Niklas Hypki
How We See, Perceive and Act:
Analysing Eye Movements in Virtual Reality
2025

Psychologie

How We See, Perceive and Act: Analysing Eye Movements in Virtual Reality

Inaugural-Dissertation zur Erlangung des Doktorgrades
im Fachbereich Psychologie und Sportwissenschaft
der Universität Münster

vorgelegt von
Niklas Hypki geb. Stein
aus Herdecke
2025

Dekanin:	Prof. Dr. Ricarda I. Schubotz
Erster Gutachter:	Prof. Dr. Markus Lappe
Zweiter Gutachter:	Prof. Dr. Heiko Wagner

Tag der mündlichen Prüfung:
Tag der Promotion:

Abstract

Following rapid technological developments in recent years, head-mounted displays (HMDs) can now track movement trajectories of the head, feet, hands, joints and even our eyes simultaneously. The possibility to rearrange and manipulate virtual stimuli based on our behaviour makes virtual reality (VR) a potentially interesting method for experiments in psychophysics. This thesis presents three empirical, peer-reviewed studies exploring this approach.

To obtain a fair assessment of the latency of eye tracking sensors in HMDs, we developed a method to measure it using simultaneous electrooculography (EOG) recording. Delays (time from an eye movement to the availability of the corresponding data) ranged from 15 ms to 52 ms, end-to-end latency (time from an eye movement to a corresponding change on the display) ranged from 45 ms to 81 ms.

Based on HMD gaze data, we then predicted future waypoints. To achieve this, 18 participants performed tasks such as walking along curved paths, avoiding or approaching objects and searching while we recorded their position, orientation and eye tracking data. Segments of 2.5 s of data were used to train a long short-term memory (LSTM) model that predicts the position the user reaches in 2.5 s. The prediction was fairly accurate with an average error of 66 cm. The extent of model improvements based on eye movement data varied depending on the task and environment. Overall, eye movements enabled more accurate prediction of locomotion behaviour, especially in situations with varying walking speeds.

Finally, we analyse the eye and head movements of 68 participants as they searched for salient and less salient targets in spatially grouped sets inside and beyond the initial field of view (FOV). Salient targets within the initial FOV facilitated search. Salient stimuli that appeared to ‘pop out’ from the periphery at the beginning of a search no longer exhibited this property when they entered the FOV due to a head movement during an ongoing search. This suggests that the influence of stimulus saliency differs between searching on a static display and in scenarios in which we can explore our surroundings using head turns.

The general discussion addresses current limitations of VR and presents approaches for reducing eye tracking sensor latency in future devices. Subsequently, ideas for improving locomotion prediction are presented. Finally, gaze behaviour is discussed in relation to the typical temporal course of the influence of top-down and bottom-up guidance.

Altogether, this thesis shows that VR can be used as a method in psychophysics. Our results demonstrate that experiments in virtual environments (VEs) have the potential to contribute to our understanding of perception as well as to the advancement of VR.

Keywords

Visual Perception, Virtual Reality, Eye Movements, Eye Tracking, Gaze, Head Movements, Visual Search, Locomotion, Prediction, Salience

Table of Contents

List of Figures

List of Tables

Abbreviations

CRT	Cathode-ray tube
EEG	Electroencephalography
EOG	Electrooculography
FOV	Field of view
GCD	Gaze-contingent display
GRU	Gated recurrent unit
HMD	Head-mounted display
IMU	Inertial measurement unit
LCD	Liquid crystal display
LED	Light-emitting diode
LGN	Lateral geniculate nucleus
LO	Lateral occipital complex
LSTM	Long short-term memory
MDE	Mean displacement error
MST	Medial superior temporal
OLED	Organic light-emitting diode
PWM	Pulse-width modulation
RDW	Redirected walking
RGB	Red, green and blue
SC	Superior colliculus
SDK	Software development kit
V1	Primary visual cortex
V2	Secondary visual cortex
V3	Third visual cortex
V3a	Third visual complex accessory
V4	Visual area V4
V5	Middle temporal visual area
VAC	Vergence accommodation conflict
VE	Virtual environment
VOG	Video-oculography
VOR	Vestibulo-ocular reflex
VR	Virtual reality

Towards Gaze-contingent Head-mounted Displays

In 1968, Ivan Sutherland and his students developed the first head-mounted display (HMD). This new see-through device made it possible to overlay rudimentary virtual objects such as triangles or cubes on the surrounding environment of the user (Sutherland, 1968). The digital objects could change in perspective according to the user's head movements, which were tracked using a heavy mechanical system. It was the first prototype of an augmented reality display. Half a century later, commercially available virtual reality (VR) headsets allow users to project very detailed virtual objects and HMDs have become everyday devices for exploring virtual environments (VEs). The key factors that have made this rapid development possible have been the astonishing technological advances in displays, sensors, motion detection algorithms, computer hardware and batteries, as well as an increasingly better understanding of our visual perception. By continuously adapting technological innovations to newly discovered perceptual principles, VEs can now be explored intuitively, using natural gestures and movements.

Today, we may have even reached a point where comparing the virtual with the real world can not only help us to create more realistic virtual experiences, but also expand our knowledge of how we perceive and behave in our environment in general.

The first part of this thesis deals with the evaluation of the basic requirements that must be met in order for HMDs to be used for the presentation of stimuli in scientific experiments in psychophysics. The introduction first presents the basic functions and technologies of current HMDs. It then provides a general overview of the anatomy of the eye and the different types of eye movements we use to perceive the world around us, which also shape our perception in VR. To accurately measure the effect of stimuli in psychophysical experiments, it is also important to know exactly which eye movements are performed at what point during the presentation. Therefore, a brief introduction to the various eye tracking methods available follows. Particular emphasis was put on the latency requirements for the presentation of gaze-dependent stimuli, which are used in many paradigms of perception research. This is followed by study I, in which we measured the latency of eye tracking in HMDs to assess whether gaze-contingent paradigms can be presented in VR. Next, an overview of how to determine three-dimensional fixation points and classify eye movements in VEs is provided.

In the following chapters, eye movements during various activities in VR are analysed. This initially serves to expand our knowledge of human perception, which could subsequently be used to improve VR applications. First, a scientific overview of typical eye movements during actions is provided. Here, particular emphasis was put on eye movements that adapt in order to solve specific tasks, as well as eye movements during locomotion. Next is study II, in which we recorded gaze, movement and orientation data in three different typical locomotion tasks. We then used this data to predict future waypoints. Here, we were particularly interested in whether eye tracking data is capable of improving prediction accuracy.

The following chapter deals with the extent to which our eye and head movements are guided by visual stimuli. It provides an introduction to the basic principles of visual search, the visual pathway and the visual cortex, and then explains how the shift of attention is possible at the neural level in the brain. In study III, we recorded and analysed eye and head movements during visual search tasks in VR.

The final chapter concludes the thesis with a general discussion of our findings. It presents different approaches for extending our results and addresses the question of the extent to which our eye movements are typically controlled by bottom-up and top-down processes. Then, we discuss how the current limitations of VR as a method in psychophysics can be addressed in research. Finally, the main conclusions of the thesis are drawn.

Head-mounted Displays

To present virtual objects, an HMD obviously needs some sort of display. To achieve stereo vision, two separate two-dimensional projections of a simulated object or scene are usually displayed. All objects in both images are slightly offset based on the position of the eyes (Hibbard et al., 2020). The brain can then extract three-dimensional information from this offset, just as it does from the retinal projections of the real world (Scarfe & Glennerster, 2019; Wheatstone, 1838). The first HMD prototypes were equipped with two cathode-ray tubes (CRTs) as screens that presented two different images, one for each eye.

Today, two types of screens are currently available in HMDs: liquid crystal displays (LCDs) and organic light-emitting diode (OLED) displays. Compared to their CRT predecessors, both technologies enable lighter and thinner VR headsets with screens that require less power and therefore also generate less heat. In both types of screen, the image is generated using red, green and blue (RGB) sub-pixels. In LCDs, these consist of a backlight source that emits light, which is then polarised and passed through colour filters and a liquid crystal matrix (Chen et al., 2018). Depending on the strength of an electric current applied to the liquid crystal layer, the crystals align themselves and become less or more translucent. As a result, sub-pixels become darker or lighter. Due to viscosity, it can take up to 10 ms for an LCD to change from black to white. In early LCDs, the backlight light source was a cold cathode fluorescent lamp. In today's displays, a light-emitting diode (LED) matrix is typically used, which consumes less power and enables splitting up the display surface in multiple dimming zones. As a result, black areas of the image appear darker and the overall contrast of the image is increased.

OLED displays have self-illuminating RGB subpixels, which consist of an LED matrix with self-emitting diodes and no backlight. This makes it possible to design even thinner and lighter displays, in which individual pixels are simply switched off to display black. It also lowers the overall power consumption and makes black to white response times of well under 1 ms possible. Yet current OLED displays have problems with brightness and stability (Liu et al., 2020): With continuous use, self-emitting LEDs become less efficient and therefore reduce their light output. This can lead to burnt-in images, especially if monochrome red, green, or blue stimuli are displayed at the same position using the same pixels over a long period of time.

To delay LED-wear, the maximum brightness in OLED displays is often reduced below the maximum brightness of the diodes by using pulse-width modulation (PWM). During active PWM, the individual LEDs constantly switch between on- and off-states (Baker, 2022). The longer each on-state (duty cycle) lasts, the higher the image brightness. For instance, a 90% duty cycle creates a bright image while a 10% duty cycle creates a dark image.

If the switching occurs at a frequency above the critical flicker fusion frequency of approximately 50 to 90 Hz, we generally do not perceive the image as flickering (Mankowska et al., 2021). Up to 500 Hz, however, we can still perceive flicker (with slight variations for different colours) when we are asked to detect it on modern displays (J. Davis et al., 2015). When PWM was first introduced in smartphone displays, frequencies of around 240 Hz were used, which some users found very unpleasant. Some can even get headaches from it, which might be related to migraineurs having slightly lower critical flicker frequencies (Kowacs et al., 2004, 2005).

Interestingly, the off-states, which are also sometimes described as artificially added black frames, have another useful property for displaying moving images: when we follow a moving rigid object with our eyes, we follow it with a continuous eye movement. By briefly inserting a black frame between two images, each stable image is visible for a shorter amount of time. Similar to the effect of a shorter exposure time in a camera, the perceived movement on the screen with black frames inserted appears sharper. This technique was also used in film projectors, which often used shutters with three blades at 24 frames per second to increase the flicker frequency to 72 Hz and achieve the same sharpening effect (Hammock, 2015).

Unlike film projectors, LCDs and OLED displays do not automatically produce off-states between individual frames. Instead, the rendered content follows a ‘sample-and-hold’ characteristic (Marsh, 2017). This means that a moving virtual object jumps forward frame by frame on a screen. Between jumps, the image remains static. This sample-and-hold image representation does not reflect the continuous movement in the real world. Each frame of the rendered content is only valid for a brief moment. In reality, at any subsequent point in time, the objects in the scene have moved to different locations. Sample-and-hold rendering presents another problem: When our eyes move across a static frame with a continuous eye movement, the content is perceived as blurry. Especially in VR, shifts of the whole FOV caused by head movements would then result in significant motion blur. Therefore, inserting artificial black images instead of presenting one frame until the next update has become a frequently used technique in HMDs. In this context, the length of each individual on-state of a frame is also referred to as the persistence of the display. In LCDs, synchronised backlight flashing is used to insert black frames. In OLED displays, all LEDs can be turned on and off directly to achieve the same effect.

Lenses

To distribute the image to the eyes of the user, a lens collects the light from the display and maps it onto the FOV (Bang et al., 2021). For this purpose, three different types of lens designs are currently used in HMDs: Fresnel lenses, aspheric lenses and pancake lenses (see Figure A). All three VR lenses have positive optical power and increase the angular size. Therefore, they make the close display look large and spread the image across the FOV.

Fresnel lenses consist of many prism-like, ring-shaped sections. With each section and from the inside to the outside, the angle at which the emerging light is diffracted increases (A. Davis & Kühnlenz, 2007). Fresnel lenses can be made of plastic and are therefore comparatively light. However, with the designs currently in use, the edges of each lens section are clearly visible. In addition, a sharp image can only be seen when the user's eyes are directly opposite the flat surface in the centre of the lens. This makes the area through which an image can be perceived sharply (sweet spot) comparably small. Moreover, light, especially from outer parts of the lens, can stray into other, darker areas of the image (Haitao et al., 2022). Sometimes the resulting visual effect is referred to as a ghost image. Additionally, Fresnel lenses introduce distortion to the transmitted image, accompanied by chromatic aberrations, often resulting in blurring and colourful aberrations on high-contrast edges. To present an undistorted image, the displayed content is preprocessed by software (Anthes et al., 2016; Nikonorov et al., 2015).

Aspherical lenses have a continuously ground surface and can be made of plastic or glass, giving the transmitted image no visible internal edges. Moreover, aspheric lenses have a larger sweet spot. On the one hand, this means the image does not become blurry when the HMD slips slightly on the user's head. On the other hand, distortions introduced by the lens still vary depending on the exact pupil position. Thus, such slippage and even small positional movements caused by an eye movement can lead to local distortions. Sometimes these effects, also known as the pupil swim effect, cause people to feel uncomfortable. To compensate them with software, very accurate and low latency eye tracking is needed (Rui et al., 2023).

In pancake lenses, the incoming light is reflected back and forth once within the lens using a half-mirror effect. As the optical path is therefore longer than the physical distance, the lens can be positioned closer to the display, which enables overall smaller HMDs (Bang et al., 2021). A side effect of the half-mirror, however, is that a part of the incoming light is reflected towards the display, resulting in a darker outgoing image. Although this can be compensated for by increasing the brightness of the screen, this in turn has other side effects, such as greater wear and tear on light sources of the displays as well as higher power consumption and greater heat generation. Like aspherical lenses, pancake lenses have a large sweet spot and can also cause pupil swim effects (Jia et al., 2023). Stray light caused by multiple surface reflections and imperfect polarisation control within the lens can also create ghost images (Luo et al., 2024).

