38) = 7.89 $p < .001$	F(2, 38) = 30.41 $p < .001$

Table A.4: Repeated measures ANOVA statistical analysis on the effect of posture and direction on head contribution.

Head Contribution Posture and Direction

5°	2-way Interaction Posture, Direction	F(4, 76) = 0.94, $p > .05$				
	Direction Main Effect	**Sitting** F(2.57, 48.83) = 2.91, $p > .05$			**Standing** F(4, 76) = 0.24, $p > .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.01 $p > .05$	**Up-Diagonal** F(1, 19) = 0.03 $p > .05$	**Horizontal** F(1, 19) = 0.52 $p > .05$	**Down-Diagonal** F(1, 19) = 1.66 $p > .05$	**Downwards** F(1, 19) = 3.51 $p > .05$
15°	2-way Interaction Posture, Direction	F(1.97, 37.40) = 1.11, $p > .05$				
	Direction Main Effect	**Sitting** F(1.98, 37.60) = 1.77, $p > .05$			**Standing** F(2.44, 46.29) = 2.03, $p > .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.54 $p > .05$	**Up-Diagonal** F(1, 19) = 0.55 $p > .05$	**Horizontal** F(1, 19) = 4.21 $p > .05$	**Down-Diagonal** F(1, 19) = 0.38 $p > .05$	**Downwards** F(1, 19) = 3.27 $p > .05$
25°	2-way Interaction Posture, Direction	F(2.82, 53.49) = 0.71, $p > .05$				
	Direction Main Effect	**Sitting** F(2.04, 38.83) = 3.86, $p < .05$			**Standing** F(1.82, 34.50) = 4.43, $p < .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.07 $p > .05$	**Up-Diagonal** F(1, 19) = 1.24 $p > .05$	**Horizontal** F(1, 19) = 1.95 $p > .05$	**Down-Diagonal** F(1, 19) = 3.74 $p > .05$	**Downwards** F(1, 19) = 0.01 $p > .05$
35°	2-way Interaction Posture, Direction	F(4, 76) = 1.32, $p > .05$				
	Direction Main Effect	**Sitting** F(4, 76) = 7.22, $p < .001$			**Standing** F(1.98, 37.63) = 6.27, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.24 $p > .05$	**Up-Diagonal** F(1, 19) = 0.05 $p > .05$	**Horizontal** F(1, 19) = 0.88 $p > .05$	**Down-Diagonal** F(1, 19) = 1.56 $p > .05$	**Downwards** F(1, 19) = 3.62 $p > .05$
45°	2-way Interaction Posture, Direction	F(4, 76) = 0.69, $p > .05$				
	Direction Main Effect	**Sitting** F(2.77, 52.68) = 5.57, $p < .05$			**Standing** F(2.28, 43.30) = 7.33, $p < .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.63 $p > .05$	**Up-Diagonal** F(1, 19) = 0.28 $p > .05$	**Horizontal** F(1, 19) = 1.60 $p > .05$	**Down-Diagonal** F(1, 19) = 0.16 $p > .05$	**Downwards** F(1, 19) = 0.28 $p > .05$
60°	2-way Interaction Posture, Direction	F(2.40, 45.61) = 1.02, $p > .05$				
	Direction Main Effect	**Sitting** F(2.52, 47.80) = 3.66, $p < .05$			**Standing** F(2.36, 44.88) = 3.05, $p < .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.22 $p > .05$	**Up-Diagonal** F(1, 19) = 8.50 $p < .05$	**Horizontal** F(1, 19) = 26.97 $p < .001$	**Down-Diagonal** F(1, 19) = 18.81 $p < .001$	**Downwards** F(1, 19) = 1.48 $p > .05$
80°	2-way Interaction Posture, Direction	F(4, 76) = 38.49, $p < .001$				
	Direction Main Effect	**Sitting** F(2.74, 51.98) = 10.22, p ¡ .001			**Standing** F(2.71, 51.56) = 26.95, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 2.31 $p > .05$	**Up-Diagonal** F(1, 19) = 45.76 $p < .001$	**Horizontal** F(1, 19) = 47.23 $p < .001$	**Down-Diagonal** F(1, 19) = 165.96 $p < .001$	**Downwards** F(1, 19) = 3.98 $p > .05$
100°	2-way Interaction Posture, Direction	F(4, 76) = 44.92, $p < .001$				
	Direction Main Effect	**Sitting** F(2.73, 51.82) = 3.82, $p < .05$			**Standing** F(2.535, 48.165) = 47.792, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 1.32 $p > .05$	**Up-Diagonal** F(1, 19) = 164.62 $p < .001$	**Horizontal** F(1, 19) = 168.92 $p < .001$	**Down-Diagonal** F(1, 19) = 187.83 $p < .001$	**Downwards** F(1, 19) = 0.09 $p > .05$

A.3 Torso Contribution

Table A.5: Repeated measures ANOVA statistical analysis on the effect of amplitude on torso contribution.

Torso Contribution Amplitude

All Amplitudes	3-way Interaction	F(5.62, 106.70) = 36.59, $p < .001$				
	2-way Interaction Amplitude, Direction	**Sitting** F(4.09, 77.61) = 1.64, $p > .05$			**Standing** F(4.96, 93.32) = 42.19, $p < .001$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(2.83, 53.73) = 1.86 $p > .05$	F(2.04, 38.67) = 56.78 $p < .001$	F(1.82, 36.64) = 78.07 $p < .001$	F(2.21, 41.99) = 123.64 $p < .001$	F(2.56, 48.70) = 0.64 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(2.32, 44.06) = 13.09 $p < .001$	F(1.77, 33.53) = 12.86 $p < .001$	F(1.74, 33.09) = 12.18 $p < .001$	F(1.32, 25.11) = 12.38 $p < .001$	F(2.27, 43.14) = 9.18 $p < .001$
	Amplitude Main Effect, Standing Posture	F(2.36, 44.76) = 6.79 $p < .05$	F(1.79, 34.04) = 59.01 $p < .001$	F(1.66, 31.50) = 90.02 $p < .001$	F(1.69, 32.10) = 179.34 $p < .001$	F(2.96, 56.31) = 12.05 $p < .001$
Within-view	3-way Interaction	F(4.52, 85.95) = 1.34, $p > .05$				
	2-way Interaction Amplitude, Direction	**Sitting** F(2.73, 51.83) = 1.38, $p > .05$			**Standing** F(3.93, 74.71) = 1.77, $p > .05$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(1.98, 37.64) = 2.12 $p > .05$	F(1.62, 30.85) = 1.11 $p > .05$	F(1.84, 35.03) = 2.65 $p > .05$	F(1.44, 46.34) = 3.13 $p > .05$	F(1.73, 32.87) = 0.69 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(1.25, 23.79) = 16.00 $p < .001$	F(1.14, 21.71) = 30.83 $p < .001$	F(1.72, 32.61) = 26.45 $p < .001$	F(1.51, 28.69) = 33.27 $p < .001$	F(1.16, 22.09) = 9.27 $p < .05$
	Amplitude Main Effect, Standing Posture	F(1.60, 30.35) = 41.84 $p < .001$	F(1.59, 30.29) = 24.21 $p < .001$	F(1.42, 26.95) = 33.96 $p < .001$	F(1.33, 25.33) = 25.85 $p < .001$	F(1.58, 29.98) = 11.77 $p < .001$
Beyond-view	3-way Interaction	F(3.80, 72.24) = 17.99, $p < .001$				
	2-way Interaction Amplitude, Direction	**Sitting** F(3.12, 59.36) = 1.44, $p > .05$			**Standing** F(3.84, 72.89) = 22.98, $p < .001$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(1.12, 21.23) = 2.03 $p > .05$	F(2, 38) = 56.17 $p < .001$	F(2, 38) = 56.84 $p < .001$	F(1.39, 26.49) = 61.35 $p < .001$	F(1.35, 25.58) = 0.64 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(1.17, 22.25) = 19.44 $p < .001$	F(1.52, 28.86) = 18.60 $p < .001$	F(1.68, 31.88) = 21.26 $p < .001$	F(1.28, 24.36) = 16.22 $p < .001$	F(1.28, 24.32) = 24.32 $p < .05$
	Amplitude Main Effect, Standing Posture	F(1.07, 20.39) = 11.59 $p < .05$	F(2, 38) = 78.53 $p < .001$	F(1.44, 27.30) = 91.44 $p < .001$	F(2, 38) = 206.04 $p < .001$	F(2, 38) = 15.51 $p < .001$

Table A.6: Repeated measures ANOVA statistical analysis on the effect of posture and direction on torso contribution.

Torso Contribution Posture and Direction

5°	2-way Interaction Posture, Direction	F(4, 76) = 1.24, $p > .05$				
	Direction Main Effect	**Sitting** F(2.16, 40.94) = 1.28, $p > .05$			**Standing** F(2.63, 49.88) = 2.18, $p > .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.13 $p > .05$	**Up-Diagonal** F(1, 19) = 3.41 $p > .05$	**Horizontal** F(1, 19) = 3.68 $p > .05$	**Down-Diagonal** F(1, 19) = 3.63 $p > .05$	**Downwards** F(1, 19) = 1.05 $p > .05$
15°	2-way Interaction Posture, Direction	F(4, 76) = 2.19, $p > .05$				
	Direction Main Effect	**Sitting** F(4, 76) = 1.25, $p > .05$			**Standing** F(2.79, 53.00) = 1.50, $p > .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 2.35 $p > .05$	**Up-Diagonal** F(1, 19) = 4.34 $p > .05$	**Horizontal** F(1, 19) = 0.09 $p > .05$	**Down-Diagonal** F(1, 19) = 3.98 $p > .05$	**Downwards** F(1, 19) = 0.09 $p > .05$
25°	2-way Interaction Posture, Direction	F(2.49, 47.32) = 0.77, $p > .05$				
	Direction Main Effect	**Sitting** F(2.82, 53.55) = 2.15, $p > .05$			**Standing** F(2.50, 47.40) = 1.34, $p > .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 2.54 $p > .05$	**Up-Diagonal** F(1, 19) = 1.06 $p > .05$	**Horizontal** F(1, 19) = 0.03 $p > .05$	**Down-Diagonal** F(1, 19) = 0.01 $p > .05$	**Downwards** F(1, 19) = 0.18 $p > .05$
35°	2-way Interaction Posture, Direction	F(1.25, 23.81) = 2.46, $p > .05$				
	Direction Main Effect	**Sitting** F(4, 76) = 1.81, $p > .05$			**Standing** F(1.18, 22.46) = 2.02, $p > .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.01 $p > .05$	**Up-Diagonal** F(1, 19) = 3.37 $p > .05$	**Horizontal** F(1, 19) = 9.70 $p > .05$	**Down-Diagonal** F(1, 19) = 4.81 $p > .05$	**Downwards** F(1, 19) = 0.73 $p > .05$
45°	2-way Interaction Posture, Direction	F(1.51, 28.76) = 1.10, $p > .05$				
	Direction Main Effect	**Sitting** F(2.29, 43.56) = 1.18, $p > .05$			**Standing** F(1.50, 28.54) = 1.35, $p > .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 3.22 $p > .05$	**Up-Diagonal** F(1, 19) = 5.93 $p < .05$	**Horizontal** F(1, 19) = 4.95 $p < .05$	**Down-Diagonal** F(1, 19) = 5.20 $p < .05$	**Downwards** F(1, 19) = 1.72 $p > .05$
60°	2-way Interaction Posture, Direction	F(2.05, 38.97) = 37.97, $p < .001$				
	Direction Main Effect	**Sitting** F(2.09, 39.74) = 2.02, $p > .05$			**Standing** F(1.90, 36.04) = 31.41, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 3.67 $p > .05$	**Up-Diagonal** F(1, 19) = 39.06 $p < .001$	**Horizontal** F(1, 19) = 53.03 $p < .001$	**Down-Diagonal** F(1, 19) = 80.64 $p < .001$	**Downwards** F(1, 19) = 0.72 $p > .05$
80°	2-way Interaction Posture, Direction	F(2.20, 41.88) = 58.12, $p < .001$				
	Direction Main Effect	**Sitting** F(1.69, 32.04) = 1.77, $p > .05$			**Standing** F(2.04, 38.75) = 57.11, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 1.64 $p > .05$	**Up-Diagonal** F(1, 19) = 58.68 $p < .001$	**Horizontal** F(1, 19) = 71.65 $p < .001$	**Down-Diagonal** F(1, 19) = 191.68 $p < .001$	**Downwards** F(1, 19) = 0.08 $p > .05$
100°	2-way Interaction Posture, Direction	F(2.79, 53.09) = 81.60, $p < .001$				
	Direction Main Effect	**Sitting** F(2.21, 42.00) = 1.68, $p > .05$			**Standing** F(2.44, 46.41) = 81.31, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 2.88 $p > .05$	**Up-Diagonal** F(1, 19) = 150.15 $p < .001$	**Horizontal** F(1, 19) = 194.13 $p < .001$	**Down-Diagonal** F(1, 19) = 201.75 $p < .001$	**Downwards** F(1, 19) = 0.38 $p > .05$