Movement Tracking in Virtual Reality

To enable a VE that can not only be displayed but also be explored using movements such as head turning and walking, a VR headset needs a range of sensors that measure the user's behaviour and transfers it into the virtual space. Modern HMDs are equipped

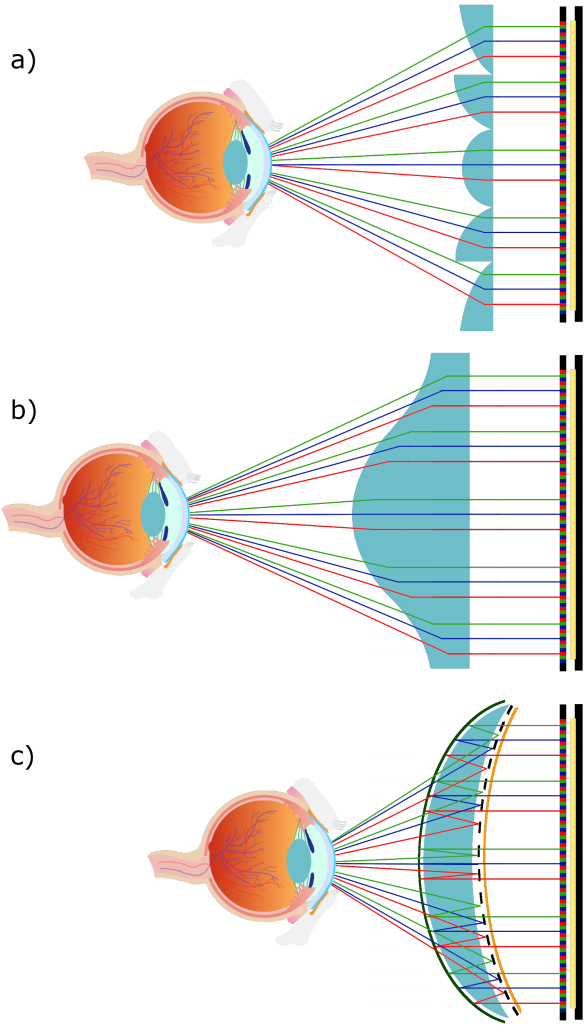


Figure A: HMD lens designs. All three lenses spread the image across the FOV. A Fresnel lens (a) consists of several rings with different optical powers. The transitions from one refractive step to the next result in a comparably small sweet spot. Aspherical lenses (b) are continuously ground and therefore thicker. In pancake lenses (c), the inner quarter-wave plate (orange) converts the linear polarisation of the light emerging from the display into circular polarisation so that it can first pass through the polarizing beam splitter (black dotted line). The outer quarter-wave plate (dark green) reverses the polarisation upon entry and reverses it again upon exit (after reflection), ensuring that the light is correctly polarised to be reflected back and then transmitted to the eye. Overall, this makes the distance travelled by the light longer. This shortens the physical distance needed between eye and lens and enables more compact HMDs.

with an inertial measurement unit (IMU) that usually includes an accelerometer, a gyroscope and a magnetometer to accurately track head rotation (Anthes et al., 2016).

The position can be tracked using external devices or cameras within the HMD, which is also known as inside-out tracking (Anthes et al., 2016; Niehorster et al., 2017). The latter is typically based on visual information from multiple front cameras that are on the outside of the headset (see Figure B) to gather data on the current surroundings and can deliver even more accurate position tracking than common external tracking methods based on infrared lighthouses (Holzwarth et al., 2021). Both types of systems can also be used to track devices, such as controllers or body trackers that can be attached to joints such as knees or feet (Anthes et al., 2016). Using a multi-stage process to estimate hand pose and finger angles, inside-out tracking systems can also be used to track the hands without additional devices in about 45 ms (Abdlkarim et al., 2024). This allows users to interact with virtual objects, for example, by pointing and grasping to select and single out targets. Moreover, the camera-based tracking can be complemented with a depth sensor, to achieve a more detailed map of the surrounding objects.

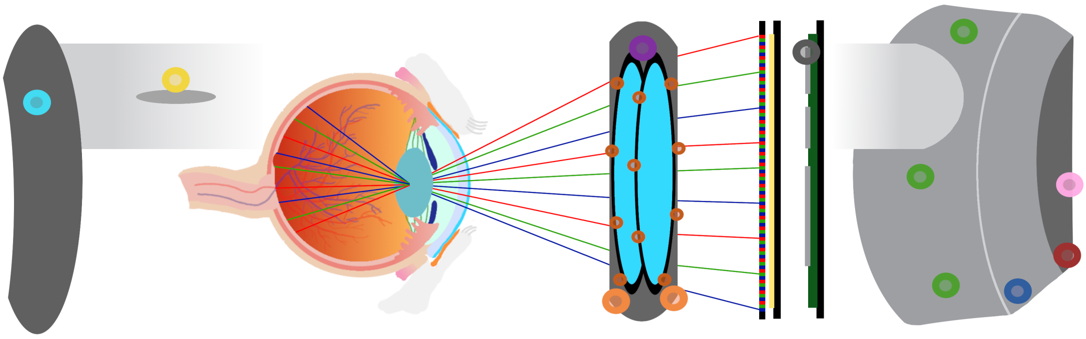


Figure B: Components of an HMD. Cameras for see-through applications (red) and microphones (pink) are often mounted on the front of the enclosure (right). In addition, there are cameras on the bottom for hand and controller tracking (blue). Some models also have infrared LEDs for external tracking (green). Inside the headset are the main circuit board, an IMU sensor for tracking head orientation (grey), a display and two lenses. Depending on the design, the display may have a backlight (as shown here) or not. A light sensor (purple) is often attached to the optical module to determine whether the HMD is currently in use. In addition, many models have two eye tracking cameras (orange) and infrared light-emitting diodes (brown) distributed around the lens, which emit infrared light towards the eyes which cause reflections necessary for eye tracking. The head strap at the back of the HMD (left) often contains headphones (yellow) and batteries (cyan) and allows the display to be worn securely on the head.

On top of a depth sensor, additional sensors aimed at the user's face, can be used to track subtle facial expressions, and together with live audio recordings from microphones, this information can be used to create ever more realistic virtual avatars that enable immersive conversations with other users. In addition, computer-generated three-dimensional avatars are now being developed that can even interact directly with the user via speech (Virk, 2025).

In summary, today's HMDs enable full body tracking while also being small and light. In addition to improved displays and lenses, all necessary sensors, computing and graphics processors can now be integrated into a single head-mounted device. As a result, VEs can now be rendered directly on the wearable device, enabling the exploration of VEs without restrictions.

Eye Movements

After an image is displayed on an HMD, it is passed through the lens and reaches the user. When the eyelids are open, light enters through the cornea, pupil and the lens of the eye (see Figure C). The perceived brightness of the image is not only influenced by the display brightness alone: The diameter of the pupil also plays a role. When looking at bright light or increasingly brighter areas, the iris contracts, causing the pupil to narrow so that less light may enter the eye (Ferree et al., 1933). In dim light, the pupil dilates to increase the incoming photons. Therefore, to a certain extent, our eyes adapt to the given input (which also means that a dark image can be perceived as brighter after an adaptation period).

To see objects at varying distances sharply, the cornea and lens focus the incoming light beam directly onto the retina through contraction of the ciliary eye muscles. With increased age, the eye lens becomes less flexible (presbyopia), which makes it impossible to focus on near objects (Mordi & Ciuffreda, 1998). In case of myopia, distant objects always appear blurred because the light rays reflected from these objects are focused in front of the retina, even when the ciliary muscles are relaxed (Baird et al., 2020).

When the incoming photons reach the retina, around 92 million rod- and approximately 4.6 million cone photoreceptors convert the incident light into action potentials, which are then propagated to the entire visual system (Curcio et al., 1990). Human eyes have three types of cones and one type of rod: the different cone photoreceptors each react optimally to certain wavelengths of light (red, blue and green), whereas the rods are not able to distinguish colours, but are much more sensitive to light and are therefore more suitable for night vision (Prasad & Galetta, 2011).

The density of the cells varies across the FOV: The area with the highest density of cone cells (15,000 cells/degree^2), and thus the highest possible visual acuity of 150 cycles per degree, is the fovea centralis in the centre of the visual field (Miller et al., 1996; Tuten & Harmening, 2021). This means that we can see up to 150 black and white lines spread over 1° of visual angle in the centre of our FOV. Therefore, a minimum resolution of 300 pixels per degree of visual angle is required to display objects with foveal acuity. Currently available HMDs offer resolutions below 50 pixels per degree. However, this is still sufficient to display many small details. Acuity defined as standard vision (20/20) when using a Snellen chart is 30 cycles per degree, or 60 pixels per degree (Caltrider et al., 2023). Towards the periphery, cone density and visual acuity rapidly decreases (Hirsch & Curcio, 1989). Thus, we tend to bring objects of interest to the centre of our FOV by moving our eyes.

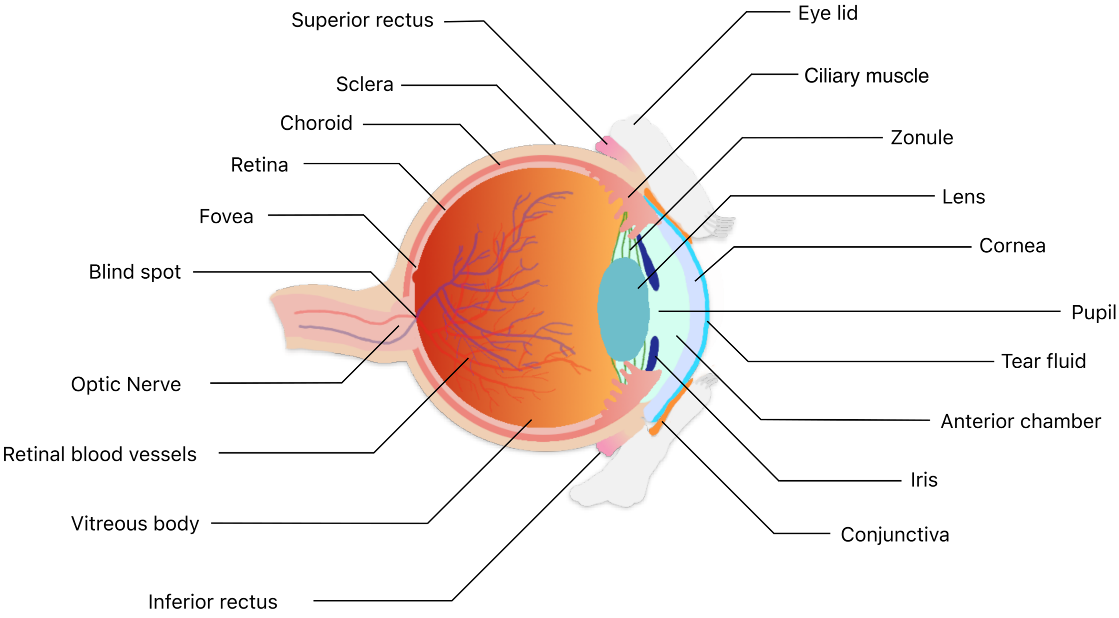


Figure C: Anatomy of the eye. From right to left, light can travel through the tear fluid and the cornea into the anterior chamber of the eye. Through the pupil, whose size can be adjusted using the intraocular muscles, the light reaches the lens, which can change its focus using the ciliary muscles and the zonule. Ideally, it focuses the incoming light precisely on the retina at the back of the vitreous body. Light from the centre of our FOV falls on the fovea, which has the highest density of cone cells. The optic nerve, which transmits visual information to the brain, is located at the blind spot. This is why we do not perceive light that falls on this point. Extraocular muscles enable us to move our eyes, while the eyelid enables us to close them.

Fixating Objects

While we inspect an object with foveal accuracy, our eyes are in a comparatively still fixation state. In general, we fixate for about 90% of the total viewing time (Duchowski, 2007). In free viewing, typical fixations have durations of between 0.1 and 0.4 s (Gajewski et al., 2005). The duration of a fixation can vary greatly, from a few hundredths to a few tenths of a second. In certain real-world situations they can also last several seconds (Hayhoe et al., 2003; Land et al., 1999; Tatler et al., 2011). In other experiments with memory tasks, the fixation time depended on the time required to obtain the necessary visual information for the current action (Droll et al., 2005; Hayhoe et al., 1998, 2003; Land et al., 1999; Tatler et al., 2011).

When we fixate on an object in the far distance, the horizontal angles of the left and right eye are almost parallel. In a situation in which the fixated object comes closer to us, our eyes turn continuously in the direction of the centre of the FOV. Our eyes therefore turn in opposite directions towards each other simultaneously. This movement is called vergence and is usually accompanied by a lens adjustment of the eye by the ciliary muscles, to keep the now closer object in focus. Both movements automatically happen at the same time through the accommodation-convergence reflex. While the visual offset between both eyes can be simulated by presenting stereo images, the screens of current HMDs are displaying images at a fixed focal length. This means that regardless of how far away a virtual object appears (which determines the convergence of the eyes), our eyes must always adjust to the fixed distance of the screen. As a result, unlike in reality, the entire field of view appears to have the same depth of field. Therefore, in VR the usual relationship between accommodation and convergence, to which we are accustomed, is disrupted. This phenomenon is called vergence accommodation conflict (VAC).

Fixational Eye Movements

Our eyeballs can rotate on three rotation-axes using six different muscles: the medial and lateral recti for lateral movements along the yaw axis, the superior and inferior recti for upward and downward movements on the pitch axis and the superior and inferior obliques for twist movements along the roll axis (Davson, 1990). In addition, the eyes can translate in the eye socket, which happens when closing the eyelid (Davson, 1990; Kirchner, Watson, & Lappe, 2022).

When examining fixations carefully and with very high spatial and temporal resolution, it becomes clear that the eyes are actually moving on a very small scale during fixations (Alexander & Martinez-Conde, 2019). Fixational eye movements contribute to fixation stability and allow us to inspect different small details within an object with maximum foveal accuracy (Poletti et al., 2013; Rolfs, 2009). These movements are categorised in three different types: tremor, ocular drift and microsaccades. Tremor is the smallest type of movement with amplitudes of 4/1000° and frequencies that normally range between 30 and 100 Hz (Alexander & Martinez-Conde, 2019; Ditchburn & Ginsborg, 1953; Martinez-Conde et al., 2004). Tremors are likely caused by the firing of motor neurons or the balancing process of the antagonistic eye muscles (Alexander & Martinez-Conde, 2019; Ezenman et al., 1985; Spauschus et al., 1999). Ocular drift moves the eyes minimally and slowly in frequently changing directions (Rucci & Poletti, 2015). Ocular drift's amplitude is usually less than 45 arcmin and its speed is typically around 50 arcmin/s.

Microsaccades are involuntary, short, jagged eye movements that occur once or twice a second during fixation (Martinez-Conde et al., 2013; Rolfs, 2009). Some definitions of microsaccades include eye movements with an amplitude up to 1° or even 2° (Rolfs, 2009). However, the term originally referred to movements with an amplitude of less than 30 arcmin that stabilise an image within the foveola, a tiny area of the retina where cones are the most densely packed (Rucci & Poletti, 2015).

Saccades

To move our gaze from one fixation point to the next, we use saccades with much higher amplitudes. These saccades often reach peak speeds of over 300°/s, making them the fastest eye movements we can perform. Our most common eye movement pattern can be described as a saccade and fixation strategy, in which we regularly switch between looking at one object and making a quick, jerky eye movement to another location (Land et al., 1999). In general, saccade amplitude can vary widely, depending on the ongoing situation (Land & Tatler, 2009a): When reading, small saccades are sufficient, whereas when preparing a meal, most fixations fall on task-relevant objects. Thus, the amplitudes of the saccades depend on the arrangement of the environment (Hayhoe et al., 2003; Land et al., 1999; Schütz et al., 2011). Even objects at a distance of 50° can normally be reached within one eye movement (Land, 2009; Tatler et al., 2011). Up to a certain amplitude range, the peak velocity increases linearly with increasing amplitude, which is also referred to as the saccadic main sequence (Bahill, Clark, & Stark, 1975; Gibaldi & Sabatini, 2021; Smit et al., 1987). Beyond this range, peak velocity saturates.

Saccadic targeting is not perfectly accurate. When looking closely at the saccade landing points of saccades directed at small peripheral objects, it can often be observed that saccades slightly undershoot their target location. Typically, this is then corrected by a second short saccade so that the fovea is then focused precisely on the new fixation object (Tatler et al., 2011).

Saccadic Suppression

During saccades the retinal image is moving. Surprisingly, however, our visual perception is never interrupted by smeared phases during saccades (as compared to a moving camera with a slow shutter speed). Instead, we only perceive a continuous, clear stream of the objects we fixate. While a saccade takes place, visual input is suppressed (Ditchburn, 1955; Wallach & Lewis, 1966). Thus, a light flash that is only presented during a saccade cannot be seen (Bridgeman et al., 1975; Latour, 1962; Zuber & Stark, 1966). The pupil also does not react to intrasaccadic brightness changes (Zuber et al., 1966). Interestingly, saccadic suppression usually occurs a few tenths of a second before the saccade onset reaches its maximum, right at the start of the saccade and disappears gradually a few tenths of a second after it ends (Diamond et al., 2000; Volkmann, 1986). This time course changes depending on various factors, such as luminance characteristics of the background and the presented stimuli, the current light adaptation of the eye, characteristics of the stimulus, the position of the stimulus on the retina and the amplitude of the saccade (Stevenson et al., 1986; Volkmann, 1986).

In voluntary saccades, saccadic suppression is stronger than in reactive saccades

(Gremmler & Lappe, 2017). Nevertheless, in most situations, we are able to recognise a previously perceived object after a voluntary saccade that has then shifted to a different position in the FOV. This allows us, for example, to select one of many apples on a tree, move our gaze, and immediately recognise the same apple again, even though it is now in a completely different position relative to the retina. Thus, our mental representation of our environment is not tied to the object's absolute position in space or to the available visual information during the movement (Irwin, 1991). Instead, it seems we build our mental representation on assumptions that are constructed from visual information after the saccade: For example, we do not notice when an object is displaced during a saccade if the displacement is less than a third of the amplitude of the saccade (Bridgeman et al., 1975). Interestingly the same effect can be observed when the target is already displaced just before the saccade is initiated. In other words, if a fitting post-saccadic target is found just after the saccade at a location not too far away from the object's previous location, we assume that our environment is stationary (Deubel et al., 1996). If no suitable target is found in the required spatio-temporal window, we seem to recalibrate the mental representation of our environment based on other information such as extraretinal signals (Deubel et al., 1996). This recalibration leads to effects such as perceived spatial compression (Burr et al., 2010) This was the case in an experiment in which participants first saw a bar and a ruler. During a saccade, the bar was moved while the ruler disappeared. After the saccade, the ruler reappeared in the same place. The participants perceived the bar as immobile and assumed that the ruler had been moved after the saccade (Matsumiya & Uchikawa, 2003). In general, the compression effect occurs if visual references are available immediately after, rather than before or during, the saccade (Lappe et al., 2000). It therefore seems to be mainly based on visual information that is available after the saccade.

A side effect of spatial compression can be that the number of elements perceived in the environment is reduced (Morrone et al., 2005; Ross et al., 1997). This is because their perceived positions overlap due to the compression. Furthermore, the compression effect is not only limited to space, but can also influence time perception: During a saccade, the perception of time intervals can also be compressed. This can shorten the perceived duration by about half and even reverse the perceived temporal sequence of successive stimuli (Burr et al., 2010; Morrone et al., 2005).

Blinks

Another obvious example of suppressed visual perception is blinking. During blinking, the eyelid closes, the eyeballs rotate downwards and slightly translate towards the skull (Kirchner, Watson, & Lappe, 2022). Although blinking, like saccades, occurs regularly and is easy to observe from the outside, we do not normally perceive the interruptions to our stream of visual perception. Instead, again it seems as if our visual system skips parts of our input. Suppression of visual input during blinking appears to be associated with an inhibitory signal from the brain triggered by efferent discharge during eyelid closure (Volkmann et al., 1982). After 100-150 ms, the eye usually returns to its previous position and the eyelid opens again (Volkmann, 1986). The recovery of the suppression of the visual input during the blink occurs gradually over a period of 100-200 ms after the onset of the blink and is therefore somewhat faster compared to saccades (Volkmann, 1986). Interestingly, there does not appear to be a hard-wired physiological mechanism that ensures visual stimuli are suppressed while we keep our eyes closed, as we can still perceive visual stimuli even with our eyelids shut. Instead, suppression seems to be mainly related to the closing of the lid. Although most light is blocked when the eyes are closed, it is thus possible to detect differences in brightness (Bierman et al., 2011; Sakai, 2023). This works especially well with light sources with long wavelengths and can be tested easily when being outdoors by moving the head towards or away from the sun with the eyes closed.

In general, blinks can be clustered in different subtypes. During reflexive blinking, the eyelid is closed to protect the eye or to moisten it with tears (Evinger, 1995). This blink reflex is, for example, triggered when air is blown into the eye or an object is quickly moving towards the face. Typically, we make 10–15 spontaneous blinks per minute, which is more than necessary to just prevent a dry cornea. Thus, we seem to also blink in situations without the need to protect the eyes. Moreover, our blink rate varies with different tasks (Doughty, 2001; Karson et al., 1981). A consciously performed closing of one or both eyelids, is called a voluntary blink or wink. While we blink, we can also perform other eye movements. For example, it is possible to perform saccades during a blink (Kirchner, Watson, Busch, & Lappe, 2022; Rottach et al., 1998). Although the kinematics of these saccades are significantly altered, their parallel execution could be useful in reducing the time during which only very limited visual stimuli are available.

Gaze

To look at objects more closely, we move not only our eyes but also our head. In this thesis, combined eye-head movements will be referred to as gaze movements. Horizontally, eye amplitudes can reach values around 50°, whereas head rotation amplitudes can even reach 175° (T. C. Wu & Tsotsos, 2025). Thus, gaze shifts are useful to reach objects beyond our FOV (Einhäuser et al., 2007; Foulsham, 2015; Thier & Ilg, 2005). In this case a coordinated head movement extends the saccade amplitude. However, a large proportion of eye and head movements are performed in opposite directions (Einhäuser et al., 2007; T. C. Wu & Tsotsos, 2025). This is, for example, the case, when we fixate a stationary object while moving our head. Then, vestibular movement signals that can detect head movements are transmitted to the eye movement control circuits (Cullen & Roy, 2004). This signal is the basis for the vestibulo-ocular reflex (VOR), through which our head movements are compensated by counter-rotating eye movements with low latency, so that a more stable retinal image is generated (Fetter, 2007). However, the image cannot be perfectly stabilised across the entire FOV because the eye position also shifts due to head movement, as the eyes are not positioned in the centre of the head's axis of rotation (Harris, 1994). Therefore, this translational shift must either be ignored or stabilisation must be optimised for a specific object of interest.

To coordinate gaze movements accurately, the neural circuits for head and eye movement are linked closely: A case report on a patient with eye muscle paralysis even showed that when eye movements are not possible, saccade-like head movements are performed instead (Gilchrist et al., 1997). During head movements, the speed of eye drift increases threefold, suggesting that drift also helps to compensate for head movements (Aytekin et al., 2014; Rucci & Poletti, 2015; Skavenski et al., 1979). The properties of the microsaccades, however, do not change, so it appears that they capture fine spatial details of objects independently of head movements (Jolly et al., 2023).

Gaze Stabilisation

When we try to fixate objects that are moving through the environment, we move our eyes to compensate for their motion. Such ocular following responses are evoked with very short latencies by sudden visible motion (Miladinović et al., 2024). The following continuous slow eye movement that tracks a moving target is called smooth pursuit (Drewes et al., 2021; Haarmeier & Thier, 1999; Haarmeier et al., 1997). To reach a moving target, our eyes can first adjust their velocity to the objects' speed and direction, then a saccade can bring the eyes to the target's position while the smooth pursuit continues (Lisberger, 2015; Rashbass, 1961). During pursuit, the eyes can smoothly compensate for speeds of up to 30°/s. At higher speeds, smooth pursuit movements are usually accompanied by catch-up saccades in motion direction. Interestingly, tracking an object using pursuit eye movements requires a centrally-derived motion percept of the target (Steinbach, 1976). This can be based on visual signal of the object's motion, such as the visible edge of the tracked object. We can also follow acoustic or sensory signals, such as when our eyes follow our own hand movements or a sound in the dark (Berryhill et al., 2006; Hashiba et al., 1996; Lisberger, 2015; Rashbass, 1961). If perception of the object movement fails, we can no longer follow it smoothly with our eyes. Instead, the object is tracked through a series of saccades and smooth pursuit phases in an attempt to follow it, while the target appears to jump in space (Koerfer et al., 2024).

When we perceive motion anywhere across our FOV, for example when looking out of the window of a train, a similar reflex called optokinetic nystagmus is triggered. Initially, we tend to again initiate a smooth eye movement, whose dynamic is similar to pursuit (Magnusson et al., 1986). Next, however, our eyes do a fast eye movement in the opposite direction of the visible motion, where the cycle restarts and another smooth movement in motion direction begins (Drewes et al., 2021). As a result, we perceive the object clearly, although everything in our FOV is in motion.

Eye Tracking

Eye movements are an essential part of our visual system. When we observe another person's eyes, we can easily detect whether a saccade or blink is occurring. Although we intuitively have a rough idea of where the other person is looking, it is difficult to analyse the rapidly changing movements more precisely based on manual observation alone. Consequently, the first attempts at automated, objective measurement of eye movements were made over a century ago and have been supplemented over the decades by numerous further developments and new methods of eye and gaze tracking (S. B. Hutton, 2019).

Initially, both Huey and Delabarre used a mechanical device consisting of a contact lens with rods attached to it, which marked the ongoing eye movements on a kymograph cylinder (Delabarre, 1898; Huey, 1898; S. B. Hutton, 2019). With this method, slow eye movements could be recorded, although the weight of the measurement equipment probably added noticeable strain on the eye muscles and therefore influenced the measured eye movements to some extent. The weight could be reduced further, when Robinson introduced contact lenses with wire coils inside (Robinson, 1963). Their movement could be measured using an electromagnetic field surrounding the head of the subject. To this day, this method offers a comparatively high degree of precision, with approximately 5 arc seconds over a limited range of approximately 5° (Young & Sheena, 1975a). Although wearing comfort has improved significantly and today's lenses no longer require anaesthetic drugs such as cocaine to be applied to the cornea before insertion, wearing a comparatively thick lens can still be unpleasant for the subject and insertion requires practice (Duchowski, 2007).

A less invasive approach is electrooculography (EOG). This eye tracking method makes use of the standing electric potential between the anterior positive and posterior negative pole of the eye. Using electrodes placed around the eye, one can determine the eyeball orientation with respect to the head position, although it lacks some spatial accuracy, especially on the vertical axis (Duchowski, 2007). This data can be collected at very high sampling rates. Unfortunately, the signal often drifts slightly over a short amount of time since the electric potential of the pigmented epithelium also changes based on the amount of light that enters the eye. Therefore, although accurately measuring the eye direction with EOG requires very frequent recalibration, the method is well suited to detect saccades and blinks with low latency (Bolte & Lappe, 2015; Kirchner, Watson, Busch, & Lappe, 2022).