Appendix B

Total movements

B.1 Total Head Movement

Table B.1: Repeated measures ANOVA statistical analysis on the effect of amplitude on total head movement.

Total Head Movement Amplitude

All Amplitudes	3-way Interaction	F(6.39, 121.32) = 1.80, $p > .05$				
	2-way Interaction Amplitude, Direction	**Sitting** F(7.35, 139.64) = 6.69, $p < .001$			**Standing** F(5.12, 97.22) = 4.93, $p < .001$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(3.24, 61.61) = 2.22 $p > .05$	F(2.42, 46.03) = 1.37 $p > .05$	F(1.12, 53.84) = 1.12 $p > .05$	F(2.05, 38.88) = 0.51 $p > .05$	F(2.75, 52.31) = 1.48 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(1.75, 33.15) = 5.30 $p < .05$	F(1.41, 26.71) = 7.95 $p < .05$	F(1.74, 33.01) = 17.64 $p < .001$	F(1.58, 30.04) = 27.86 $p < .001$	F(2.29, 43.54) = 24.49 $p < .001$
	Amplitude Main Effect, Standing Posture	F(1.76, 33.41) = 3.38 $p < .05$	F(1.49, 28.29) = 5.14 $p < .05$	F(1.96, 37.20) = 19.28 $p < .001$	F(1.93, 36.58) = 19.22 $p < .001$	F(2.17, 41.21) = 14.29 $p < .001$
Within-view	3-way Interaction	F(6.05, 114.93) = 1.65, $p > .05$				
	2-way Interaction Amplitude, Direction	**Sitting** F(6.48, 123.09) = 1.10, $p > .05$			**Standing** F(4.95, 94.07) = 1.54, $p > .05$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(2.59, 49.26) = 2.22 $p > .05$	F(1.90, 36.18) = 1.47 $p > .05$	F(2.51, 47.71) = 1.09 $p > .05$	F(1.56, 29.71) = 0.22 $p > .05$	F(2.45, 46.62) = 1.59 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(2.24, 42.46) = 1.09 $p > .05$	F(2.58, 26.80) = 2.58 $p > .05$	F(1.75, 33.28) = 1.68 $p > .05$	F(1.53, 29.10) = 3.40 $p > .05$	F(2.06, 39.04) = 2.71 $p > .05$
	Amplitude Main Effect, Standing Posture	F(1.59, 30.16) = 2.98 $p > .05$	F(1.43, 27.15) = 3.10 $p > .05$	F(1.94, 36.94) = 3.14 $p > .05$	F(1.61, 30.53) = 2.61 $p > .05$	F(1.86, 35.33) = 2.53 $p > .05$
Beyond-view	3-way Interaction	F(3.72, 70.65) = 2.85, $p < .05$				
	2-way Interaction Amplitude, Direction	**Sitting** F(3.31, 62.85) = 8.96, $p < .001$			**Standing** F(3.82, 72.51) = 1.74, $p > .05$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(2, 38) = 1.18 $p > .05$	F(1.45, 28.19) = 6.99 $p < .05$	F(2, 38) = 6.30 $p < .05$	F(1.40, 26.57) = 4.05 $p < .05$	F(2, 38) = 1.20 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(2, 38) = 3.13 $p > .05$	F(2, 38) = 14.53 $p < .001$	F(1.39, 26.37) = 48.27 $p < .001$	F(1.24, 23.56) = 9.46 $p < .05$	F(2, 38) = 2.88 $p > .05$
	Amplitude Main Effect, Standing Posture	F(1.39, 26.33) = 1.20 $p > .05$	F(1.29, 24.51) = 0.33 $p > .05$	F(1.33, 25.26) = 4.14 $p > .05$	F(1.44, 27.27) = 1.07 $p > .05$	F(1.36, 25.91) = 1.16 $p > .05$

Table B.2: Repeated measures ANOVA statistical analysis on the effect of posture and direction on total head movement.

		Total Head Movement Posture and Direction				
5°	2-way Interaction Posture, Direction	F(2.47, 46.96) = 0.57, $p > .05$				
	Direction Main Effect	**Sitting** F(2.51, 47.59) = 2.19, $p > .05$			**Standing** F(1.92, 36.39) = 2.71, $p > .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 3.69 $p > .05$	**Up-Diagonal** F(1, 19) = 4.30 $p > .05$	**Horizontal** F(1, 19) = 2.89 $p > .05$	**Down-Diagonal** F(1, 19) = 1.30 $p > .05$	**Downwards** F(1, 19) = 1.28 $p > .05$
15°	2-way Interaction Posture, Direction	F(4, 76) = 2.08, $p > .05$				
	Direction Main Effect	**Sitting** F(2.48, 47.19) = 8.52, $p < .001$			**Standing** F(2.22, 42.17) = 1.22, $p > .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.34 $p > .05$	**Up-Diagonal** F(1, 19) = 2.50 $p > .05$	**Horizontal** F(1, 19) = 0.11 $p > .05$	**Down-Diagonal** F(1, 19) = 3.39 $p > .05$	**Downwards** F(1, 19) = 3.46 $p > .05$
25°	2-way Interaction Posture, Direction	F(4, 76) = 1.00, $p > .05$				
	Direction Main Effect	**Sitting** F(2.22, 42.21) = 8.52, $p < .05$			**Standing** F(2.58, 48.96) = 5.44, $p < .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 1.95 $p > .05$	**Up-Diagonal** F(1, 19) = 3.23 $p > .05$	**Horizontal** F(1, 19) = 0.61 $p > .05$	**Down-Diagonal** F(1, 19) = 4.07 $p > .05$	**Downwards** F(1, 19) = 4.01 $p > .05$
35°	2-way Interaction Posture, Direction	F(4, 76) = 0.40, $p > .05$				
	Direction Main Effect	**Sitting** F(2.37, 45.07) = 9.64, $p < .001$			**Standing** F(2.00, 38.05) = 8.26, $p < .05$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.65 $p > .05$	**Up-Diagonal** F(1, 19) = 2.91 $p > .05$	**Horizontal** F(1, 19) = 1.01 $p > .05$	**Down-Diagonal** F(1, 19) = 4.30 $p > .05$	**Downwards** F(1, 19) = 0.81 $p > .05$
45°	2-way Interaction Posture, Direction	F(2.24, 42.55) = 3.04, $p > .05$				
	Direction Main Effect	**Sitting** F(1.91, 36.27) = 21.77, $p < .001$			**Standing** F(2.12, 40.36) = 15.37, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.18 $p > .05$	**Up-Diagonal** F(1, 19) = 4.92 $p < .05$	**Horizontal** F(1, 19) = 5.65 $p < .05$	**Down-Diagonal** F(1, 19) = 10.45 $p < .05$	**Downwards** F(1, 19) = 0.16 $p > .05$
60°	2-way Interaction Posture, Direction	F(2.19, 41.59) = 2.65, $p > .05$				
	Direction Main Effect	**Sitting** F(2.08, 39.48) = 10.74, $p < .001$			**Standing** F(2.18, 41.55) = 14.33, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.45 $p > .05$	**Up-Diagonal** F(1, 19) = 17.50 $p < .05$	**Horizontal** F(1, 19) = 15.57 $p < .05$	**Down-Diagonal** F(1, 19) = 5.75 $p < .05$	**Downwards** F(1, 19) = 2.76 $p > .05$
80°	2-way Interaction Posture, Direction	F(1.76, 33.49) = 7.08, $p < .05$				
	Direction Main Effect	**Sitting** F(1.52, 28.91) = 4.43, $p < .05$			**Standing** F(2.44, 46.39) = 33.30, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 1.57 $p > .05$	**Up-Diagonal** F(1, 19) = 29.43 $p < .001$	**Horizontal** F(1, 19) = 46.09 $p < .001$	**Down-Diagonal** F(1, 19) = 21.34 $p < .001$	**Downwards** F(1, 19) = 1.21 $p > .05$
100°	2-way Interaction Posture, Direction	F(1.75, 33.19) = 9.20, $p < .05$				
	Direction Main Effect	**Sitting** F(1.71, 32.50) = 4.13, $p < .05$			**Standing** F(2.02,38.45) = 18.78, $p < .001$	
	Posture Main Effect	**Upwards** F(1, 19) = 0.37 $p > .05$	**Up-Diagonal** F(1, 19) = 44.26 $p < .001$	**Horizontal** F(1, 19) = 50.72 $p < .001$	**Down-Diagonal** F(1, 19) = 30.18 $p < .001$	**Downwards** F(1, 19) = 0.08 $p > .05$

B.2 Total Torso Movement

Table B.3: Repeated measures ANOVA statistical analysis on the effect of amplitude on total torso movement.