The first predecessor of today's most common eye tracking method was the photochronograph by Dodge and Cline (Dodge & Cline, 1901; S. B. Hutton, 2019). Instead of attaching measuring equipment directly on to or close to the eye, this method measured movements remotely. First, a light source was directed at the user's eyes. Since eyes are not perfectly spherical, the reflection of light on the cornea changes depending on the direction of gaze. The reflected light was then exposed to a film plate that was later analysed. Therefore, this method was only suitable for visualizing the accumulated eye movements over a certain period of time. In revised devices, horizontal and vertical components of the eye reflection could be separated and exposed with higher temporal resolution on photographic tape or film (Buswell, 1935; Judd et al., 1905; Macele & Mueggenburg, 2024).

Later, Yarbus (1967) developed contact lenses that were attached to the eye with suction cups and to which either eye-fixed visual stimuli or tiny mirrors were attached, which also made it possible to record eye movements on film (S. B. Hutton, 2019; Tatler et al., 2010).

Eventually, Yarbus' approach led to a number of different remote eye trackers, all of which used different types of light reflections from the eye to determine the direction of gaze. To allow the subject to still see a stimulus while using such an eye tracker, the remote light source was soon replaced by infrared light. Infrared limbal eye trackers exploit the fact that the white sclera of the eye reflects more light than the iris, using measurements from a high-frequency infrared sensor as a signal for the direction of gaze on the horizontal or vertical axis (S. B. Hutton, 2019).

On closer inspection, not just one, but four visible reflections of incoming light (also known as Purkinje reflections) can be observed in the eye: these come from the anterior and posterior surfaces of the cornea and the lens of the eye. Dual Purkinje eye trackers measure the disparity between the first reflection from the cornea and the fourth reflection from the eye lens (Cornsweet & Crane, 1973; S. B. Hutton, 2019; Richardson & Spivey, 2008). In the first Purkinje eye trackers, this measurement was done mechanically: Adjustable mirrors were aligned in such a way that the two reflections overlapped. This calibration procedure allowed accurate (up to 1 min of arc) and fast eye direction estimations and was therefore suited to measure microsaccades, if the head was stabilised carefully.

As soon as camera equipment enabled real-time video recordings of eye movements, video-oculography (VOG) became the most popular method of estimating gaze direction based on light reflections of the eye (Merchant et al., 1974; Richardson & Spivey, 2008; Young & Sheena, 1975b). The most common technique in VOG is dark pupil tracking. As in limbal eye trackers, this approach makes use of the fact that different features of the eyes reflect different amounts of light. In this case, since the pupil reflects even less light than the iris, the pupil can be detected as a dark small ellipse in each frame of the recording. This ellipse, can then be compared with the first corneal reflection that is induced using an infrared emitter. The distance between the two can be calibrated to estimate the gaze direction. In some devices, multiple infrared light sources and therefore multiple first order Purkinje reflections are used (see Figure D). Moreover, the dual Purkinje eye tracker has been further developed with the help of camera technology: Instead of analogue mechanical parts, a digital dual Purkinje eye tracker uses an optical setup (with no moving components) and a digital imaging module (R.-J. Wu et al., 2023). Again, to track eye movements meticulously, the offset of the first and fourth Purkinje reflections are used, now obtained from a digital image.

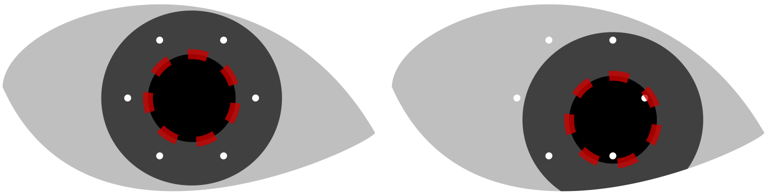


Figure D: Video-based eye tracking. With every eye movement, the pupil changes its position slightly (left vs. right). By comparing the tracked pupil (red dotted circle) with the tracked corneal reflection of several infrared emitters (white dots), the pupil position can be mapped to positions throughout the FOV through calibration.

Thanks to technological advances in cameras, processing hardware and batteries, VOG devices can be so lightweight that they can be integrated into wearable devices to track eye movements in natural and virtual environments. However, apart from contact lenses with wire coils, the presented tracking methods only measure the eye movements relative to the head. To track the gaze relative to a person's surroundings, the orientation and position of the head must also be taken into account. In the case of wearable eye trackers, this can be done using an outward-facing camera, from whose recordings of the FOV the tracked gaze direction can later be visualised. When researching human behaviour in VEs, the data from the position tracking system, the IMU and the eye tracker can be combined to obtain a gaze direction signal. This signal can be both recorded and also used to make the VE react to the user's eye movements.

Gaze-contingent Displays

Even before Sutherland built his first HMD prototype, he envisioned the capabilities of the ultimate display of the future (Sutherland, 1965). Interestingly, despite the limited technology of the 1960s, his vision included a display that was capable of eye tracking. In his essay, Sutherland laid out an experiment in which the device would display a triangle whose edges would round off as soon as a user fixated them. He hoped that this approach would contribute to a better understanding of the mechanisms of human vision. To implement this idea, the eye tracking signal must be available in near real time so that visual stimuli can be updated in a timely manner. Such so-called gaze-contingent displays (GCDs) were developed a few years later in the 1970s, initially for reading research (Duchowski et al., 2004; McConkie & Rayner, 1975; Rayner, 2014). However, these GCDs were not attached to the head, but worked with remote eye trackers and screens positioned on desks. Later, the same principle was applied again with the aim of improving system performance when displaying computationally intensive stimuli (Murphy & Duchowski, 2002). In foveated rendering, for example, the level of detail or resolution of visual content in the periphery is reduced. It is therefore important that eye movements with a large amplitude in particular are detected quickly so that the image can be updated accordingly, giving users a detailed, high-resolution image in the centre of their FOV (Duchowski et al., 2004; O'Sullivan et al., 2003; Patney et al., 2016).

In addition, GCDs met the expectation that they could deepen our understanding of our visual system. They enabled experiments on the time span of the perceptual window, which showed that we process up to 15 letters in reading and about 5° in searching (Bertera & Rayner, 2000; McConkie & Rayner, 1975). With the help of GCDs, it became possible to measure how vision degrades from the fovea to the periphery. It was found that this can vary depending on image characteristics such as spatial and temporal resolution, colour, luminance and contrast (Duchowski et al., 2004; Loschky & Wolverton, 2007). Further experiments with desktop-based GCDs showed that contrast sensitivity for natural stimuli is significantly reduced compared to static images (Dorr & Bex, 2011), gaze-contingent contrast influences colour perception (Mauderer et al., 2016) and high cognitive load leads to a loss of accuracy evenly distributed over the eccentricities of the retina (Gaspar et al., 2016).

GCDs were also used in retinal visual impairments simulations that mimicked macular degeneration due to ageing (Geisler & Perry, 2002; Vinnikov et al., 2008). With the help of such applications, circumvention strategies of the visually impaired could be better understood and environments suitable for the disabled could be evaluated more easily. In addition, personally experiencing impairment simulations could help to raise awareness and increase empathy among the public. Finally, gaze-contingent HMDs could compensate the pupil swim effect and improve visible cues for the perception of depth in VR (Konrad et al., 2020; Rui et al., 2023).

An ideal gaze-contingent HMD would have a refresh rate above the fusion flicker perception threshold and an eye tracking latency below the frame to frame interval. This would require very fast and expensive sensors that consume a lot of energy, heat up quickly and cannot be operated for long periods on battery power. However, because we use saccades to move our eyes from one fixation to another and vision is suppressed, during and shortly after each saccade, it is possible to present gaze-contingent, world-fixated stimuli using hardware with higher end-to-end latencies.

The maximum possible latency of the GCD at which the user cannot see the adjustment of the stimuli also depends on the saccadic main sequence. This is because the phase of maximum suppression and the recovery of vision take longer with saccades with higher amplitudes. (Stevenson et al., 1986; Volkmann, 1986). Thus, a voluntary saccade with an amplitude of 8° results in suppression times around 40 ms after saccade onset, whereas a voluntary saccade with an amplitude of 16° results in suppression times around 80 ms after onset (Stevenson et al., 1986).

This, however, does not mean that larger saccades are always helpful when setting up a gaze-contingent experiment. For example, in foveated rendering, objects in the periphery are displayed with a lower resolution, which is why a degradation factor must be implemented that adjusts the resolution of a stimulus based on the distance to the centre of the FOV. Since saccades with longer amplitudes land further away from the original fixation, gaze-dependent manipulations such as low resolution may appear more obvious to the user after a saccade, if updated even shortly after the saccadic suppression period, while small saccades may still fall in a higher resolution area where the manipulation is less likely to be detected even if the visual content is updated too late (Loschky & Wolverton, 2007). Therefore, the end-to-end latency from an eye movement to the change in display content, the expected saccade amplitude and the gradient of the manipulation must be carefully considered when setting up gaze-contingent manipulations that should not be perceived.

Requirements for Gaze-contingent HMDs

Since some currently available HMDs are equipped with eye trackers, it may be possible to use them as gaze-contingent HMDs. In contrast to desktop-based gaze-dependent setups, these could enable stimuli to be presented at specific positions in the FOV, independent of both eye movements and head movements, which would allow the presentation of gaze-contingent stimuli in VEs.

In order to assess whether this is possible, it is first necessary to estimate the minimum latency requirements for GCDs. When using a remote eye tracker and a desktop screen to create a gaze-contingent setup, no differences in eye movements were observed between an artificial latency of 5 ms and 15 ms (Loschky & McConkie, 2000). At an added latency of 45 ms, however, Loschky and McConkie (2000) observed slightly longer fixation durations, although even then, the performance in a search task was not altered. In a detection task of gaze-contingent peripheral blur, performance did not significantly decrease until a latency of 60 ms was reached (Loschky & Wolverton, 2007). Although these observations do not lead to a clear maximum latency for a GCD, they can give an idea of the required latency range which is between 15 and 60 ms.

Displays in currently available HMDs have a frame rate of about 90 Hz (11.1 ms). However, to implement a VR-GCD, not only the frame-to-frame interval, but the overall end-to-end latency from a gaze shift to an image that is displayed on the HMD needs to be considered. In that time frame, multiple processing steps need to be done. First, motions from all available tracking sensors need to be recorded. Then, the tracked signal needs to be used to estimate the current state of, head and eyes with accuracy and precision. Next, the VE needs to be updated based on this estimate. Finally, the updated VE needs to be rendered to a stereo-image that is then displayed on the HMD.

The latency between a change in position and a corresponding change on the display can be determined externally by simultaneously recording the start of the HMD's position change and the display with a high-speed camera. For head tracking, current HMDs have motion to photon latencies between 21 and 42 ms, which is below the required latency range for GCDs (Niehorster et al., 2017; Warburton et al., 2023).

Because the HMD occludes the eyes, a different method is required to measure the latency of VR eye trackers. We present this method in the following chapter.

The development of HMDs has come a long way since 1968. Prototypes have become commercially available products that are used for research. At the same time, by continuously innovating new eye tracking methods, we have increased our knowledge about our visual perception and how it is influenced by the way we move our eyes. In order to investigate the potential use of VR with eye tracking in psychophysical research, for example to present strictly controlled visual stimuli while simultaneously measuring head, eye and walking movements, we must be able to estimate the latency of the tracking sensors. In the following study, we address this need by measuring the time required by currently available HMDs to register an actual saccade, record the corresponding eye tracking data and finally display a change on the screen depending on the recorded data.

A Comparison of Eye Tracking Latencies Among Several Commercial Head-mounted Displays

NIKLAS STEIN, DIEDERICK C. NIEHORSTER, TAMARA WATSON,
FRANK STEINICKE, KATHARINA RIFAI , SIEGFRIED WAHL &
MARKUS LAPPE

This chapter was published in i-Perception (vol. 12) in 2021

Abstract

A number of VR HMDs with integrated eye trackers have recently become commercially available. If their eye tracking latency is low and reliable enough for gaze-contingent rendering, this may open up many interesting opportunities for researchers. We measured eye tracking latencies for the Fove-0, the Varjo VR-1 and the HTC Vive Pro Eye using simultaneous electrooculography measurements.

We determined the time from the occurrence of an eye position change to its availability as a data sample from the eye tracker (delay) and the time from an eye position change to the earliest possible change of the display content (latency). For each test and each device, participants performed 60 saccades between two targets 20° of visual angle apart. The targets were continuously visible in the HMD and saccades were instructed by an auditory cue. Data collection and eye tracking calibration was done using the recommended scripts for each device in Unity3D. The Vive Pro Eye was recorded twice, once using the Steam VR software development kit (SDK) and once using the Tobii XR SDK.

Our results show clear differences between the HMDs. Delays ranged from 15 to 52 ms and latencies from 45 ms to 81 ms. The Fove-0 appears to be the fastest device and best suited for gaze-contingent rendering.

Introduction

In the past decade, the rapid commercialisation of VR equipment has made HMDs available to a much larger number of researchers. Head position tracking is included in most devices and has opened interesting avenues to study perception with naturally moving observers in small areas (Bansal et al., 2019; Niehorster et al., 2017; Scarfe & Glennerster, 2015; Warren et al., 2001; Wexler & van Boxtel, 2005). Now, the integration of eye trackers in HMDs promises to be the next technological step, allowing even more immersion in virtual environments (VEs).

Since the first objective measurement of eye movements by Delabarre (1898) and Huey in 1898, eye tracking techniques have come a long way (Wade, 2010). In the 1950s, the first contact lenses including coils were used to precisely record where a subject is looking (Holmqvist et al., 2011). EOG, a less invasive method, was the most applied method in the 1970s (Young & Sheena, 1975b). By measuring the changes of the electric field evoked by changes of the position of the pigmented epithelium using electrodes placed around the eye, one can determine the eyeball orientation with respect to the head position. Although the data can be collected at high sampling rates, EOG lacks spatial accuracy, especially on the vertical axis (Duchowski, 2007). Thus, today's most used method is video-based eye tracking, combining a non-invasive method with sampling rates up to 2000Hz and an advertised spatial accuracy down to 0.15° (SR Research, 2020). Progress in camera technology made it possible to create head-mounted eye trackers (including cameras filming the wearer's FOV) that estimate eye positions in the real world without explicitly recording head movements (Niehorster, Hessels, & Benjamins, 2020; Niehorster, Santini, et al., 2020). The technology is now included in HMDs, promising advances in wearing comfort by software that helps the user to adjust the position of the HMD on the head to optimise optical parameters to the location of the eyeballs. Moreover, it allows analysing the location and length of fixations in VEs and the implementation of new interaction methods (Piumsomboon et al., 2017).