Total Torso Movement Amplitude

All Amplitudes	3-way Interaction	F(6.19, 117.52) = 18.58, $p < .001$				
	2-way Interaction Amplitude, Direction	**Sitting** F(3.85, 73.05) = 3.32, $p < .05$			**Standing** F(6.09, 115.68) = 21.13, $p < .001$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(2.40, 45.56) = 3.01 $p > .05$	F(2.66, 50.51) = 40.25 $p < .001$	F(2.17, 41.19) = 36.22 $p < .001$	F(2.27, 43.13) = 47.68 $p < .001$	F(2.52, 47.90) = 1.27 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(2.43, 46.15) = 20.00 $p < .001$	F(1.77, 33.70) = 22.72 $p < .001$	F(1.71, 32.43) = 23.05 $p < .001$	F(2.10, 39.85) = 18.66 $p < .001$	F(2.43, 46.24) = 17.19 $p < .001$
	Amplitude Main Effect, Standing Posture	F(1.91, 36.22) = 19.93 $p < .001$	F(2.39, 45.44) = 50.04 $p < .001$	F(1.93, 36.75) = 45.07 $p < .001$	F(2.63, 50.00) = 70.72 $p < .001$	F(2.66, 50.60) = 20.95 $p < .001$
Within-view	3-way Interaction	F(4.53, 86.03) = 1.41, $p > .05$				
	2-way Interaction Amplitude, Direction	**Sitting** F(4.23, 80.44) = 1.13, $p > .05$			**Standing** F(4.35, 82.71) = 1.83, $p > .05$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(1.31, 24.95) = 3.22 $p > .05$	F(2.10, 39.97) = 2.65 $p > .05$	F(1.68, 31.94) = 1.77 $p > .05$	F(1.72, 32.61) = 3.27 $p > .05$	F(1.30, 24.75) = 2.30 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(1.55, 29.38) = 43.41 $p < .001$	F(1.20, 22.78) = 62.63 $p < .001$	F(1.13, 37.27) = 37.27 $p < .001$	F(1.20, 22.74) = 36.60 $p < .001$	F(1.15, 21.92) = 24.32 $p < .001$
	Amplitude Main Effect, Standing Posture	F(1.16, 22.06) = 39.42 $p < .001$	F(2.30, 43.77) = 30.82 $p < .001$	F(1.93, 36.68) = 16.24 $p < .001$	F(1.54, 29.33) = 21.77 $p < .001$	F(1.27, 24.06) = 30.34 $p < .001$
Beyond-view	3-way Interaction	F(3.42, 65.05) = 3.73, $p < .05$				
	2-way Interaction Amplitude, Direction	**Sitting** F(2.11, 40.14) = 3.46, $p < .05$			**Standing** F(3.64, 69.22) = 6.48, $p < .001$	
		Upwards	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
	2-way Interaction Amplitude, Posture	F(1.09, 20.68) = 3.54 $p > .05$	F(1.28, 24.23) = 3.49 $p > .05$	F(1.54, 29.29) = 5.81 $p < .05$	F(1.43, 27.11) = 0.65 $p > .05$	F(1.23, 23.30) = 0.56 $p > .05$
	Amplitude Main Effect, Sitting Posture	F(1.17, 22.17) = 10.76 $p < .05$	F(1.60, 30.35) = 26.42 $p < .001$	F(1.37, 26.00) = 27.54 $p < .001$	F(1.57, 29.82) = 27.17 $p < .001$	F(1.26, 23.85) = 15.82 $p < .001$
	Amplitude Main Effect, Standing Posture	F(1.08, 20.46) = 13.79 $p < .05$	F(1.28, 24.36) = 17.84 $p < .001$	F(2, 38) = 0.14 $p > .05$	F(1.38, 26.15) = 5.01 $p < .05$	F(1.32, 25.16) = 15.37 $p < .001$

Table B.4: Repeated measures ANOVA statistical analysis on the effect of posture and direction on total torso movement.

Total Torso Movement Posture and Direction

5°	2-way Interaction Posture, Direction	F(4, 76) = 0.38, $p > .05$				
	Direction Main Effect	**Sitting** F(4, 76) = 0.87, $p > .05$			**Standing** F(3.02, 57.40) = 0.13, $p > .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 4.23 $p > .05$	F(1, 19) = 26.11 $p < .001$	F(1, 19) = 3.61 $p > .05$	F(1, 19) = 33.76 $p < .001$	F(1, 19) = 4.05 $p > .05$
15°	2-way Interaction Posture, Direction	F(2.12, 40.26) = 5.11, $p < .05$				
	Direction Main Effect	**Sitting** F(4, 76) = 3.39, $p < .05$			**Standing** F(1.95, 37.07) = 4.71, $p < .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 0.15 $p > .05$	F(1, 19) = 8.09 $p < .05$	F(1, 19) = 12.72 $p < .05$	F(1, 19) = 7.44 $p < .05$	F(1, 19) = 3.24 $p > .05$
25°	2-way Interaction Posture, Direction	F(1.60, 30.33) = 3.88, $p < .05$				
	Direction Main Effect	**Sitting** F(2.36, 44.81) = 1.27, $p > .05$			**Standing** F(1.65, 31.28) = 4.04, $p < .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 3.50 $p > .05$	F(1, 19) = 10.68 $p < .05$	F(1, 19) = 10.30 $p < .05$	F(1, 19) = 10.67 $p < .05$	F(1, 19) = 3.69 $p > .05$
35°	2-way Interaction Posture, Direction	F(1.55, 29.47) = 8.17, $p < .05$				
	Direction Main Effect	**Sitting** F(4, 76) = 1.14, $p > .05$			**Standing** F(1.63, 30.88) = 8.74, $p < .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 5.51 $p < .05$	F(1, 19) = 10.87 $p < .05$	F(1, 19) = 14.42 $p < .05$	F(1, 19) = 13.01 $p < .05$	F(1, 19) = 14.91 $p < .05$
45°	2-way Interaction Posture, Direction	F(1.52, 28.78) = 8.85, $p < .05$				
	Direction Main Effect	**Sitting** F(3.09, 58.75) = 1.66, $p > .05$			**Standing** F(1.45, 27.60) = 9.85, $p < .05$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 11.77 $p < .05$	F(1, 19) = 11.59 $p < .05$	F(1, 19) = 13.24 $p < .05$	F(1, 19) = 15.23 $p < .05$	F(1, 19) = 13.24 $p < .05$
60°	2-way Interaction Posture, Direction	F(2.54, 47.57) = 99.97, $p < .001$				
	Direction Main Effect	**Sitting** F(4, 76) = 3.14, $p < .05$			**Standing** F(2.30, 43.60) = 83.92, $p < .001$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 6.14 $p < .05$	F(1, 19) = 84.13 $p < .001$	F(1, 19) = 147.06 $p < .001$	F(1, 19) = 145.96 $p < .001$	F(1, 19) = 14.57 $p < .05$
80°	2-way Interaction Posture, Direction	F(4, 76) = 117.62, $p < .001$				
	Direction Main Effect	**Sitting** F(2.39, 45.34) = 3.05, $p < .05$			**Standing** F(2.07, 39.36) = 96.42, $p < .001$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 8.25 $p < .05$	F(1, 19) = 133.45 $p < .001$	F(1, 19) = 148.69 $p < .001$	F(1, 19) = 273.97 $p < .001$	F(1, 19) = 8.64 $p < .05$
100°	2-way Interaction Posture, Direction	F(2.34, 44.41) = 42.78, p ¡ .001				
	Direction Main Effect	**Sitting** F(1.69, 32.10) = 4.13, $p < .05$			**Standing** F(2.08, 39.52) = 35.44, $p < .001$	
	Posture Main Effect	**Upwards**	**Up-Diagonal**	**Horizontal**	**Down-Diagonal**	**Downwards**
		F(1, 19) = 6.08 $p < .05$	F(1, 19) = 231.92 $p < .001$	F(1, 19) = 221.07 $p < .001$	F(1, 19) = 183.91 $p < .001$	F(1, 19) = 7.30 $p < .05$

References

[1] Johnny Accot and Shumin Zhai. "More than Dotting the i's — Foundations for Crossing-Based Interfaces". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '02. Minneapolis, Minnesota, USA: Association for Computing Machinery, 2002, pp. 73–80. ISBN: 1581134533. DOI: 10.1145/503376.503390.

[2] Sunggeun Ahn, Stephanie Santosa, Mark Parent, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. "StickyPie: A Gaze-Based, Scale-Invariant Marking Menu Optimized for AR/VR". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445297.

[3] Sunggeun Ahn, Jeongmin Son, Sangyoon Lee, and Geehyuk Lee. "Verge-It: Gaze Interaction for a Binocular Head-Worn Display Using Modulated Disparity Vergence Eye Movement". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–7. ISBN: 9781450368193. DOI: 10.1145/3334480.3382908.

[4] Rawan Alghofaili, Yasuhito Sawahata, Haikun Huang, Hsueh-Cheng Wang, Takaaki Shiratori, and Lap-Fai Yu. "Lost in Style: Gaze-driven Adaptive Aid for VR Navigation". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, 348:1–348:12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300578.

[5] Sean Andrist, Tomislav Pejsa, Bilge Mutlu, and Michael Gleicher. "A Head-eye Coordination Model for Animating Gaze Shifts of Virtual Characters". In: *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*. Gaze-In '12. Santa Monica, California: ACM, 2012, 4:1–4:6. ISBN: 978-1-4503-1516-6. DOI: 10.1145/2401836.2401840.

[6] Georg Apitz and François Guimbretière. “CrossY: A Crossing-Based Drawing Application”. In: *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*. UIST ’04. Santa Fe, NM, USA: Association for Computing Machinery, 2004, pp. 3–12. ISBN: 1581139578. DOI: 10.1145/1029632.1029635.

[7] Georg Apitz, François Guimbretière, and Shumin Zhai. “Foundations for Designing and Evaluating User Interfaces Based on the Crossing Paradigm”. In: *ACM Trans. Comput.-Hum. Interact.* 17.2 (May 2008). ISSN: 1073-0516. DOI: 10.1145/1746259.1746263.

[8] Ferran Argelaguet and Carlos Andujar. “A survey of 3D object selection techniques for virtual environments”. In: *Computers & Graphics* 37.3 (May 2013), pp. 121–136. ISSN: 00978493. DOI: 10.1016/j.cag.2012.12.003.

[9] Per Bækgaard, John Paulin Hansen, Katsumi Minakata, and I. Scott MacKenzie. “A Fitts’ Law Study of Pupil Dilations in a Head-Mounted Display”. In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ETRA ’19. Denver, Colorado: Association for Computing Machinery, 2019. ISBN: 9781450367097. DOI: 10.1145/3314111.3319831.

[10] A. Terry Bahill, Michael R. Clark, and Lawrence Stark. “Glissades-eye movements generated by mismatched components of the saccadic motoneuronal control signal”. In: *Mathematical Biosciences* 26.3-4 (1975), pp. 303–318.

[11] Gilles Bailly, Eric Lecolinet, and Laurence Nigay. “Visual Menu Techniques”. In: *ACM Comput. Surv.* 49.4 (Dec. 2016), 60:1–60:41. ISSN: 0360-0300. DOI: 10.1145/3002171.

[12] Dana H. Ballard, Mary M. Hayhoe, Feng Li, Steven D. Whitehead, J.p. Frisby, J. G. Taylor, R. B. Fisher, Horace Basil Barlow, J. P. Frisby, George Adrian Horridge, and M. A. Jeeves. “Hand-eye coordination during sequential tasks”. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 337.1281 (1992), pp. 331–339. DOI: 10.1098/rstb.1992.0111.

[13] Albert E. Bartz. “Eye and Head Movements in Peripheral Vision: Nature of Compensatory Eye Movements”. In: *Science* 152.3729 (1966), pp. 1644–1645. ISSN: 0036-8075. DOI: 10.1126/science.152.3729.1644.

[14] Michael Barz, Florian Daiber, Daniel Sonntag, and Andreas Bulling. “Error-Aware Gaze-Based Interfaces for Robust Mobile Gaze Interaction”. In: *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ETRA ’18. Warsaw, Poland: Association for Computing Machinery, 2018. ISBN: 9781450357067. DOI: 10.1145/3204493.3204536.

[15] Richard Bates and Howell Istance. "Why are Eye Mice Unpopular? A Detailed Comparison of Head and Eye Controlled Assistive Technology Pointing Devices". In: *Universal Access in the Information Society* 2.3 (Oct. 2003), pp. 280–290. ISSN: 1615-5289. DOI: 10.1007/s10209-003-0053-y.