Now that eye trackers in HMDs have become available from multiple vendors, we may ask for what kinds of application or experiment they are suitable for. From their low sampling rate (generally around 100Hz) it is clear that these eye trackers cannot be used for analyses of trial by trial saccade dynamics (Mack et al., 2017), although it should be noted that successful attempts have been made to determine saccade dynamics through a modelling approach using low sampling rate low noise data, when averaging over many saccades (Gibaldi & Sabatini, 2021; Wierts et al., 2008). It remains an open question whether these methods would work reliably on the offline data provided by the eye trackers examined in the current study.

In contrast, for the study of looking behaviour by means of recording fixations (periods during which the image of an object is held stable on the retina (Hessels et al., 2018)) the lower sampling rates do not pose significant issues (Andersson et al., 2010). For measures based on fixations, the important factors are accuracy and precision (Niehorster, Zemblys, et al., 2020), which may affect the minimum size of objects to which gaze can be reliably resolved (Hessels et al., 2016, 2017; Holmqvist et al., 2012). Assessment of these aspects of eye trackers in HMDs is thus an important topic for those types of applications (Adhanom et al., 2020).

In the present study, we are concerned with a third type of application, gaze-contingent display paradigms in perceptual research. Gaze-contingent displays rely on low-latency access to eye tracking data and the ability to rapidly adapt the visual stimulation presented in the HMD in response to an eye movement. In such conditions, a host of perceptual effects related to eye movements can be studied and used in applications: Examples include, first, blinks which can be detected (Alsaeedi & Wloka, 2019) and used to apply non-perceivable modifications to a VE (Langbehn, Steinicke, et al., 2018). Second, saccadic omission, often also called saccadic suppression, refers to the phenomenon that stimuli presented during a saccade are often not consciously perceived (Campbell & Wurtz, 1978; Diamond et al., 2000; Holt, 1903). This includes the self-generated motion on the retina that occurs as the eye sweeps across the scene during the saccade (Ilg & Hoffmann, 1993). Saccadic suppression of displacement refers to the failure to perceive a displacement of a persistent stimulus in the scene during the saccade, or even the saccade target itself (Bridgeman et al., 1975). Change blindness refers to the failure to notice changes to objects in the scene, including their colour or position, as well as the introduction or removal of objects (Rensink et al., 1997; Simons & Levin, 1997). In addition, there are also several phenomena of changes to the perceived position of objects briefly presented at the time of a saccade (Lappe et al., 2000; Matin & Pearce, 1965; Ross et al., 1997). Next to the spatial aspects of scene vision, there are also changes to temporal processing associated with saccades. Chronostasis, i.e., the stopped clock illusion, refers to an illusion in which time appears to briefly halt during the saccade (Yarrow et al., 2001), likely because of deployment of attention towards the saccade target (Georg & Lappe, 2007). Moreover, temporal intervals are misperceived and the temporal order of successively presented stimuli may even be reversed (Morrone et al., 2005).

These perceptual phenomena are commonly studied with gaze-contingent experimental paradigms. The ability to use such paradigms with HMDs in VR would open up many new research avenues with more naturalistic settings, for example in the combination of eye and head movements (Anderson et al., 2020). Moreover, perceptual phenomena can also be applied to enhance VR scenarios (Bolte & Lappe, 2015; Bruder et al., 2013; Sun et al., 2018). There are several applications that make use of these techniques, for example, environments with artificial scotoma (Bertera, 1988) and gaze-contingent foveated rendering techniques (Patney et al., 2016). In gaze-contingent rendering, the resolution of the rendered image can be imperceptibly reduced in the periphery since peripheral visual resolution is much lower than foveal visual resolution. Using eye tracking, the location of a high resolution area in the image can be updated according to gaze direction.

Besides perception, another common use of gaze-contingent paradigms is the study of calibration of motor functions of the brain, involving stimuli that change during saccades. Changes applied during a saccade can have long-lasting effects mediated by implicit learning. Consistently applied displacement of the saccade target leads to adjustments in saccade amplitude known as saccadic adaptation (McLaughlin, 1967; Pélisson et al., 2010). These adjustments can be specific for different depth planes in the three-dimensional space (Chaturvedi & Van Gisbergen, 1997). Moreover, recalibration can also occur in opposite directions in each eye at the same time, suggesting the existence of independent monoculomotor and binoculomotor plasticities for each eye (Maiello et al., 2016). Saccadic adaptation involves not only motor learning but also persistent changes in perception of visual space (Collins et al., 2007; Zimmermann & Lappe, 2016). Trans-saccadic changes of the size of a target object lead to a re-calibration of size perception in peripheral vision (Bosco et al., 2015; Valsecchi & Gegenfurtner, 2016) and to a re-calibration of grasping action (Bosco et al., 2015). Also, peripheral shape or colour perception can be modified by trans-saccadic changes (Herwig & Schneider, 2014; Paeye et al., 2018).

For all of these applications, correct timing of the respective stimulation or display change in relation to the saccade is essential. These phenomena are transient and occur typically only up to the end of the saccade or some few tens of milliseconds thereafter, often with a common time course (Burr et al., 2010; Ross et al., 2001). For example, saccadic suppression fades out by the end of the saccade (Diamond et al., 2000), as do possible impacts on localisation or timing. Saccadic suppression of displacement occurs if the displaced stimulus is visible at the end of the saccade, but not if it is instead presented 100 ms later (Deubel et al., 1996). Saccadic adaptation also works best if the displaced target is present at saccade onset and less well if it is delayed (Fujita et al., 2002). From the time courses of these effects it is clear that effective methods for gaze-contingent displays for vision research, although dependent on the specific paradigm, should ideally be made with stimulus delays that do not exceed the duration of the saccade by more than a few tens of milliseconds (Loschky & McConkie, 2000). Similar requirements exist for foveated rendering in VR (Albert et al., 2017).

In the present study, we evaluated eye tracking latency of currently available HMDs to investigate their suitability for GCD paradigms. We recorded delays of eye trackers in three commercially available HMDs and compared them to simultaneously recorded EOG. As a current gold-standard for comparison, we perform the same analysis on a stationary, screen based eye tracker (Eyelink 1000) that is well established in the vision research community.

Setup

In this section, we describe the hardware, software and calibration method for each setup, including versions and configurations. All recordings were done on the same computer with an Intel Core i9 9900K (3.6 GHz) processor, 8 GB RAM and an NVIDIA RTX2080 graphics card (driver version 26.21.14.3086). The operating system was Windows 10 (Version 1809, Build 17763.914) with the NVIDIA driver set to maximum performance and minimum latency mode. To obtain consistent timestamps, we implemented and compiled a CPU-based high precision clock script that could be used across all software packages. In all tests, we used the first timestamp after a sample was available to the recording software (Unity3D, 2019.1.8f1). Recordings were made using the following devices:

Head-mounted Displays

Fove-0

The Fove-0 was the first commercially available HMD that included eye tracking. The first devices were delivered in January 2017 and included a camera based eye tracking system with an advertised spatial tracking accuracy of less than 1°. The maximum eye tracking sampling rate is 120Hz. The display has a refresh rate of 70Hz. We used it with the FoveInterface prefab of the Fove-0 Unity package (Version 0.16). This approach reduces the rate of eye tracking samples to 70Hz.

Although this approach does not use the full 120Hz capability of the eye tracker, the method of synchronizing to the display frame rate allows receiving the most recent data point at the start of each frame and to manipulate the next frame accordingly. Access to data at higher sampling rates would not allow changing the effective latency of the whole system since manipulation of the visual stimuli can only be done frame by frame. Unity's XR settings were set to Stereo Display.

Varjo VR-1

The Varjo VR-1 was introduced in October 2019 and includes a secondary (foveal) display to provide a higher resolution in the centre of the main 87° FOV display. The eye tracker samples at 100Hz and the display has a refresh rate of 90Hz. Our recordings used Unity3D (2019.1.8f1), the Varjo Base Software (Version 2.2.1.17) and Steam VR. The Varjo API provides for each frame a list with one or two eye data samples. We always used the most recent sample to manipulate the next frame. Unity's XR Settings were set to Open VR.

Vive Pro Eye

The Vive Pro Eye was introduced in June 2019 and is an upgraded version of HTCs Vive Pro that includes the Tobii eye tracking system that is also present in other headsets (e.g., Pico Neo 2 Eye and Tobii HTC Vive Devkit). The headset has a declared eye tracking accuracy of 0.5° – 1.1° at 120Hz, a 110° FOV and a display with a refresh rate of 90Hz. We used it with Unity3D (2019.1.8f1), SRanipal (v1.1.0.1) and Steam VR. Unity's XR Settings were set to Open VR.

Comparison Devices

EOG

For EOG recordings simultaneously to data collection from the HMD eye trackers, we used the same setup as described in (Bolte & Lappe, 2015). It consisted of a BioVision EOG amplifier (BioVision, Germany) used with two self-adhesive electrodes next to each eye and a third electrode as ground on the left cheek (see I.1). Amplified EOG signals were read in via analogue ports of a Meilhaus RedLab-1208FS PLUS sampling device. Because the pigmented epithelium in the retina evokes changes of an electric field when saccades are made, the difference between the electrodes next to the subject's eyes provide zero-latency signals of horizontal eye movement. Our setup collected EOG data at sampling rates of approximately 800Hz. Sampled EOG data was recorded in Matlab R2018b and Psychtoolbox Version 3.0.14 concurrently with the HMD's eye tracking data.

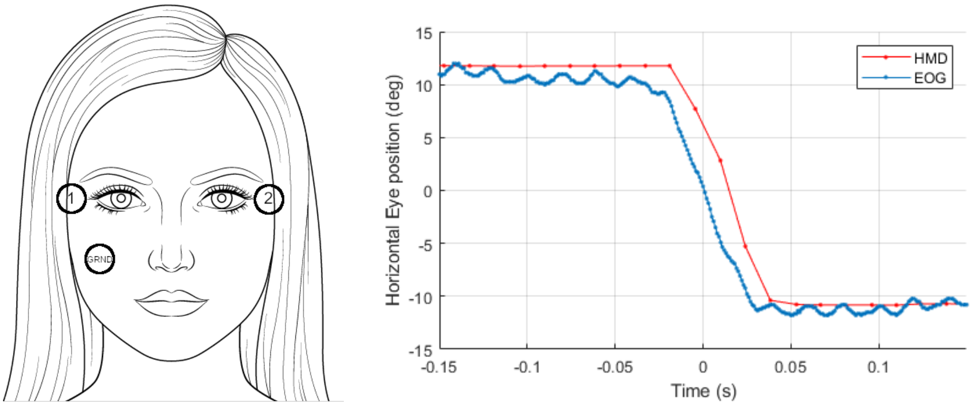


Figure I.1: Electrooculography (EOG). Left: placement of the electrodes to record the electronic signal. Right: shifted and aligned EOG and HMD signals, where the EOG crossed 0° is defined as 0 ms.

Eyelink 1000

For comparison to the HMD eye trackers, we included the Eyelink 1000 from SR Research as a baseline device for its high temporal resolution (1000Hz) and low latency. It has become a gold standard of eye tracking research in the past decades and is extensively used in latency critical experiments. The stationary eye tracker was connected via an Intel onboard Ethernet port (I219-V). Stimuli were presented on a FlexScan 930 screen from EIZO with a resolution of 1280 × 1024 pixels. Movements of the head were minimised by using a chin and head rest, allowing a declared latency of below 1.8 ms (SR Research Ltd., 2009). The device was set to monocular recording to achieve the highest possible sampling rate (1000Hz).

Study 1: Eye Tracking Delay

In this study, we measured the time it takes the HMD's eye tracker to register a change in eye position and deliver this information to the experimental software. We call this the delay of the eye tracker. We measure it by registering the same eye movement in EOG and HMD and calculate the time difference between the two signals.

Methods

Participants

We measured all setups with four participants (3 male, 1 author, all employees of the University of Münster, normal vision, no contact lenses, no glasses, all experienced with eye tracking experiments, age 25–46). Informed, written consent was obtained from all participants.

Procedures

In four sessions, one for each of the HMDs (Eyelink 1000, Fove-0, Varjo VR-1, Vive Pro Eye), the participants made 60 horizontal saccades of 20° in-between two persistently visible targets from left to right and vice versa. The start of all trials was indicated by an auditory cue, which was played alternating between 0.5s and 1s after the start of the previous trial. Participants were instructed to fixate a target and make a saccade to the opposite target after an auditory cue.

Data analysis

For the measurement of the delay of the HMD eye tracker, we matched saccade data from the EOG and the HMD (only left eye used) and determined the temporal offset between them (Figure I.1). The EOG signal, which is recorded as a voltage, was scaled for each saccade to match the repeated gaze direction at start and end points of the matching saccade from the HMD tracker, provided in degrees of visual angle. Then, the point in time at which the eye position computed from the EOG signal crossed the centre of the screen (0°) was taken as the reference time point (0 ms) (see Figure I.1). Next, data samples (time stamps and eye positions) were averaged, separately for EOG and HMD data, across all valid saccades to derive a mean saccade time course for each subject. Saccades smaller than 7.5° were discarded (4.5% of all saccades, see Table I.1). All saccades were manually checked to ascertain their validity.

Finally, to calculate the tracker delay, time lags of HMD samples of eye positions during the saccade (i.e., within a range from 20% to 80% of the mean saccade amplitude) were taken relative to the same eye position samples of the EOG.