[16] Hrvoje Benko, Andrew D. Wilson, and Patrick Baudisch. "Precise Selection Techniques for Multi-Touch Screens". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '06. Montréal, Québec, Canada: Association for Computing Machinery, 2006, pp. 1263–1272. ISBN: 1595933727. DOI: 10.1145/1124772.1124963.

[17] Emilio Bizzi. "The coordination of eye-head movements". In: *Scientific American* 231.4 (Oct. 1974), pp. 100–109.

[18] Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. "Advantages of Eye-gaze over Head-gaze-based Selection in Virtual and Augmented Reality Under Varying Field of Views". In: *Proceedings of the Workshop on Communication by Gaze Interaction*. COGAIN '18. Warsaw, Poland: ACM, 2018, 1:1–1:9. ISBN: 978-1-4503-5790-6. DOI: 10.1145/3206343.3206349.

[19] Richard A. Bolt. "Gaze-orchestrated Dynamic Windows". In: *Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '81. Dallas, Texas, USA: ACM, 1981, pp. 109–119. ISBN: 0-89791-045-1. DOI: 10.1145/800224.806796.

[20] Riccardo Bovo, Daniele Giunchi, Ludwig Sidenmark, Hans Gellersen, Enrico Costanza, and Thomas Heinis. "Real-Time Head-Based Deep-Learning Model for Gaze Probability Regions in Collaborative VR". In: *2022 Symposium on Eye Tracking Research and Applications*. ETRA '22. Seattle, WA, USA: Association for Computing Machinery, 2022. ISBN: 9781450392525. DOI: 10.1145/3517031.3529642. URL: https://doi.org/10.1145/3517031.3529642.

[21] Andreas Bulling, Daniel Roggen, and Gerhard Tröster. "Wearable EOG Goggles: Eye-Based Interaction in Everyday Environments". In: *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '09. Boston, MA, USA: Association for Computing Machinery, 2009, pp. 3259–3264. ISBN: 9781605582474. DOI: 10.1145/1520340.1520468.

[22] J. Callahan, D. Hopkins, M. Weiser, and B. Shneiderman. "An Empirical Comparison of Pie vs. Linear Menus". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '88. Washington, D.C., USA: ACM, 1988, pp. 95–100. ISBN: 0-201-14237-6. DOI: 10.1145/57167.57182.

[23] Xiang Cao, Jacky Jie Li, and Ravin Balakrishnan. “Peephole Pointing: Modeling Acquisition of Dynamically Revealed Targets”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’08. Florence, Italy: ACM, 2008, pp. 1699–1708. ISBN: 978-1-60558-011-1. DOI: 10.1145/1357054.1357320.

[24] Stuart K. Card, Thomas P. Moran, and Allen Newell. “The Keystroke-level Model for User Performance Time with Interactive Systems”. In: *Commun. ACM* 23.7 (July 1980), pp. 396–410. ISSN: 0001-0782. DOI: 10.1145/358886.358895.

[25] Debaleena Chattopadhyay and Davide Bolchini. “Touchless Circular Menus: Toward an Intuitive UI for Touchless Interactions with Large Displays”. In: *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. AVI ’14. Como, Italy: Association for Computing Machinery, 2014, pp. 33–40. ISBN: 9781450327756. DOI: 10.1145/2598153.2598181.

[26] Jixu Chen and Qiang Ji. “Probabilistic gaze estimation without active personal calibration”. In: *CVPR 2011*. 2011, pp. 609–616. DOI: 10.1109/CVPR.2011.5995675.

[27] Myungguen Choi, Daisuke Sakamoto, and Tetsuo Ono. “Kuiper Belt: Utilizing the “Out-of-Natural Angle” Region in the Eye-Gaze Interaction for Virtual Reality”. In: *CHI Conference on Human Factors in Computing Systems*. CHI ’22. New Orleans, LA, USA: Association for Computing Machinery, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3517725.

[28] Hugo Chow-Wing-Bom, Tessa M. Dekker, and Pete R. Jones. “The worse eye revisited: Evaluating the impact of asymmetric peripheral vision loss on everyday function”. In: *Vision Research* 169 (2020), pp. 49–57. ISSN: 0042-6989. DOI: 10.1016/j.visres.2019.10.012. URL: https://www.sciencedirect.com/science/article/pii/S0042698920300304.

[29] Nathan Cournia, John D. Smith, and Andrew T. Duchowski. “Gaze- vs. Hand-Based Pointing in Virtual Environments”. In: *CHI ’03 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’03. Ft. Lauderdale, Florida, USA: Association for Computing Machinery, 2003, pp. 772–773. ISBN: 1581136374. DOI: 10.1145/765891.765982.

[30] K. Das and C. W. Borst. “An evaluation of menu properties and pointing techniques in a projection-based VR environment”. In: *2010 IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2010, pp. 47–50. DOI: 10.1109/3DUI.2010.5444721.

[31] Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. "Towards Gaze-Based Prediction of the Intent to Interact in Virtual Reality". In: *ACM Symposium on Eye Tracking Research and Applications*. ETRA '21 Short Papers. Virtual Event, Germany: Association for Computing Machinery, 2021. ISBN: 9781450383455. DOI: 10.1145/3448018.3458008.

[32] Matthew M. Davis, Joseph L. Gabbard, Doug A. Bowman, and Dennis Gracanin. "Depth-based 3D gesture multi-level radial menu for virtual object manipulation". In: *2016 IEEE Virtual Reality (VR)*. IEEE, Mar. 2016, pp. 169–170. DOI: 10.1109/VR.2016.7504707.

[33] William Delamare, Teng Han, and Pourang Irani. "Designing a Gaze Gesture Guiding System". In: *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '17. Vienna, Austria: Association for Computing Machinery, 2017. ISBN: 9781450350754. DOI: 10.1145/3098279.3098561.

[34] Anup Doshi and Mohan M. Trivedi. "Head and eye gaze dynamics during visual attention shifts in complex environments". In: *Journal of Vision* 12.2 (2012), p. 9. DOI: 10.1167/12.2.9.

[35] Heiko Drewes, Alexander De Luca, and Albrecht Schmidt. "Eye-Gaze Interaction for Mobile Phones". In: *Proceedings of the 4th International Conference on Mobile Technology, Applications, and Systems and the 1st International Symposium on Computer Human Interaction in Mobile Technology*. Mobility '07. Singapore: Association for Computing Machinery, 2007, pp. 364–371. ISBN: 9781595938190. DOI: 10.1145/1378063.1378122.

[36] Heiko Drewes, Mohamed Khamis, and Florian Alt. "Smooth Pursuit Target Speeds and Trajectories". In: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*. MUM 2018. Cairo, Egypt: Association for Computing Machinery, 2018, pp. 139–146. ISBN: 9781450365949. DOI: 10.1145/3282894.3282913.

[37] Heiko Drewes and Albrecht Schmidt. "Interacting with the Computer Using Gaze Gestures". In: *Human-Computer Interaction – INTERACT 2007*. Ed. by Cécilia Baranauskas, Philippe Palanque, Julio Abascal, and Simone Diniz Junqueira Barbosa. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 475–488. ISBN: 978-3-540-74800-7. DOI: 10.1007/978-3-540-74800-7_43.

[38] Andrew Duchowski and Roel Vertegaal. "Eye-based interaction in graphical systems : Theory & practice". In: *ACM SIGGRAPH 2000 Course Notes* (2000).

[39] Wolfgang Einhäuser, Frank Schumann, Stanislavs Bardins, Klaus Bartl, Guido Böning, Erich Schneider, and Peter König. "Human eye-head co-ordination in natural exploration". In: *Network: Computation in Neural Systems* 18.3 (Jan. 2007), pp. 267–297. ISSN: 0954898X. DOI: 10.1080/09548980701671094.

[40] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. "Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements". In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST '15. Charlotte, NC, USA: Association for Computing Machinery, 2015, pp. 457–466. ISBN: 9781450337793. DOI: 10.1145/2807442.2807499.

[41] João Marcelo Evangelista Belo, Anna Maria Feit, Tiare Feuchtner, and Kaj Grønbæk. "XRgonomics: Facilitating the Creation of Ergonomic 3D Interfaces". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445349. URL: https://doi.org/10.1145/3411764.3445349.

[42] Ribel Fares, Shaomin Fang, and Oleg Komogortsev. "Can We Beat the Mouse with MAGIC?" In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. Paris, France: Association for Computing Machinery, 2013, pp. 1387–1390. ISBN: 9781450318990. DOI: 10.1145/2470654.2466183.

[43] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. "Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design". In: CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 1118–1130. ISBN: 9781450346559. DOI: 10.1145/3025453.3025599.

[44] Virgilio F. Ferrario, Chiarella Sforza, Graziano Serrao, GianPiero Grassi, and Erio Mossi. "Active range of motion of the head and cervical spine: a three-dimensional investigation in healthy young adults". In: *Journal of Orthopaedic Research* 20.1 (Jan. 2002), pp. 122–129. ISSN: 0736-0266. DOI: 10.1016/S0736-0266(01)00079-1.

[45] Edward G. Freedman. "Coordination of the eyes and head during visual orienting". In: *Experimental Brain Research* 190.4 (Oct. 2008), pp. 369–387. ISSN: 00144819. DOI: 10.1007/s00221-008-1504-8.

[46] Edward G. Freedman and David L. Sparks. "Coordination of the eyes and head: movement kinematics". In: *Experimental Brain Research* 131.1 (Mar. 2000), pp. 22–32. ISSN: 0014-4819. DOI: 10.1007/s002219900296.

[47] James H. Fuller. "Head movement propensity". In: *Experimental Brain Research* 92.1 (Dec. 1992), pp. 152–164. ISSN: 0014-4819. DOI: 10.1007/BF00230391.

[48] Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M. Angela Sasse. "The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '03. Ft. Lauderdale, Florida, USA: Association for Computing Machinery, 2003, pp. 529–536. ISBN: 1581136307. DOI: 10.1145/642611.642703.

[49] Christoph Gebhardt, Brian Hecox, Bas van Opheusden, Daniel Wigdor, James Hillis, Otmar Hilliges, and Hrvoje Benko. "Learning Cooperative Personalized Policies from Gaze Data". In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. UIST '19. New Orleans, LA, USA: Association for Computing Machinery, 2019, pp. 197–208. ISBN: 9781450368162. DOI: 10.1145/3332165.3347933.

[50] S. Gebhardt, S. Pick, F. Leithold, B. Hentschel, and T. Kuhlen. "Extended Pie Menus for Immersive Virtual Environments". In: *IEEE Transactions on Visualization and Computer Graphics* 19.4 (Apr. 2013), pp. 644–651. ISSN: 2160-9306. DOI: 10.1109/TVCG.2013.31.

[51] D. Gerber and D. Bechmann. "The spin menu: a menu system for virtual environments". In: *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* Mar. 2005, pp. 271–272. DOI: 10.1109/VR.2005.1492790.

[52] H. H. L. M. Goossens and A. John Van Opstal. "Human eye-head coordination in two dimensions under different sensorimotor conditions". In: *Experimental Brain Research* 114.3 (May 1997), pp. 542–560. ISSN: 0014-4819. DOI: 10.1007/PL00005663.