	Fove-0	Varjo VR-1	Vive Pro Eye	Tobii XR	Eyelink 1000
Study 1					
Eye tracking delay (ms)	15 (2.5)	36 (3.3)	52 (3.2)	-	0 (3.4)
Number of saccades	236 (98.33%)	218 (90.83%)	232 (96.66%)	-	231 (96.25%)
Sampling rate	70.1	99.1	91.8	-	1000
Study 2					
Eye tracking delay (ms)	16 (1.3)	36 (2.7)	50 (1.8)	51 (1.2)	-
End-to-end latency (ms)	45 (6.1)	57 (6.6)	79 (5.4)	81 (2.6)	-
Display delay (ms)	29	21	29	30	-
Number of saccades	177 (98.33%)	170 (94.44%)	171 (95%)	174 (96.66%)	-
Sampling rate	69.7	98.1	88.3	82.5	-

Table I.1: Overview of eye tracking comparison results. Standard deviations and % of planned saccades are shown in brackets. The analysis only includes saccades that could be detected in both the optical eye tracker and the EOG. The EOG achieved a mean sampling rate of 805Hz (study 1) and 1196Hz (study 2). Display delay in study 2 was determined as the difference between end-to-end latency and eye tracking delay for each device. The display delay can include things such as rendering time, transmission time of the signal and latency of the display.

Results

Figure I.2 shows the averaged eye positions during saccades for each subject in each of the devices. The comparison between the Eyelink 1000 and the EOG shows no delay between the two methods (0 ms). Among the three HMDs, the Fove-0 shows the lowest delay (15 ms), followed by the Varjo VR-1 (36 ms) and the Vive Pro Eye (52 ms). Figure I.3 gives an overview of averaged rightward saccades of all the devices.

A linear model was fitted to predict the *Mean Delay* between each eye tracker and the EOG using *Device* and *Saccade direction* as factors and the Eyelink data as baseline (adjusted $R^2 = 0.98$, $RMSE = 0.0031$, $df = 3, 28$). All devices showed significantly larger delays than the Eyelink baseline. Tukey-Kramer corrected post hoc t-tests revealed significant differences between all device pairs ($p < .001$). There was no significant difference between saccade directions.

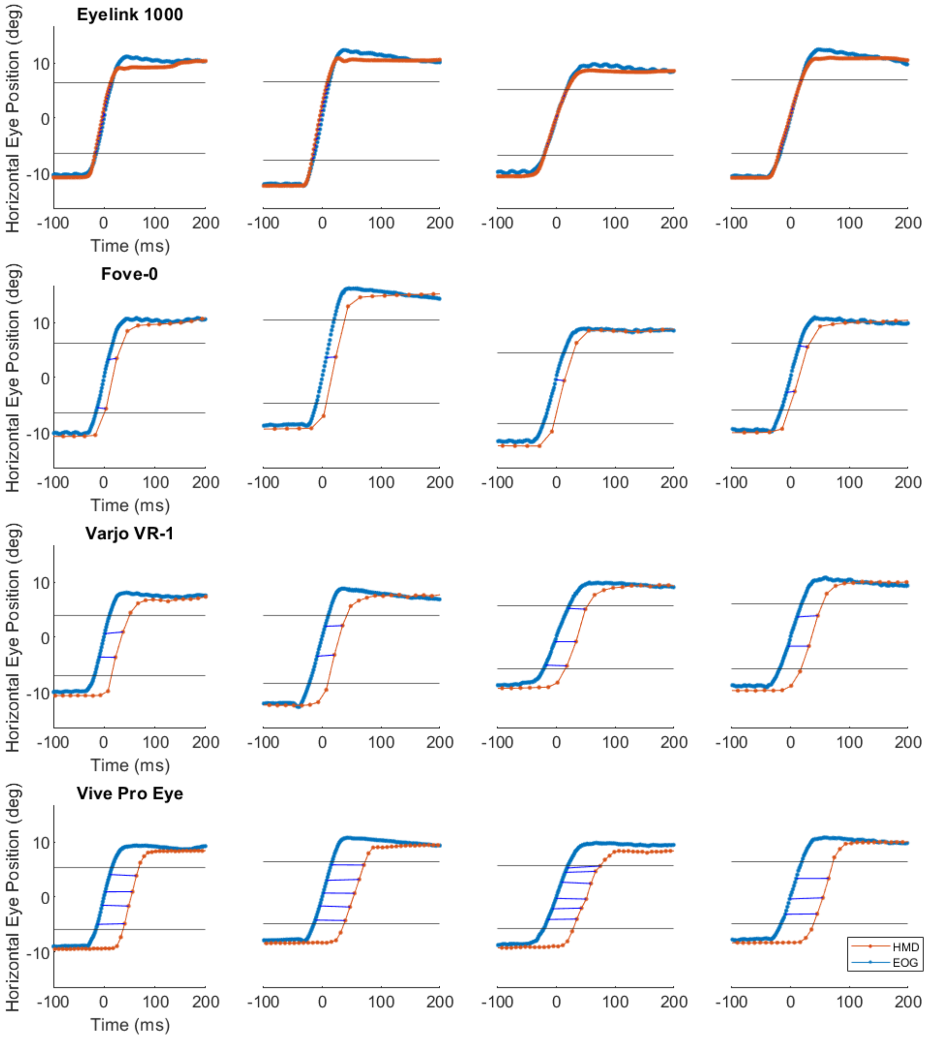


Figure I.2: Rightward mean saccades. Averaged saccades recorded with the HMDs (red) and the EOG (blue) of all four participants. Eye tracking delays are included as blue lines. Horizontal black lines show 20% to 80% of the mean amplitude.

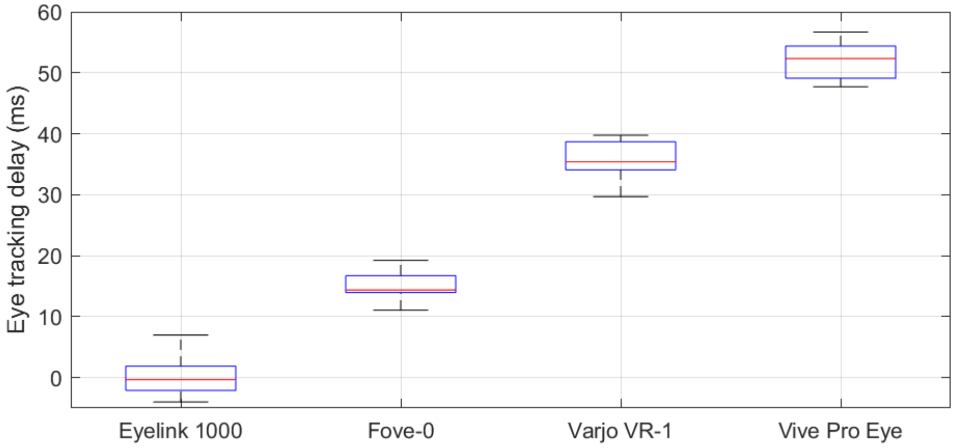


Figure I.3: Average eye tracking delays of all tested HMDs in comparison to the EOG. Measurements of the Eyelink 1000 are included as a baseline. The digitisation of the EOG signal leads to some delay. The Eyelink 1000 has an advertised latency of below 2 ms. Since EOG and Eyelink have a relative delay of 0 ms, the EOG has a similar low delay.

Study 2: Latency for Saccade-Contingent Display Update

If the HMD is used as a gaze-contingent display, the screen needs to be updated as soon as possible after the onset of a saccade. The latency of saccade-contingent update of the HMD depends on at least two further factors in addition to the tracker delay. The first is the speed of the detection of the saccade from the eye tracking data, which varies with the algorithm used, but may also depend on the sampling frequency. The second is the time it takes to update the screen, which depends on rendering time, frame rate and video hardware. Additionally, communication standards and bus architectures, as well as hard- and software used for transmitting eye tracker and video signals between the HMD and the computer, can add latency. We aimed to compare the end-to-end latency of saccade-contingent display update between the HMDs, using in each case the same saccade detection algorithm and the same video hardware, to measure the end-to-end latency in comparable minimum latency conditions. For this purpose, we used a simple Unity script that changed the background colour of the virtual environment in the HMD from white to black as soon as the eye position samples crossed the midline between the two targets (i.e., 0°). The black background was then present for the next 50 frames before it switched back to white for the next trial. The pre-rendering setting in Unity was turned off.

Methods

Setup

The setup was the same as described above, using the same HMDs. To record the time of the display change, a photodiode was fixed to the right display of each HMD (see example in Figure I.4). The photodiode was connected to an analogue input port of a RedLab-1208FS PLUS that was used to record the EOG signals. To prevent any sampling problems, we increased the maximum recording frequency of the input ports to 1500Hz.

Furthermore, we measured the Vive Pro Eye twice in this experiment, once with the SRanipal SDK provided by the device as before, and, second, with the Tobii XR SDK. We included this second measurement with a different SDK in the hope that the Tobii XR SDK might produce shorter delays than the native SDK.

Participants

We performed all measurements with three participants (2 male, 1 author, no contact lenses, no glasses, age 20–26, all experienced with eye tracking experiments). Informed, written consent was obtained from all participants.

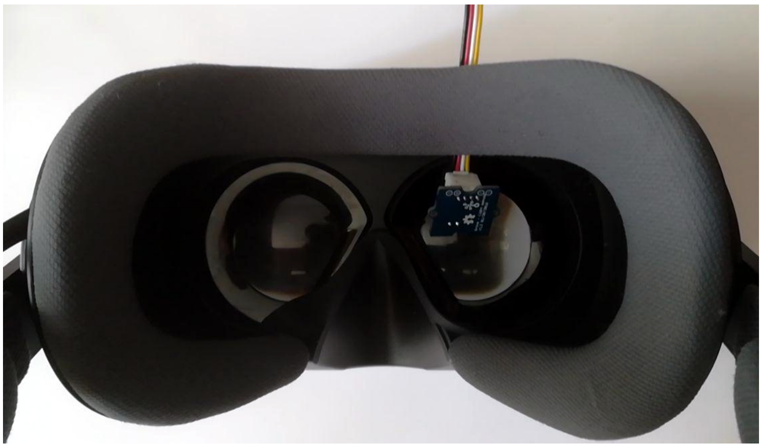


Figure I.4: Varjo VR-1 with photodiode placed on the right screen.

Procedures

The procedures were the same as in Study 1. In addition, when the eye tracker from the left eye returned to a gaze position that crossed the vertical midline of the FOV, the background colour of the VE in the HMD changed from white to black. To determine the time at which the change of the display from white to black occurred, we applied a change point detection algorithm (Killick et al., 2012) to the sampled photodiode signal. In case of multiple candidates, we used the last change point before the screen became and stayed black. Trials in which EOG saccades did not trigger a photodiode change in a time-window of (0,+200 ms) were excluded. All saccades were then manually checked to ascertain their validity. Overall, 692 of the 720 saccades (96.11%) were included in the analysis (see Table I.1 for an overview). From this data, we calculated mean saccades for each subject, device and direction, as before.

Results

Figure I.5 shows average end-to-end latencies for the tested devices. For each mean saccade, we calculated the mean end-to-end latency of saccade-contingent display update from the time difference between the middle of the saccade in the EOG signal and the time of display change determined from the photodiode signal. Among the four HMDs, the Fove-0 showed the lowest latency (45 ms), followed by the Varjo VR-1 (57 ms), the Vive Pro Eye with the native SDK (79 ms) and the Vive Pro Eye using the Tobii XR SDK (80 ms). Using the resulting 24 mean latencies for each subject, direction and device, an ANOVA showed a significant difference between devices ($F = 62$, $p < .001$, df = 3, 20). Moreover, Tukey-Kramer corrected post hoc t-tests revealed significant differences between the latencies of all devices except for the two conditions using the Vive

Pro Eye (Fove-0 compared to Varjo VR-1 $p < .01$, all others $p < .001$).

Since the setup and procedure was largely the same as in the prior measurement of tracker delay, we also evaluated tracker delay in this measurement series to confirm previous results in the separate delay measurements of test 1. Figure I.6 shows tracker delay in this measurement series. Using the resulting 24 mean saccades with delays for each subject, direction and device, an ANOVA revealed significant differences for devices ($F = 454$, $p < .001$, $df = 3, 20$), but not saccade direction. Moreover, Tukey-Kramer corrected post hoc t-tests revealed significant differences between the delays of all device conditions ($p < .001$), except for the two conditions using the Vive Pro Eye, which were not significantly different from each other. Display latency was determined by subtracting end-to-end latency from eye tracking delay (see Table I.1).

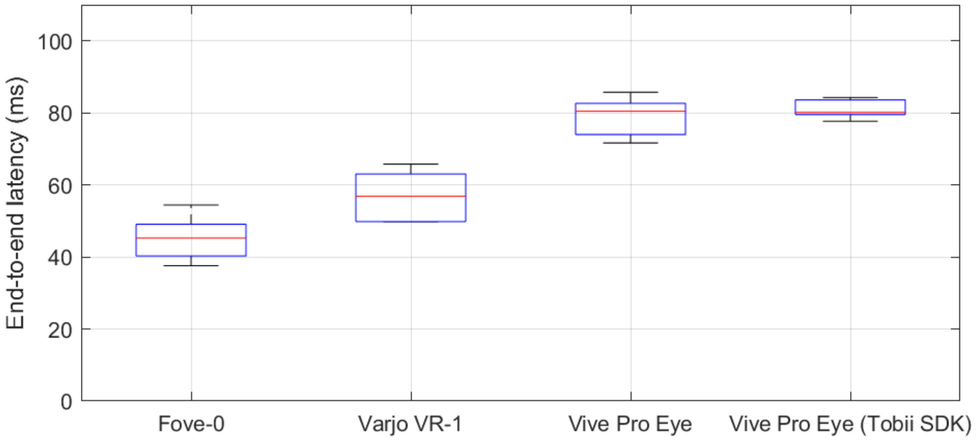


Figure I.5: Average saccade-contingent latency of all tested HMD eye trackers in comparison to the EOG. End-to-end latencies of all tested HMD eye trackers based on the EOG signal and a photodiode capturing changes on the HMD display.