[53] Tovi Grossman and Ravin Balakrishnan. "The Bubble Cursor: Enhancing Target Acquisition by Dynamic Resizing of the Cursor's Activation Area". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. Portland, Oregon, USA: ACM, 2005, pp. 281–290. ISBN: 1-58113-998-5. DOI: 10.1145/1054972.1055012.

[54] Daniel Guitton and Michel Volle. "Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range". In: *Journal of neurophysiology* 58.3 (1987), pp. 427–459.

[55] John Paulin Hansen, Anders Sewerin Johansen, Dan Witzner Hansen, Kenji Itoh, and Satoru Mashino. "Command Without a Click: Dwell Time Typing by Mouse and Gaze Selections". In: *Human-Computer Interaction - INTERACT 2003*. Amsterdam, The Netherlands: IOS Press, 2003, pp. 121–128.

[56] John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. "A Fitts' Law Study of Click and Dwell Interaction by Gaze, Head and Mouse with a Head-mounted Display". In: *Proceedings of the Workshop on Communication by Gaze Interaction*. COGAIN '18. Warsaw, Poland: ACM, 2018, 7:1–7:5. ISBN: 978-1-4503-5790-6. DOI: 10.1145/3206343.3206344.

[57] Craig Hennessey, Borna Noureddin, and Peter Lawrence. "A Single Camera Eye-gaze Tracking System with Free Head Motion". In: *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications*. ETRA '06. San Diego, California: ACM, 2006, pp. 87–94. ISBN: 1-59593-305-0. DOI: 10.1145/1117309.1117349.

[58] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. "Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 1063–1072. ISBN: 9781450324731. DOI: 10.1145/2556288.2557130. URL: https://doi.org/10.1145/2556288.2557130.

[59] Teresa Hirzle, Fabian Fischbach, Julian Karlbauer, Pascal Jansen, Jan Gugenheimer, Enrico Rukzio, and Andreas Bulling. "Understanding, Addressing, and Analysing Digital Eye Strain in Virtual Reality Head-Mounted Displays". In: *ACM Trans. Comput.-Hum. Interact.* 29.4 (Mar. 2022). ISSN: 1073-0516. DOI: 10.1145/3492802. URL: https://doi.org/10.1145/3492802.

[60] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. "A Design Space for Gaze Interaction on Head-Mounted Displays". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300855.

[61] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Jarodzka Halszka, and Joost van de Weijer. *Eye Tracking : A Comprehensive Guide to Methods and Measures*. eng. Oxford University Press, 2011, p. 560. ISBN: 9780199697083.

[62] Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. "Eye Tracker Data Quality: What It is and How to Measure It". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. Santa Barbara, California: Association for Computing Machinery, 2012, pp. 45–52. ISBN: 9781450312219. DOI: 10.1145/2168556.2168563.

[63] Don Hopkins. "The Design and Implementation of Pie Menus". In: *Dr. Dobb's J.* 16.12 (Dec. 1991), pp. 16–26. ISSN: 1044-789X.

[64] Brian Hu, Ishmael Johnson-Bey, Mansi Sharma, and Ernst Niebur. "Head movements are correlated with other measures of visual attention at smaller spatial scales". In: *Information Sciences and Systems (CISS), 2018 52nd Annual Conference on*. IEEE, 2018, pp. 1–6. DOI: 10.1109/CISS.2018.8362264.

[65] Brian Hu, Ishmael Johnson-bey, Mansi Sharma, and Ernst Niebur. "Head movements during visual exploration of natural images in virtual reality". In: *2017 51st Annual Conference on Information Sciences and Systems (CISS)*. IEEE, Apr. 2017, pp. 1–6. ISBN: 9781509047802. DOI: 10.1109/CISS.2017.7926138.

[66] Anke Huckauf and Mario H. Urbina. "Gazing with pEYEs: Towards a Universal Input for Various Applications". In: *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*. ETRA '08. Savannah, Georgia: ACM, 2008, pp. 51–54. ISBN: 978-1-59593-982-1. DOI: 10.1145/1344471.1344483.

[67] Felix Hülsmann, Timo Dankert, and Thies Pfeiffer. "Comparing gaze-based and manual interaction in a fast-paced gaming task in Virtual Reality". In: *Proceedings of the Workshop Virtuelle & Erweiterte Realität 2011*. Ed. by Christian-A. Bohn and Sina Mostafawy. Shaker Verlag, 2011, pp. 1–12.

[68] Thomas E. Hutchinson, K. Preston White, Worthy N. Martin, Kelly C. Reichert, and Lisa A. Frey. "Human-Computer Interaction Using Eye-Gaze Input". In: *IEEE Transactions on Systems, Man, and Cybernetics* 19.6 (Nov. 1989), pp. 1527–1534. ISSN: 0018-9472. DOI: 10.1109/21.44068.

[69] Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. "Gaze Gestures or Dwell-based Interaction?" In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. Santa Barbara, California: ACM, 2012, pp. 229–232. ISBN: 978-1-4503-1221-9. DOI: 10.1145/2168556.2168602.

[70] Aulikki Hyrskykari, Päivi Majaranta, Antti Aaltonen, and Kari-Jouko Räihä. "Design Issues of IDICT: A Gaze-Assisted Translation Aid". In: *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*. ETRA '00. Palm Beach Gardens, Florida, USA: Association for Computing Machinery, 2000, pp. 9–14. ISBN: 1581132808. DOI: 10.1145/355017.355019.

[71] *Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices*. Standard. International Organization for Standardization, Feb. 2000.

[72] Laurent Itti, Nitin Dhavale, and Fréderic Pighin. "Photorealistic attention-based gaze animation". In: *2006 IEEE International Conference on Multimedia and Expo*. IEEE, 2006, pp. 521–524. DOI: 10.1109/ICME.2006.262440.

[73] Rob Jacob and Sophie Stellmach. "What You Look at is What You Get: Gaze-Based User Interfaces". In: *Interactions* 23.5 (Aug. 2016), pp. 62–65. ISSN: 1072-5520. DOI: 10.1145/2978577.

[74] Robert J. K. Jacob. "What You Look at is What You Get: Eye Movement-based Interaction Techniques". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '90. Seattle, Washington, USA: ACM, 1990, pp. 11–18. ISBN: 0-201-50932-6. DOI: 10.1145/97243.97246.

[75] Robert J. K. Jacob. "Eye movement-based human-computer interaction techniques: Toward non-command interfaces". In: *Advances in human-computer interaction* 4 (1993), pp. 151–190.

[76] Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. "MAGIC Pointing for Eyewear Computers". In: *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ISWC '15. Osaka, Japan: ACM, 2015, pp. 155–158. ISBN: 978-1-4503-3578-2. DOI: 10.1145/2802083.2802094.

[77] Shahram Jalaliniya, Mardanbegi, Thomas Pederson, and Dan Witzner Hansen. "Head and Eye Movement as Pointing Modalities for Eyewear Computers". In: *2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops*. IEEE, June 2014, pp. 50–53. DOI: 10.1109/BSN.Workshops.2014.14.

[78] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. "Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 3328–3339. ISBN: 9781450346559. DOI: 10.1145/3025453.3025523. URL: https://doi.org/10.1145/3025453.3025523.

[79] Allison Jing, Kieran William May, Mahnoor Naeem, Gun Lee, and Mark Billinghurst. "EyemR-Vis: Using Bi-Directional Gaze Behavioural Cues to Improve Mixed Reality Remote Collaboration". In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021. ISBN: 9781450380959. DOI: 10.1145/3411763.3451844.

[80] Florian Jungwirth, Michael Haslgrübler, and Alois Ferscha. "Contour-Guided Gaze Gestures: Using Object Contours as Visual Guidance for Triggering Interactions". In: *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ETRA '18. Warsaw, Poland: Association for Computing Machinery, 2018. ISBN: 9781450357067. DOI: 10.1145/3204493.3204530.

[81] Jari Kangas, Deepak Akkil, Jussi Rantala, Poika Isokoski, Päivi Majaranta, and Roope Raisamo. "Gaze Gestures and Haptic Feedback in Mobile Devices". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 435–438. ISBN: 9781450324731. DOI: 10.1145/2556288.2557040.

[82] Christina Katsini, Yasmeen Abdrabou, George E. Raptis, Mohamed Khamis, and Florian Alt. "The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–21. ISBN: 9781450367080. DOI: 10.1145/3313831.3376840. URL: https://doi.org/10.1145/3313831.3376840.

[83] Mohamed Khamis, Carl Oechsner, Florian Alt, and Andreas Bulling. "VRpursuits: Interaction in Virtual Reality Using Smooth Pursuit Eye Movements". In: *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*. AVI '18. Castiglione della Pescaia, Grosseto, Italy: ACM, 2018, 18:1–18:8. ISBN: 978-1-4503-5616-9. DOI: 10.1145/3206505.3206522.

[84] Tobit Kollenberg, Alexander Neumann, Dorothe Schneider, Tessa-Karina Tews, Thomas Hermann, Helge Ritter, Angelika Dierker, and Hendrik Koesling. "Visual search in the (un)real world: how head-mounted displays affect eye movements, head movements and target detection". In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ETRA '10. Austin, Texas: ACM, 2010, pp. 121–124. ISBN: 978-1-60558-994-7. DOI: 10.1145/1743666.1743696.

[85] Manu Kumar, Terry Winograd, and Andreas Paepcke. "Gaze-Enhanced Scrolling Techniques". In: *CHI '07 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '07. San Jose, CA, USA: Association for Computing Machinery, 2007, pp. 2531–2536. ISBN: 9781595936424. DOI: 10.1145/1240866.1241036.

[86] Andrew Kurauchi, Wenxin Feng, Carlos Morimoto, and Margrit Betke. "HMAGIC: Head Movement and Gaze Input Cascaded Pointing". In: *Proceedings of the 8th ACM International Conference on PErvasive Technologies*

Related to Assistive Environments. PETRA '15. Corfu, Greece: ACM, 2015, 47:1–47:4. ISBN: 978-1-4503-3452-5. DOI: 10.1145/2769493.2769550.

[87] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. "Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: ACM, 2018, 81:1–81:14. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173655.

[88] Michael Land and Benjamin Tatler. *Looking and Acting: Vision and Eye Movements in Natural Behaviour*. United Kingdom: Oxford University Press, 2012. ISBN: 9780198570943. DOI: 10.1093/acprof:oso/9780198570943.001.0001.

[89] Michael F. Land. "The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations". In: *Experimental Brain Research* 159.2 (Nov. 2004), pp. 151–160. ISSN: 1432-1106. DOI: 10.1007/s00221-004-1951-9.

[90] Michael F. Land, Neil Mennie, and Jennifer Rusted. "The Roles of Vision and Eye Movements in the Control of Activities of Daily Living". In: *Perception* 28.11 (1999). PMID: 10755142, pp. 1311–1328. DOI: 10.1068/p2935.

[91] Chris Lankford. "Effective Eye-gaze Input into Windows". In: *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*. ETRA '00. Palm Beach Gardens, Florida, USA: ACM, 2000, pp. 23–27. ISBN: 1-58113-280-8. DOI: 10.1145/355017.355021.

[92] Otto Lappi. "Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference". In: *Neuroscience & Biobehavioral Reviews* 69 (2016), pp. 49–68. ISSN: 0149-7634. DOI: 10.1016/j.neubiorev.2016.06.006.

[93] Joseph J. LaViola. "A Discussion of Cybersickness in Virtual Environments". In: *SIGCHI Bull.* 32.1 (Jan. 2000), pp. 47–56. ISSN: 0736-6906. DOI: 10.1145/333329.333344. URL: https://doi.org/10.1145/333329.333344.

[94] Zhen Li, Michelle Annett, Ken Hinckley, Karan Singh, and Daniel Wigdor. "HoloDoc: Enabling Mixed Reality Workspaces That Harness Physical and Digital Content". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, 687:1–687:14. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300917.

[95] Jonathan Liebers, Patrick Horn, Christian Burschik, Uwe Gruenefeld, and Stefan Schneegass. "Using Gaze Behavior and Head Orientation for Implicit Identification in Virtual Reality". In: *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. VRST '21. Osaka, Japan: Association for Computing Machinery, 2021. ISBN: 9781450390927. DOI: 10.1145/3489849.3489880. URL: https://doi.org/10.1145/3489849.3489880.

[96] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. "Context-Aware Online Adaptation of Mixed Reality Interfaces". In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. UIST '19. New Orleans, LA, USA: Association for Computing Machinery, 2019, pp. 147–160. ISBN: 9781450368162. DOI: 10.1145/3332165.3347945. URL: https://doi.org/10.1145/3332165.3347945.

[97] Chang Liu, Jason Orlosky, and Alexander Plopski. "Eye Gaze-Based Object Rotation for Head-Mounted Displays". In: *Symposium on Spatial User Interaction*. SUI '20. Virtual Event, Canada: Association for Computing Machinery, 2020. ISBN: 9781450379434. DOI: 10.1145/3385959.3418444.

[98] Chang Liu, Alexander Plopski, and Jason Orlosky. "OrthoGaze: Gaze-based three-dimensional object manipulation using orthogonal planes". In: *Computers & Graphics* 89 (2020), pp. 1–10. ISSN: 0097-8493. DOI: 10.1016/j.cag.2020.04.005.

[99] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. "Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 930–939. DOI: 10.1109/VR46266.2020.00113.

[100] Yuexing Luo and Daniel Vogel. "Pin-and-Cross: A Unimanual Multitouch Technique Combining Static Touches with Crossing Selection". In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST '15. Charlotte, NC, USA: Association for Computing Machinery, 2015, pp. 323–332. ISBN: 9781450337793. DOI: 10.1145/2807442.2807444.

[101] Christof Lutteroth, Moiz Penkar, and Gerald Weber. "Gaze vs. Mouse: A Fast and Accurate Gaze-Only Click Alternative". In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST '15. Charlotte, NC, USA: ACM, 2015, pp. 385–394. ISBN: 978-1-4503-3779-3. DOI: 10.1145/2807442.2807461.

[102] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. "Fast Gaze Typing with an Adjustable Dwell Time". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. Boston, MA, USA: ACM, 2009, pp. 357–360. ISBN: 978-1-60558-246-7. DOI: 10.1145/1518701.1518758.

[103] Diako Mardanbegi, Christopher Clarke, and Hans Gellersen. "Monocular Gaze Depth Estimation Using the Vestibulo-ocular Reflex". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ETRA '19. Denver, Colorado: ACM, 2019, 20:1–20:9. ISBN: 978-1-4503-6709-7. DOI: 10.1145/3314111.3319822.

[104] Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. "Eye-based Head Gestures". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. Santa Barbara, California: ACM, 2012, pp. 139–146. ISBN: 978-1-4503-1221-9. DOI: 10.1145/2168556.2168578.

[105] Diako Mardanbegi, Tobias Langlotz, and Hans Gellersen. "Resolving Target Ambiguity in 3D Gaze Interaction Through VOR Depth Estimation". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, 612:1–612:12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300842.

[106] Diako Mardanbegi, Ken Pfeuffer, Alexander Perzl, Benedikt Mayer, Shahram Jalaliniya, and Hans Gellersen. "EyeSeeThrough: Unifying Tool Selection and Application in Virtual Environments". In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Mar. 2019, pp. 474–483. DOI: 10.1109/VR.2019.8797988.

[107] Sebastian Marwecki, Andrew D. Wilson, Eyal Ofek, Mar Gonzalez Franco, and Christian Holz. "Mise-Unseen: Using Eye Tracking to Hide Virtual Reality Scene Changes in Plain Sight". In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. UIST '19. New Orleans, LA, USA: Association for Computing Machinery, 2019, pp. 777–789. ISBN: 9781450368162. DOI: 10.1145/3332165.3347919.

[108] Ken I. McAnally and Russell L. Martin. "Sound localization with head movement: implications for 3-d audio displays". In: *Frontiers in Neuroscience* 8 (2014). ISSN: 1662-453X. DOI: 10.3389/fnins.2014.00210. URL: https://www.frontiersin.org/articles/10.3389/fnins.2014.00210.

[109] Zeljka Mihajlovic, Sinisa Popovic, Karla Brkic, and Kresimir Cosic. "A system for head-neck rehabilitation exercises based on serious gaming and virtual reality". In: *Multimedia Tools and Applications* 77.15 (Aug. 2018), pp. 19113–19137. ISSN: 1573-7721. DOI: 10.1007/s11042-017-5328-z. URL: https://doi.org/10.1007/s11042-017-5328-z.

[110] Katsumi Minakata, John Paulin Hansen, I. Scott MacKenzie, Per Bækgaard, and Vijay Rajanna. "Pointing by Gaze, Head, and Foot in a Head-Mounted Display". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ETRA '19. Denver, Colorado: Association for

Computing Machinery, 2019. ISBN: 9781450367097. DOI: 10.1145/3317956.3318150.

[111] Mark R Mine. "Virtual environment interaction techniques". In: *UNC Chapel Hill CS Dept* (1995).

[112] Pallavi Mohan, Wooi Boon Goh, Chi-Wing Fu, and Sai-Kit Yeung. "DualGaze: Addressing the Midas Touch Problem in Gaze Mediated VR Interaction". In: *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. Oct. 2018, pp. 79–84. DOI: 10.1109/ISMAR-Adjunct.2018.00039.

[113] Emilie Møllenbach, Martin Lillholm, Alastair Gail, and John Paulin Hansen. "Single Gaze Gestures". In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ETRA '10. Austin, Texas: ACM, 2010, pp. 177–180. ISBN: 978-1-60558-994-7. DOI: 10.1145/1743666.1743710.

[114] Carlos H. Morimoto and Marcio R. M. Mimica. "Eye Gaze Tracking Techniques for Interactive Applications". In: *Computer Vision and Image Understanding* 98.1 (Apr. 2005), pp. 4–24. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2004.07.010.

[115] Orval H. Mowrer. "XXXII. Concerning the Normal Function of the Vestibular Apparatus". In: *Annals of Otology, Rhinology & Laryngology* 41.2 (1932), pp. 412–421. DOI: 10.1177/000348943204100209.

[116] Florian Müller, Joshua McManus, Sebastian Günther, Martin Schmitz, Max Mühlhäuser, and Markus Funk. "Mind the Tap: Assessing Foot-Taps for Interacting with Head-Mounted Displays". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, 477:1–477:13. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300707.

[117] Diederick C. Niehorster, Thiago Santini, Roy S. Hessels, Ignace T. C. Hooge, Enkelejda Kasneci, and Marcus Nyström. "The impact of slippage on the data quality of head-worn eye trackers". In: *Behavior Research Methods* 52.3 (June 2020), pp. 1140–1160. ISSN: 1554-3528. DOI: 10.3758/s13428-019-01307-0.

[118] Nahal Norouzi, Austin Erickson, Kangsoo Kim, Ryan Schubert, Joseph LaViola, Gerd Bruder, and Greg Welch. "Effects of Shared Gaze Parameters on Visual Target Identification Task Performance in Augmented Reality". In: *Symposium on Spatial User Interaction*. SUI '19. New Orleans, LA, USA: Association for Computing Machinery, 2019. ISBN: 9781450369756. DOI: 10.1145/3357251.3357587.

[119] Tomi Nukarinen, Jari Kangas, Oleg Špakov, Poika Isokoski, Deepak Akkil, Jussi Rantala, and Roope Raisamo. "Evaluation of HeadTurn: An Interaction Technique Using the Gaze and Head Turns". In: *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. NordiCHI '16. Gothenburg, Sweden: ACM, 2016, 43:1–43:8. ISBN: 978-1-4503-4763-1. DOI: 10.1145/2971485.2971490.

[120] Brian S. Oommen, Ryan M. Smith, and John S. Stahl. "The influence of future gaze orientation upon eye-head coupling during saccades". In: *Experimental Brain Research* 155.1 (Mar. 2004), pp. 9–18. DOI: 10.1007/s00221-003-1694-z.

[121] Nelusa Pathmanathan, Michael Becher, Nils Rodrigues, Guido Reina, Thomas Ertl, Daniel Weiskopf, and Michael Sedlmair. "Eye vs. Head: Comparing Gaze Methods for Interaction in Augmented Reality". In: *ACM Symposium on Eye Tracking Research and Applications*. ETRA '20 Short Papers. Stuttgart, Germany: Association for Computing Machinery, 2020. ISBN: 9781450371346. DOI: 10.1145/3379156.3391829.

[122] Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. "Towards Foveated Rendering for Gaze-tracked Virtual Reality". In: *ACM Trans. Graph.* 35.6 (Nov. 2016), 179:1–179:12. ISSN: 0730-0301. DOI: 10.1145/2980179.2980246.

[123] Tomislav Pejsa, Daniel Rakita, Bilge Mutlu, and Michael Gleicher. "Authoring Directed Gaze for Full-body Motion Capture". In: *ACM Trans. Graph.* 35.6 (Nov. 2016), 161:1–161:11. ISSN: 0730-0301. DOI: 10.1145/2980179.2982444.

[124] Jeff Pelz, Mary Hayhoe, and Russ Loeber. "The coordination of eye, head, and hand movements in a natural task". In: *Experimental Brain Research* 139.3 (Aug. 2001), pp. 266–277. ISSN: 1432-1106. DOI: 10.1007/s002210100745.

[125] Kevin Pfeil, Eugene M. Taranta II, Arun Kulshreshth, Pamela Wisniewski, and Joseph J. LaViola Jr. "A Comparison of Eye-head Coordination Between Virtual and Physical Realities". In: *Proceedings of the 15th ACM Symposium on Applied Perception*. SAP '18. Vancouver, British Columbia, Canada: ACM, 2018, 18:1–18:7. ISBN: 978-1-4503-5894-1. DOI: 10.1145/3225153.3225157.

[126] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. "Gaze-touch: Combining Gaze with Multi-touch for Interaction on the Same Surface". In: *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. Honolulu, Hawaii, USA: ACM, 2014, pp. 509–518. ISBN: 978-1-4503-3069-5. DOI: 10.1145/2642918.2647397.

[127] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. "Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze". In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST '15. Charlotte, NC, USA: ACM, 2015, pp. 373–383. ISBN: 978-1-4503-3779-3. DOI: 10.1145/2807442.2807460.

[128] Ken Pfeuffer, Jason Alexander, and Hans Gellersen. "Partially-Indirect Bimanual Input with Gaze, Pen, and Touch for Pan, Zoom, and Ink Interaction". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 2845–2856. ISBN: 9781450333627. DOI: 10.1145/2858036.2858201.

[129] Ken Pfeuffer, Matthias J. Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian Alt. "Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300340.