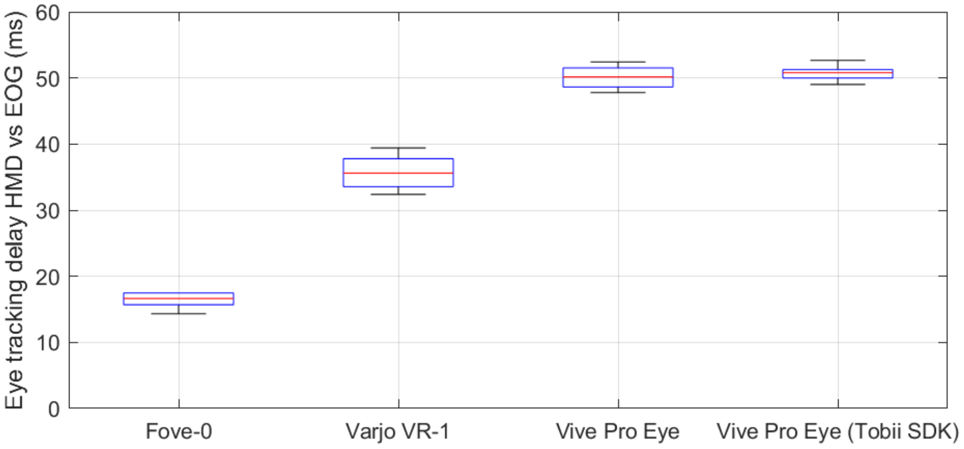


Figure I.6: Average eye tracking delays of all tested HMD eye trackers in comparison to the EOG. The eye tracking delays of the second test confirm previous results. The Tobii XR SDK did not make a difference for the delays of the Vive Pro Eye.

Discussion

We measured tracker delays and end-to-end latencies of three HMD eye trackers in comparison to EOG. The Fove-0 had the lowest tracker delay in both tests (15 and 16 ms) and an end-to-end latency of 45 ms. The Varjo VR-1 had an eye tracking delay of 36 and 35 ms and an end-to-end latency of 57 ms. The Vive Pro Eye, either with its native SDK or with the Tobii SDK, was considerably slower with tracker delay of 50 ms and an end-to-end latency of 80 ms. The values of the Vive are out of the range suitable for gaze-contingent experiments (Loschky & McConkie, 2000). The latencies achieved by the Fove-0 or the Varjo VR-1 appear more suitable to such tasks (Loschky & Wolverton, 2007).

We do not know what causes the differences in tracker delay between the devices. One possibility, other than hardware differences, might be extensive low-pass filtering of the eye position data in the SDK. Such filtering reduces spatial noise (which might allow more accurate gaze position estimation) at the cost of latency. It may thus be that different manufactures put different emphasis on spatial vs. temporal accuracy for different use-cases. Indeed, we noticed informally that a recent software update (version 0.17) of the Fove-0 SDK, which was advertised to improve spatial accuracy, produced longer delays than the ones we measured here. Any trade-off between spatial and temporal accuracy would be different for different research applications. Because using an HMD as gaze-contingent display has different requirements regarding spatial filtering and latency than using it for the analysis of fixation distribution, for researchers and developers it is highly desirable to get full access to filtering and raw data from each eye tracking sample with minimal delay.

Our measurements of delay and latency were taken in comparison to EOG. The EOG operates in principle without any delay since it measures changes in electrical voltage directly produced by the change in orientation of the eyeball. However, the digitisation of the EOG signal also adds a small delay to the EOG data. For accurate delay estimates of the HMDs, the delay introduced by the EOG needs to be taken into account. We cannot truly measure this delay, but based on the sampling rates we used it should be on the order of one or two milliseconds. This is consistent with the delay measured between the EOG and the Eyelink, our reference system, which came to a mean of 0 ms. As the Eyelink is expected to have a delay of 1.8 ms (SR Research Ltd., 2009) the delay of the EOG sampling should be in the same range. Given the size of the measured effects and the assumption that the delay of the EOG should be the same for all HMDs, this does not affect our results.

In placing the photodiode for the measurement of end-to-end latency, we aimed for similar layouts of the photodiode in each subject and device. However, there is some variance resulting from different head shapes of the participants who took part in the study. Moreover, the pixel matrix of different HMDs might be refreshed with timing differences between the first and the last pixel. However, this could create at maximum a difference between two systems of one inter frame interval (for example 11ms at 90 Hz) and thus cannot explain the differences we found between devices.

All three of our tested HMDs show rather large end-to-end latencies when compared to eye tracking devices currently used in research (Gibaldi et al., 2017). However, applications using off-line analysis of fixations in VR can be done with all devices. In this context, considerations regarding spatial resolution and data quality are more important than online latency. Off-line analysis of fixations, even if time-critical, seems to be possible with all HMDs since the delays and latencies had small variance and thus were quite stable. However, any analyses of saccadic reaction times or times of first fixation, etc. will have a constant bias due to the eye tracking delay. Researchers interested in such measures should therefore measure the delay in their setup and correct for it in their analysis.

Based on our results, it appears that the Fove-0 and the Varjo VR-1 can be used as gaze-contingent display to a certain extent. However, it is important to note that for our measurement of end-to-end latency for gaze-contingent display change, we used a rather large (20°) saccade and a very simple detection algorithm (crossing of the screen centre). These choices were made to allow a robust and comparable measurement of delay. Gaze-contingent research and applications are likely to involve different saccade amplitudes and more complex detection algorithms. Saccade amplitudes are highly dependent on task, for example, when looking at a natural environment the typical saccade varies between 2° and 20° (Bahill, Adler, & Stark, 1975; Gibaldi & Banks, 2019; Land et al., 1999; McConkie et al., 1984; Sprague et al., 2015). Smaller saccades are more difficult to detect than larger saccades. Moreover, smaller saccades are also shorter in duration and might last for only 30 ms or much less. This might reduce the applicability of the tested HMDs for gaze-contingent research with small saccades. Therefore, researchers should do their own tests for their specific setup and scenario to examine whether the HMD eye tracker is suitable for their needs.

A further issue for online saccade detection concerns the sampling rate of the eye tracker. Reliable onset detection can be hard using sampling rates of the tested systems ($70-90$Hz). The eye trackers in our test HMDs are capable of higher sampling rates than frame rates, which might be helpful for implementing reliable saccade detection algorithms. Moreover, more advanced methods might allow achieving lower latency for saccade onset detection (Arabadzhiyska et al., 2017). Still, as eye tracking in HMDs may become more advanced in the near future, access to faster eye tracking data with lower delay and lower end-to-end latencies would constitute an important step for VR and perception research.

Acknowledgements

This work was supported by the German Research Foundation (DFG La 952-4-3 and La 952-8) and has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 734227.

Virtual Reality Eye Tracking Data

Using the method presented in study I, various HMDs can be evaluated with regard to their ability to display visual stimuli at a specific position in the FOV independently of the current viewing direction. To obtain a good assessment of the overall eye tracking quality, other metrics have to be considered as well. In addition to sensor latency and delay, these include precision and accuracy of the estimated eye direction (Aziz et al., 2024; Lohr et al., 2019; Sipatchin et al., 2021). To measure precision, the root-mean-square of the inter-sample angular distances between successive gaze directions during a fixation is usually calculated (Blignaut & Beelders, 2012). Eye tracking accuracy is often defined as the average error between the position of a fixation target and the measured fixation, while precision describes the dispersion of gaze data while the gaze is fixed on a static position in space (Clemotte et al., 2014; Nyström et al., 2025). Under optimal conditions, the HTC Vive Pro Eye shows eye tracking data delays of about 50 ms, an accuracy below 1° in central, but up to 10° in strong peripheral positions (27°), while eye tracking precision varied between 1.4 and 3.6° (Sipatchin et al., 2021; Stein et al., 2021).

Low eye tracking precision can sometimes be corrected at the expense of additional latency, by applying a time-wise filter on the eye tracking signal (Holmqvist et al., 2011). However, to detect rapid eye movements such as saccades as early as possible, it is best to collect eye tracking data at a high temporal sampling rate and with the lowest possible latency.

Accuracy depends on the calibration of the eye tracker. During this process, the eye tracking system creates a mapping of the distance between the tracked pupil and the visible infrared reflections of the cornea, which were extracted from the recorded image of the eyes and the angle of the distance vector to positions on the screen. (Adhanom et al., 2023; Harezlak et al., 2014). In most cases, the user is shown a series of (usually 5-16) small targets that must be looked at in sequence (Adhanom et al., 2023). To ensure consistent accuracy of the calibrated assignment, it is also important to check whether the worn eye tracker changes position during the experiment (Niehorster, Santini, et al., 2020). If there is too much movement, recalibration must be performed. Although they are not (yet) widely used, there are also some VR algorithms for calibrating eye tracking that can be run in the background without the user having to perform a separate calibration process. For example, gaze movements can be compared with the user's interactions with the user interface in order to continuously update the mapping (Hou et al., 2025). Alternatively, it is also possible to use an algorithm to find connections between eye movements and visible motion on the screen (Tripathi & Guenter, 2017).

Determining Gaze Position in VR

After calibration, current VR eye trackers usually provide two three-dimensional eye direction vectors. These can easily be converted into vertical and horizontal angles, by comparing the x and y-coordinates with the head direction. The gaze direction, which is calculated by combining the eye and head orientation vectors, can then be assigned to virtual objects on individual images of the VE.

Although the vectors of eye direction are often represented as three-dimensional vectors, they usually do not contain any information about the gaze distance. Instead, the third coordinate is typically the result of normalizing the vectors to a length of 1. To estimate the viewing distance by exclusively using eye tracking data, the vergence angle between the left and right gaze direction can be calculated and calibrated to different gaze depths. However, we tend to make large vergence angle errors, especially when looking at near targets (Cornell et al., 2003). The same errors can be observed in VR and can not be reduced even when using optical lenses specifically implemented to reduce the VAC (McAnally et al., 2024). An alternative method of obtaining a rough estimate of the user's viewing distance is to measure the interpupillary distance resulting from the estimated eye positions (Arefin et al., 2022).

Another way to determine the three-dimensional viewing position is to combine the eye movement data with the structure of the VE. Once the head and eye signals have been combined, a virtual ray can be created from the eye position in the direction of gaze. If it crosses a surface of one of the virtual objects in the VE, the collision point can be used as the current three-dimensional gaze position, with the virtual object serving as a label. Since these objects update their position frame by frame, the rendering frequency of the VE influences the recorded three-dimensional gaze positions. Without further adjustments, the rendering frequency is limited by the maximum refresh rate of the display and how well the available hardware copes with the complexity of the scene. Thus, the quality of the resulting gaze position depends on the quality of the eye tracking and the quality of the head and position tracking. By default, the gaze ray method returns the first object the ray collides with. Since the accuracy and precision are not perfect in current VR eye trackers, it is possible that a gaze position from an object behind the current object of interest is selected, especially if the gaze target is small. This limitation also applies to desktop eye tracking configurations. In VR however, a small measurement error in vertical and horizontal eye direction angle can lead to comparably large errors in three-dimensional gaze points. To some extent, this problem can be detected by examining the distance between the user and the fixation position over time. Unusual spikes while no saccades are present in the eye movement signal may indicate that a gaze signal is bouncing between two objects in the line of sight.

Eye Movement Classification

In the next step, each of the continuous data points from the stream of gaze data is classified as a type of eye movement. In experiments, in which visual stimuli are presented on a fixed screen on a desktop, fixations and saccades are usually classified depending on threshold algorithms that are based on the eye movement velocity (Holmqvist et al., 2011) and therefore can in theory also be applied when using mobile eye tracking in natural environments. However, most of the commonly used algorithms are based on velocity thresholds and are optimised to work in static environments in scenarios without head movements. In experiments with head movements, differentiating clearly between smooth pursuit, saccades and eye movements that stabilise an object on the retina during an ongoing head movement can be challenging since these movements interact and their conceptual separation becomes less clear (Drewes et al., 2021; Lappi, 2015; Steinman & Collewijn, 1980). In VR, data from VE, such as fixation marks based on the positions of virtual objects can also be taken into account. For example, one criterion for classifying a series of eye movements as fixation or smooth pursuit could be that they are continuously directed at the same virtual object. In addition, the positions and rotations of the user's head, eyes and the virtual object being looked at could be used to distinguish smooth pursuit to track a moving object from eye movements to stabilise a fixed object during a head movement. This means that contextual information is used to classify the gaze data. Thus, additional efforts should be made to show that the gaze patterns identified using this method are transferable to different types of VEs and are not specific to eye movements in VR.

Manual classification of fixations has proven to be an unreliable standard since untrained coders often disagree with each other (Hooge et al., 2018). Nevertheless, the manual inspection of eye tracking data is a useful step and can be especially helpful in finding and improving incorrect classifications when using algorithms.

After classification, we can observe, replay and analyse how VR users are looking around in the presented virtual worlds. This setup is particularly interesting for research purposes since it offers a simple way to measure eye, head and thus also gaze movements while strictly controlling the (visual) inputs that participants receive. In comparison to real-world eye tracking experiments, gaze fixation points can be automatically linked to the inspected virtual objects. The resulting data therefore contains a detailed stream of when certain virtual objects were in the user's FOV and which of these objects the user viewed while solving various VR tasks. VR experiments thus make it possible to collect a large amount of labelled data that describes how participants explore their environment using the same movements as in the real world. As a result, VR research can help us better understand complex human behaviours and could be a helpful addition to field experiments where both the control of visual stimuli and the labelling of eye movement data involve significantly more effort.

Where Do We Look While We Act?