[130] Ken Pfeuffer and Hans Gellersen. "Gaze and Touch Interaction on Tablets". In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. UIST '16. Tokyo, Japan: ACM, 2016, pp. 301–311. ISBN: 978-1-4503-4189-9. DOI: 10.1145/2984511.2984514.

[131] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. "Gaze + Pinch Interaction in Virtual Reality". In: *Proceedings of the 5th Symposium on Spatial User Interaction*. SUI '17. Brighton, United Kingdom: ACM, 2017, pp. 99–108. ISBN: 978-1-4503-5486-8. DOI: 10.1145/3131277.3132180.

[132] Ken Pfeuffer, Lukas Mecke, Sarah Delgado Rodriguez, Mariam Hassib, Hannah Maier, and Florian Alt. "Empirical Evaluation of Gaze-Enhanced Menus in Virtual Reality". In: *26th ACM Symposium on Virtual Reality Software and Technology*. VRST '20. Virtual Event, Canada: Association for Computing Machinery, 2020. ISBN: 9781450376198. DOI: 10.1145/3385956.3418962.

[133] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. "The Effects of Sharing Awareness Cues in Collaborative Mixed Reality". In: *Frontiers in Robotics and AI* 6 (2019). ISSN: 2296-9144. DOI: 10.3389/frobt.2019.00005.

[134] Thammathip Piumsomboon, Gun Lee, Robert W. Lindeman, and Mark Billinghurst. "Exploring Natural Eye-gaze-based Interaction for Immersive Virtual Reality". In: *2017 IEEE Symposium on 3D User Interfaces*. 3DUI.

IEEE, 2017, pp. 36–39. ISBN: 9781509067169. DOI: 10.1109/3DUI.2017.7893315.

[135] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. "The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-Worn Extended Reality". In: *ACM Comput. Surv.* 55.3 (Mar. 2022). ISSN: 0360-0300. DOI: 10.1145/3491207.

[136] Marco Porta, Alice Ravarelli, and Giovanni Spagnoli. "ceCursor, a Contextual Eye Cursor for General Pointing in Windows Environments". In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ETRA '10. Austin, Texas: ACM, 2010, pp. 331–337. ISBN: 978-1-60558-994-7. DOI: 10.1145/1743666.1743741.

[137] Majid Pourmemar and Charalambos Poullis. "Visualizing and Interacting with Hierarchical Menus in Immersive Augmented Reality". In: *The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry*. VRCAI '19. Brisbane, QLD, Australia: Association for Computing Machinery, 2019. ISBN: 9781450370028. DOI: 10.1145/3359997.3365693.

[138] Yuan Yuan Qian and Robert J. Teather. "The Eyes Don't Have It: An Empirical Comparison of Head-based and Eye-based Selection in Virtual Reality". In: *Proceedings of the 5th Symposium on Spatial User Interaction*. SUI '17. Brighton, United Kingdom: ACM, 2017, pp. 91–98. ISBN: 978-1-4503-5486-8. DOI: 10.1145/3131277.3132182.

[139] Yuan Yuan Qian and Robert J. Teather. "Look to Go: An Empirical Evaluation of Eye-Based Travel in Virtual Reality". In: *Proceedings of the Symposium on Spatial User Interaction*. SUI '18. Berlin, Germany: Association for Computing Machinery, 2018, pp. 130–140. ISBN: 9781450357081. DOI: 10.1145/3267782.3267798.

[140] Yashas Rai, Jesús Gutiérrez, and Patrick Le Callet. "A Dataset of Head and Eye Movements for 360 Degree Images". In: *Proceedings of the 8th ACM on Multimedia Systems Conference*. MMSys'17. Taipei, Taiwan: ACM, 2017, pp. 205–210. ISBN: 978-1-4503-5002-0. DOI: 10.1145/3083187.3083218.

[141] Vijay Rajanna and John Paulin Hansen. "Gaze Typing in Virtual Reality: Impact of Keyboard Design, Selection Method, and Motion". In: *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ETRA '18. Warsaw, Poland: ACM, 2018, 15:1–15:10. ISBN: 978-1-4503-5706-7. DOI: 10.1145/3204493.3204541.

[142] Argenis Ramirez Ramirez Gomez, Christopher Clarke, Ludwig Sidenmark, and Hans 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*. ETRA '21 Full Papers. Virtual Event, Germany: Association for Computing Machinery, 2021. ISBN: 9781450383448. DOI: 10.1145/3448017.3457381. URL: https://doi.org/10.1145/3448017.3457381.

[143] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. "A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception". In: *Computer Graphics Forum* 34.6 (Sept. 2015), pp. 299–326. ISSN: 0167-7055. DOI: 10.1111/cgf.12603.

[144] Sohrab Saeb, Cornelius Weber, and Jochen Triesch. "Learning the optimal control of coordinated eye and head movements". In: *PLoS Computational Biology* 7.11 (Nov. 2011). Ed. by Jörn Diedrichsen, e1002253. ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1002253.

[145] Krystian Samp and Stefan Decker. "Supporting Menu Design with Radial Layouts". In: *Proceedings of the International Conference on Advanced Visual Interfaces*. AVI '10. Roma, Italy: Association for Computing Machinery, 2010, pp. 155–162. ISBN: 9781450300766. DOI: 10.1145/1842993.1843021.

[146] A. Santos, T. Zarraonandia, P. Díaz, and I. Aedo. "A Comparative Study of Menus in Virtual Reality Environments". In: *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. ISS '17. Brighton, United Kingdom: Association for Computing Machinery, 2017, pp. 294–299. ISBN: 9781450346917. DOI: 10.1145/3132272.3132277.

[147] Sayan Sarcar, Prateek Panwar, and Tuhin Chakraborty. "EyeK: An Efficient Dwell-Free Eye Gaze-Based Text Entry System". In: *Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction*. APCHI '13. Bangalore, India: Association for Computing Machinery, 2013, pp. 215–220. ISBN: 9781450322539. DOI: 10.1145/2525194.2525288.

[148] Jeffrey S. Shell, Roel Vertegaal, Daniel Cheng, Alexander W. Skaburskis, Changuk Sohn, A. James Stewart, Omar Aoudeh, and Connor Dickie. "ECSGlasses and EyePliances: Using Attention to Open Sociable Windows of Interaction". In: *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*. ETRA '04. San Antonio, Texas: ACM, 2004, pp. 93–100. ISBN: 1-58113-825-3. DOI: 10.1145/968363.968384.

[149] Gwanseob Shin and Sudeep Hegde. “User-Preferred Position of Computer Displays: Effects of Display Size”. In: *Human Factors* 52.5 (2010), pp. 574–585. DOI: 10.1177/0018720810380405.

[150] Ludwig Sidenmark, Christopher Clarke, Xuesong Zhang, Jenny Phu, and Hans Gellersen. “Outline Pursuits: Gaze-Assisted Selection of Occluded Objects in Virtual Reality”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376438.

[151] Ludwig Sidenmark and Hans Gellersen. “Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality”. In: *ACM Trans. Comput.-Hum. Interact.* 27.1 (Dec. 2019). ISSN: 1073-0516. DOI: 10.1145/3361218.

[152] Ludwig Sidenmark and Hans Gellersen. “Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection”. In: *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. UIST ’19. New Orleans, LA, USA: Association for Computing Machinery, 2019, pp. 1161–1174. ISBN: 9781450368162. DOI: 10.1145/3332165.3347921.

[153] Ludwig Sidenmark, Nicolas Kiefer, and Hans Gellersen. “Subtitles in interactive virtual reality: Using gaze to address depth conflicts”. In: *Workshop on Emerging Novel Input Devices and Interaction Techniques*. 2019.

[154] Ludwig Sidenmark and Anders Lundström. “Gaze Behaviour on Interacted Objects During Hand Interaction in Virtual Reality for Eye Tracking Calibration”. In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ETRA ’19. Denver, Colorado: ACM, 2019, 6:1–6:9. ISBN: 978-1-4503-6709-7. DOI: 10.1145/3314111.3319815.

[155] Ludwig Sidenmark, Diako Mardanbegi, Argenis Ramirez Gomez, Christopher Clarke, and Hans Gellersen. “BimodalGaze: Seamlessly Refined Pointing with Gaze and Filtered Gestural Head Movement”. In: *ACM Symposium on Eye Tracking Research and Applications*. ETRA ’20 Full Papers. Stuttgart, Germany: Association for Computing Machinery, 2020. ISBN: 9781450371339. DOI: 10.1145/3379155.3391312.

[156] Ludwig Sidenmark, Mark Parent, Chi-Hao Wu, Joannes Chan, Michael Glueck, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. “Weighted Pointer: Error-aware Gaze-based Interaction through Fallback Modalities”. In: *IEEE Transactions on Visualization and Computer Graphics* 28.11 (2022), pp. 3585–3595. DOI: 10.1109/TVCG.2022.3203096.

[157] Ludwig Sidenmark, Dominic Potts, Bill Bapisch, and Hans Gellersen. "Radi-Eye: Hands-Free Radial Interfaces for 3D Interaction using Gaze-Activated Head-Crossing". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021, pp. 1–11. DOI: 10.1145/3411764.3445697.

[158] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. "Saliency in VR: how do people explore virtual environments?" In: *IEEE Transactions on Visualization and Computer Graphics* 24.4 (Apr. 2018), pp. 1633–1642. ISSN: 1077-2626. DOI: 10.1109/TVCG.2018.2793599.

[159] Oleg Špakov, Poika Isokoski, Jari Kangas, Deepak Akkil, and Päivi Majaranta. "PursuitAdjuster: An Exploration into the Design Space of Smooth Pursuit –Based Widgets". In: *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. ETRA '16. Charleston, South Carolina: Association for Computing Machinery, 2016, pp. 287–290. ISBN: 9781450341257. DOI: 10.1145/2857491.2857526.

[160] Oleg Špakov, Poika Isokoski, and Päivi Majaranta. "Look and Lean: Accurate Head-assisted Eye Pointing". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '14. Safety Harbor, Florida: ACM, 2014, pp. 35–42. ISBN: 978-1-4503-2751-0. DOI: 10.1145/2578153.2578157.

[161] Oleg Špakov and Päivi Majaranta. "Enhanced Gaze Interaction Using Simple Head Gestures". In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. UbiComp '12. Pittsburgh, Pennsylvania: ACM, 2012, pp. 705–710. ISBN: 978-1-4503-1224-0. DOI: 10.1145/2370216.2370369.

[162] John S. Stahl. "Amplitude of human head movements associated with horizontal saccades". In: *Experimental Brain Research* 126.1 (Apr. 1999), pp. 41–54. ISSN: 00144819. DOI: 10.1007/s002210050715.

[163] John S. Stahl. "Eye-head coordination and the variation of eye-movement accuracy with orbital eccentricity." In: *Experimental Brain Research* 136.2 (Jan. 2001), pp. 200–210. ISSN: 0014-4819. DOI: 10.1007/s002210000593.

[164] India Starker and Richard A. Bolt. "A Gaze-Responsive Self-Disclosing Display". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '90. Seattle, Washington, USA: Association for Computing Machinery, 1990, pp. 3–10. ISBN: 0201509326. DOI: 10.1145/97243.97245.

[165] Sophie Stellmach and Raimund Dachselt. "Investigating Gaze-Supported Multimodal Pan and Zoom". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. Santa Barbara, California: Association for Computing Machinery, 2012, pp. 357–360. ISBN: 9781450312219. DOI: 10.1145/2168556.2168636.

[166] Sophie Stellmach and Raimund Dachselt. "Look & Touch: Gaze-supported Target Acquisition". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: ACM, 2012, pp. 2981–2990. ISBN: 978-1-4503-1015-4. DOI: 10.1145/2207676.2208709.

[167] Sophie Stellmach and Raimund Dachselt. "Still Looking: Investigating Seamless Gaze-Supported Selection, Positioning, and Manipulation of Distant Targets". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. Paris, France: Association for Computing Machinery, 2013, pp. 285–294. ISBN: 9781450318990. DOI: 10.1145/2470654.2470695.

[168] Sophie Stellmach, Sebastian Stober, Andreas Nürnberger, and Raimund Dachselt. "Designing Gaze-Supported Multimodal Interactions for the Exploration of Large Image Collections". In: *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications*. NGCA '11. Karlskrona, Sweden: Association for Computing Machinery, 2011. ISBN: 9781450306805. DOI: 10.1145/1983302.1983303.

[169] Ahmed N. Sulaiman and Patrick Olivier. "Attribute Gates". In: *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*. UIST '08. Monterey, CA, USA: Association for Computing Machinery, 2008, pp. 57–66. ISBN: 9781595939753. DOI: 10.1145/1449715.1449726.

[170] Vildan Tanriverdi and Robert J. K. Jacob. "Interacting with Eye Movements in Virtual Environments". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '00. The Hague, The Netherlands: ACM, 2000, pp. 265–272. ISBN: 1-58113-216-6. DOI: 10.1145/332040.332443.

[171] Zachary C. Thumser, Brian S. Oommen, Igor S. Kofman, and John S. Stahl. "Idiosyncratic variations in eye–head coupling observed in the laboratory also manifest during spontaneous behavior in a natural setting". In: *Experimental Brain Research* 191.4 (Dec. 2008), pp. 419–434. ISSN: 1432-1106. DOI: 10.1007/s00221-008-1534-2.

[172] Zachary C. Thumser and John S. Stahl. "Eye-head coupling tendencies in stationary and moving subjects". In: *Experimental Brain Research* 195.3 (May 2009), pp. 393–401. ISSN: 00144819. DOI: 10.1007/s00221-009-1803-8.

[173] Subarna Tripathi and Brian Guenter. "A Statistical Approach to Continuous Self-Calibrating Eye Gaze Tracking for Head-Mounted Virtual Reality Systems". In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2017, pp. 862–870. DOI: 10.1109/WACV.2017.101.

[174] Huawei Tu, Susu Huang, Jiabin Yuan, Xiangshi Ren, and Feng Tian. "Crossing-Based Selection with Virtual Reality Head-Mounted Displays". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019. ISBN: 9781450359702. DOI: 10.1145/3290605.3300848.

[175] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. "Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 4179–4188. ISBN: 9781450331456. DOI: 10.1145/2702123.2702355.

[176] Jayson Turner, Jason Alexander, Andreas Bulling, Dominik Schmidt, and Hans Gellersen. "Eye Pull, Eye Push: Moving Objects between Large Screens and Personal Devices with Gaze and Touch". In: *Human-Computer Interaction – INTERACT 2013*. Ed. by Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 170–186. ISBN: 978-3-642-40480-1. DOI: 10.1007/978-3-642-40480-1_11.

[177] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. "Eye Drop: An Interaction Concept for Gaze-Supported Point-to-Point Content Transfer". In: *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*. MUM '13. Luleå, Sweden: Association for Computing Machinery, 2013. ISBN: 9781450326483. DOI: 10.1145/2541831.2541868.

[178] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. "Cross-device Gaze-supported Point-to-point Content Transfer". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '14. Safety Harbor, Florida: ACM, 2014, pp. 19–26. ISBN: 978-1-4503-2751-0. DOI: 10.1145/2578153.2578155.

[179] Jayson Turner, Andreas Bulling, and Hans Gellersen. "Combining Gaze with Manual Interaction to Extend Physical Reach". In: *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction*. PETMEI '11. Beijing, China: Association for Computing Machinery, 2011, pp. 33–36. ISBN: 9781450309301. DOI: 10.1145/2029956.2029966.

[180] Jayson Turner, Andreas Bulling, and Hans Gellersen. "Extending the Visual Field of a Head-mounted Eye Tracker for Pervasive Eye-based Interaction". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. Santa Barbara, California: ACM, 2012, pp. 269–272. ISBN: 978-1-4503-1221-9. DOI: 10.1145/2168556.2168613.

[181] D. Tweed, B. Glenn, and T. Vilis. "Eye-head Coordination during Large Gaze Shifts". In: *Journal of Neurophysiology* 73.2 (Feb. 1995), pp. 766–779. ISSN: 0022-3077. DOI: 10.1152/jn.1995.73.2.766.

[182] Mario H. Urbina, Maike Lorenz, and Anke Huckauf. "Pies with EYEs: The Limits of Hierarchical Pie Menus in Gaze Control". In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ETRA '10. Austin, Texas: ACM, 2010, pp. 93–96. ISBN: 978-1-60558-994-7. DOI: 10.1145/1743666.1743689.

[183] Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. "Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem"". In: *Human-Computer Interaction INTERACT '97: IFIP TC13 International Conference on Human-Computer Interaction, 14th–18th July 1997, Sydney, Australia*. Ed. by Steve Howard, Judy Hammond, and Gitte Lindgaard. Boston, MA: Springer US, 1997, pp. 509–516. ISBN: 978-0-387-35175-9. DOI: 10.1007/978-0-387-35175-9_77.

[184] Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. "AmbiGaze: Direct Control of Ambient Devices by Gaze". In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. DIS '16. Brisbane, QLD, Australia: ACM, 2016, pp. 812–817. ISBN: 978-1-4503-4031-1. DOI: 10.1145/2901790.2901867.

[185] Roel Vertegaal. "Attentive user interfaces". In: *Communications of the ACM* 46.3 (2003), pp. 30–33.

[186] Roel Vertegaal. "A Fitts Law Comparison of Eye Tracking and Manual Input in the Selection of Visual Targets". In: *Proceedings of the 10th International Conference on Multimodal Interfaces*. ICMI '08. Chania, Crete, Greece: Association for Computing Machinery, 2008, pp. 241–248. ISBN: 9781605581989. DOI: 10.1145/1452392.1452443.

[187] Roel Vertegaal, Connor Dickie, Changuk Sohn, and Myron Flickner. "Designing Attentive Cell Phone Using Wearable Eyecontact Sensors". In: *CHI '02 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '02. Minneapolis, Minnesota, USA: Association for Computing Machinery, 2002, pp. 646–647. ISBN: 1581134541. DOI: 10.1145/506443.506526.

[188] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. "Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets". In: *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '13. Zurich, Switzerland: ACM, 2013, pp. 439–448. ISBN: 978-1-4503-1770-2. DOI: 10.1145/2493432.2493477.

[189] Mélodie Vidal, David H. Nguyen, and Kent Lyons. "Looking at or through? Using Eye Tracking to Infer Attention Location for Wearable Transparent Displays". In: *Proceedings of the 2014 ACM International Symposium on Wearable Computers*. ISWC '14. Seattle, Washington: Association for Computing Machinery, 2014, pp. 87–90. ISBN: 9781450329699. DOI: 10.1145/2634317.2634344.

[190] Colin Ware and Harutune H. Mikaelian. "An Evaluation of an Eye Tracker As a Device for Computer Input2". In: *SIGCHI Bulletin* 17.SI (May 1986), pp. 183–188. ISSN: 0736-6906. DOI: 10.1145/30851.275627.

[191] Julie R. Williamson, Mark McGill, and Khari Outram. "PlaneVR: Social Acceptability of Virtual Reality for Aeroplane Passengers". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019. ISBN: 9781450359702. DOI: 10.1145/3290605.3300310.

[192] Jacob O. Wobbrock and Krzysztof Z. Gajos. "Goal Crossing with Mice and Trackballs for People with Motor Impairments: Performance, Submovements, and Design Directions". In: *ACM Trans. Access. Comput.* 1.1 (May 2008). ISSN: 1936-7228. DOI: 10.1145/1361203.1361207.

[193] B. R. Wooten and George Wald. "Color-Vision Mechanisms in the Peripheral Retinas of Normal and Dichromatic Observers ". In: *Journal of General Physiology* 61.2 (Feb. 1973), pp. 125–145. ISSN: 0022-1295. DOI: 10.1085/jgp.61.2.125.

[194] H.P. Wyss, R. Blach, and M. Bues. "iSith - Intersection-based Spatial Interaction for Two Hands". In: *3D User Interfaces (3DUI'06)*. 2006, pp. 59–61. DOI: 10.1109/VR.2006.93.

[195] Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Difeng Yu, and Diego Monteiro. "DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019, 444:1–444:14. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300674.

[196] Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Tianyu Zhang, Difeng Yu, and Diego Monteiro. "RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions". In: *IEEE Transactions on Visualization and Computer Graphics* 25.5 (May 2019), pp. 1991–2001. ISSN: 1077-2626. DOI: 10.1109/TVCG.2019.2898736.

[197] Yan Yan, Ke Chen, Yu Xie, Yiming Song, and Yonghong Liu. "The Effects of Weight on Comfort of Virtual Reality Devices". In: *Advances in Ergonomics in Design*. Ed. by Francisco Rebelo and Marcelo M. Soares. Cham: Springer International Publishing, 2019, pp. 239–248. ISBN: 978-3-319-94706-8. DOI: 10.1007/978-3-319-94706-8_27.

[198] Yukang Yan, Yingtian Shi, Chun Yu, and Yuanchun Shi. "HeadCross: Exploring Head-Based Crossing Selection on Head-Mounted Displays". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4.1 (Mar. 2020). DOI: 10.1145/3380983.

[199] Yukang Yan, Chun Yu, Xin Yi, and Yuanchun Shi. "HeadGesture: Hands-Free Input Approach Leveraging Head Movements for HMD Devices". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2.4 (Dec. 2018). DOI: 10.1145/3287076. URL: https://doi.org/10.1145/3287076.

[200] Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. "Gaze-Supported 3D Object Manipulation in Virtual Reality". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. Yokohama, Japan: Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445343.

[201] Robert C. Zeleznik, Andrew S. Forsberg, and Jürgen P. Schulze. *Look-That-There: Exploiting Gaze in Virtual Reality Interactions*. Tech. rep. Brown University, 2005.

[202] Shumin Zhai, Carlos Morimoto, and Steven Ihde. "Manual and Gaze Input Cascaded (MAGIC) Pointing". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '99. Pittsburgh, Pennsylvania, USA: ACM, 1999, pp. 246–253. ISBN: 0-201-48559-1. DOI: 10.1145/302979.303053.

[203] Xuan Zhang and I. Scott MacKenzie. "Evaluating Eye Tracking with ISO 9241 - Part 9". In: *Proceedings of the 12th International Conference on Human-computer Interaction: Intelligent Multimodal Interaction Environments*. HCI'07. Beijing, China: Springer-Verlag, 2007, pp. 779–788. ISBN: 978-3-540-73108-5. DOI: 10.1007/978-3-540-73110-8_85.

[204] Zhiwei Zhu and Qiang Ji. "Novel Eye Gaze Tracking Techniques Under Natural Head Movement". In: *IEEE Transactions on Biomedical Engineering* 54.12 (Dec. 2007), pp. 2246–2260. ISSN: 0018-9294. DOI: 10.1109/TBME.2007.895750.