Whenever we enter a new environment, we intuitively begin to explore using eye movements (Hayhoe et al., 2003; Land et al., 1999; Rothkopf et al., 2016). But how does our visual system know where to look next? In many everyday situations, we use our eyes to facilitate the action we are performing, such as when watching birds pick cherries, exploring a particular object in a virtual world or contemplating at a work of art in a museum. Seemingly effortless, our gaze finds its way from one object to the next, allowing us to perceive and interact with our surroundings. We can move our eyes in a certain way based on tasks and strategies we have decided on even before we have seen the environment in which we are acting by incorporating previously accumulated knowledge and our current expectations.

This chapter provides an overview of where we look when performing various actions and explains how we plan and monitor our actions with the help of precisely timed eye movements. In particular, the chapter covers where we look when walking, where we look when overcoming obstacles, how we orient ourselves toward a destination when walking, and how walking influences our perception.

Task-dependent eye and head movements have been systematically observed and analysed in a range of everyday activities (Adhanom et al., 2020; Land & Tatler, 2009a; Schütz et al., 2011). In many activities, such as making tea or reading, we look at places and objects that are related to our actions (Land et al., 1999). When reading, the amplitude of the saccades and the duration of the fixation adapt to the structure of the words and sentences we are currently looking at (Schütz et al., 2011). When we are looking for something, we tend to make slightly longer saccades than when we are just glancing at a picture (Tatler et al., 2006). Through examining eye movements during different actions, it becomes clear that the frequency of different types of eye movements change and our fixations and actions are temporally perfectly coordinated (Tatler et al., 2011). This can for example be observed when reading sheet music (Furneaux & Land, 1999), driving (Land & Lee, 1994) or motor racing (Land & Tatler, 2001).

In addition to saccade amplitude and fixation position, an ongoing task can also influence other types of eye movements (Doughty, 2001; Karson et al., 1981): For example, the blink rate is lowest when reading (7–10 blinks/minute), but increases slightly when we are in a relaxed state in a silent room (12–16 blinks/minute). This rate is also known as the spontaneous eye-blink rate. During conversation or when memorizing while listening, the blink rate increases to more than 24 blinks/minute. During repetitive tasks fixational eye movements such as drifts increase (Di Stasi et al., 2013; Friedman & Komogortsev, 2025).

Interestingly, there are also a number of examples that show that the top-down appeal and engagement of visual stimuli can have an effect on our gaze. For example, the blink rate can decrease, when watching repetitive actions or when a person leaves a scene (Andreu-Sánchez et al., 2021). When we look at less obvious characteristics of a film clip, such as the length of the shots, we can see that shorter shots slightly reduce our blink rate (Andreu-Sánchez et al., 2017). Thus, the blink rate can be used to measure changes in engagement in periods of seconds to a minute during free viewing (Ranti et al., 2020). Similarly, our saccade amplitude, fixation duration and eye movement variability change when we watch different types of outdoor scene videos with varying styles of cutting or effects such as stop-motion (Dorr et al., 2010).

Task Dependent Gaze Patterns

To understand where we look while we act, we need to understand how different tasks influence our eye movements. We also need to understand that the currently available visual input is also important since it defines most of our fixation targets. This interaction between top-down and bottom-up factors influencing our gaze was first systematically described in 1967 when Yarbus showed that the patterns of eye movements are similar when the same painting is viewed by different people, even if there is a gap of one or two days between viewings. This underlined the important influence that the visual input has on our eye movements. However, although the patterns were similar, they were not identical, with the eye movement patterns within a person being more similar than those between different viewers. Yarbus then asked one person to look at the same painting seven times while giving various viewing instructions, such as making a judgement about the depicted scene, remembering aspects of the image or simply looking at the painting freely, whereupon the eye movements changed significantly (Duchowski, 2007; Tatler et al., 2010). This showed that the instruction influenced eye movements substantially.

Later Noton and Stark (1971a) expanded these results by showing that subjects without an explicit task also fixate similar regions of interest when confronted with the identical visual stimuli. Just like Yarbus, they observed small differences between viewers and noted that the order of fixations within the regions that all participants looked at were not stable. For example, when inspecting a square, participants usually fixated on the corners of it. The order of the corners, however, varied from one observer to the other and even between successive observations by the same person (Duchowski, 2007; Noton & Stark, 1971a, 1971b).

These results show that our eye movements vary dependent on the ongoing actions. During an action, our saccade and fixation patterns seem to follow some common underlying principles that help us solve the ongoing tasks while still perceiving what is around us. In the phase before a task begins or an instruction is given, we seem to mainly use our gaze to explore our surroundings. Thus, the distribution of fixations between task-relevant and irrelevant objects is roughly equal. Shortly before and during an interaction, we mostly fixate on locations relevant to the task (Hayhoe et al., 2003; Rothkopf et al., 2016).

Task related eye movements can be classified into different types. Land et al. (1999) proposed a categorisation based on monitoring functions. They distinguished between eye movements that serve to locate an object that will be used later in the process, eye movements that monitor how we move our hand or an object in our hand to a new location, eye movements related to the approach of one object to another, and eye movements related to checking a condition or variable related to the ongoing action.

While this classification is well suited to tasks that involve interactions with one or more objects, Foulsham (2015) developed a classification of fixations that is applicable to an even broader range of situations: He divides fixations during actions into three categories depending on how they relate to the task at hand temporally.

Planning

The first category consists of fixations that are used to plan actions. These anticipatory fixations are part of task-dependent strategies and can be observed relatively early on, as they are often unrelated to the action currently being performed. In a study in which participants had to solve a modelling task, look-ahead fixations occurred before 20% of all grasping movements and took place up to 3 seconds before the actual movement (Mennie et al., 2007). Similar fixations can also be observed when making binary decisions, such as deciding which path to follow at a junction. Usually, these precede the actual decision only by a few hundred milliseconds (Wiener et al., 2011).

In addition, there are less common anticipatory fixations, which are not related to planning the next action, but can be used to memorise the position of objects for later use (Foulsham, 2015). In an experiment in which eye movements were compared between different tasks in the same environment, Pelz and Canosa (2001) found that anticipatory fixations were typically on task relevant objects.

Just-in-Time Fixations

The second category mostly includes fixations that occur when we interact with objects in our environment. These frequently occurring fixations clearly relate to the upcoming action and are called just-in-time fixations (Ballard et al., 1992; Foulsham, 2015; Hayhoe et al., 1998). In an experiment by Land et al. (1999), gaze moved to the next object an average of 0.6 seconds before the participants had finished manipulating the previous object. Foulsham observed a similar time sequence when grasping. He reported that we fixate on objects approximately 0.5 to 1 s before reaching out for them.

Brouwer et al. (2009) described the time course of eye movements during such a grasping task in more detail: The first eye fixation is normally directed towards the centre of gravity of the target object. When we simply look at an object, our gaze normally rests close to this position. Sometimes, a second saccade slightly corrects a possible undershoot of the initial landing position. Depending on the next intended action, the point of fixation may also be at a different location. For example, if we grasp an object, our second fixation is focused on regions that contain relevant information for the grasping positions of our index finger and the thumb. Thus, by looking ahead and fixating task-relevant landmarks during interaction, our eyes support our hand movements (Johansson et al., 2001). However, when we try to walk past an obstacle, we look at its outer edges rather than its centre (Rothkopf et al., 2016). This means that the additional visual information is gained just in time to carry out the next step (Ballard et al., 1995).

Thus, when putting a cup somewhere, we fixate empty areas, such as the place on the table where we want to put the cup down (Hayhoe, 2000; Land et al., 1999; Schütz et al., 2011). Interestingly, this pattern of fixating empty space between two potential future targets has also been observed in some visual search tasks (Findlay, 1997; Najemnik & Geisler, 2008). This suggests that during a search, some just-in-time preparation is also beneficial.

In addition to the fixation position, the timing of these task-related eye movements also appears to adjust depending on the task. For example, saccades towards anticipatory fixations during interaction seem to be faster and shorter than saccades used to examine visual objects (Epelboim et al., 1995, 1997). Such shortening of fixation time and acceleration of saccades can also be observed in other tasks, for example when an eye movement is performed as part of a pre-planned sequence of fixations in the same direction (Carpenter, 2001). Moreover, on the basis of an anlaysis of eye movement strategies of cricket players, Land and McLeod (2000) argue that the correct temporal placement of the eyes is more crucial for the successful execution of behaviours than fixating accurately and that skilled performance depends as much on the correct temporal as spatial allocation of gaze. Furthermore, Hacques et al. (2022) showed that the complexity of climbers' gaze paths reduced when they trained a particular route for several weeks. In addition, their ratio of observational and anticipatory eye movements adjusted depending on the training.

Another function of short-term look ahead fixations is likely to relieve our visual working memory: Ballard et al. (1992) conducted an experiment, in which participants were instructed to copy an arrangement of coloured blocks on a computer. Instead of storing and recalling the original structure in visual memory, they were constantly looking back and forth between the original figure and their copy (Ballard et al., 1992, 1995). Thus, by continuously using eye movements as a method for information seeking, we seem to reduce the need of using a large capacity in our visual memory.

Monitoring

Finally, fixations can be related to our ongoing behaviour. This can be helpful to quickly adjust our actions based on visual information if necessary. When we cut bread, for example, we first fix the point of contact with the knife and then move our gaze along the cut directly in front of the knife (Hayhoe et al., 2003). Similarly, we can adjust our pouring speed while filling a cup if we keep a close eye on the cup during this action (Land et al., 1999).

Our eye and head movements during an action are closely linked to our behaviour and are often necessary to complete a task successfully. Most of our eye movements are associated with monitoring, anticipating and planning our actions. This means that leading up to, and during a task interacting with our environment, we are constantly gathering useful and necessary visual information with our eyes and thus adapting our actions to the requirements of different tasks. This also helps us to perform more precise hand movements. However, it also means that task-related gaze patterns do not always occur in fixation blocks that only relate to one subtask. Instead, a time series of eye movements often appears chaotic and complex at first glance, as successive fixations can relate to the current, the next and then again to the current subtask. Therefore, analysing eye movements in relation to associated actions can be a helpful method for untangling the chaos.

Interestingly, describing gaze data according to this principle can shed light on how far ahead we typically plan and what information we use in different types of tasks. This means that task-related eye movements can give us a better understanding of how we act and could ultimately allow us to make predictions about future behaviour given a predefined task in a known environment.

Gaze During Locomotion

Natural locomotion allows us to constantly shift our FOV to perceive relevant visual information around us. In many situations where we want to interact with an object in our environment, we walk towards it. This is because walking can be easily integrated into many of our daily tasks, such as when we go to the kitchen, find our favourite cup, fill it with water and then go to the living room to sit down and drink. Of course, we also move our head and eyes during these subtasks. Gaze tracking recordings during such natural tasks may appear chaotic at first glance. However, many of the eye and head movements follow systematic patterns and can be attributed either to interactions with objects or to walking planning. To better understand and recognise these patterns, various typical gaze behaviours when walking are described in the following section. The next few pages also discuss how walking itself can influence our perception.

Looking Ahead

Like many other actions, walking benefits from visual information that we collect on the fly. However, the walking movements themselves make it difficult for walkers to capture this information straight away. During walking, vertical and horizontal movements of the head occur with approximately 2 Hz vertical and 1 Hz horizontal head movements (Imai et al., 2001; Moore et al., 1999; Moore et al., 2001). Thus, eye movements are compensating for the head motion to stabilise our gaze (Moore et al., 2001). In addition, our eyes compensate for our progress as we walk, allowing us to keep a fixed object more stable on the retina even as we approach (Patla & Vickers, 2003). At the same time, we align our head to a point approximately 0.8 m in front of us (Hirasaki et al., 1999; Moore et al., 1999). In walking experiments on flat terrain we usually direct slightly more than half of our fixations toward the ground (Matthis et al., 2018; Patla & Vickers, 2003; Pelz & Rothkopf, 2007).

The targets of these travel-gaze fixations are usually in locations that we reach between 0.8 and 1.5 s later, two to four steps ahead (Hollands et al., 1995; Matthis & Fajen, 2014; Matthis et al., 2018; Patla & Vickers, 2003). Most of these fixations are shorter than 0.6 s (Patla & Vickers, 2003) and tend to be mainly directed at locations that are relevant to the ongoing task (Hollands et al., 2002; D. Marigold & Patla, 2007; Rothkopf et al., 2016). In general, these fixations appear to serve to inspect specific locations in the terrain, while saccades direct the gaze back towards future waypoints further ahead ('t Hart & Einhäuser, 2012).

Brenner, Ghiani, et al. (2024) found that we apparently look less far ahead when running than when walking. They noticed that during running, we do not adjust our gaze distance to the running speed. In other words, regardless of speed and step length, we look a certain distance ahead and reach the place we are looking at earlier and in fewer steps when we are moving faster while taking larger steps. This also means, we usually do not look at the ground to place our feet. Instead, we seem to rely on the passive mechanical response of our body to remain stable (Dhawale & Venkadesan, 2023). However, when running with others, we intuitively adapt our gaze behaviour. Brenner, Janssen, et al. (2024) found that when running in groups, the average time spent looking at the path in front increases by about 10 %, presumably to make sure we do not trip over other runners' feet. At the same time, we are also able to adapt our gait to the demands of vision to a certain extent. For example, Mulavara and Bloomberg (2003) found that participants in a walking and reading task were able to extend the double support phase of the gait cycle by 10% without changing step length or step duration to stabilise their upper body and thus their gaze. The more stable a target is in relation to the surroundings along the axes perpendicular to the direction of movement, the faster and more precise the gaze fixation (Manakhov et al., 2024). In contrast, visual objects that are attached to the head are perceived less accurately during walking (Borg et al., 2015; Genç et al., 2016).

Overcoming Obstacles & Rough Terrain

If no obstacle is in our way, we mostly look at the path ahead of us and occasionally at our walking target (Patla & Vickers, 2003). In contrast to reaching movements, we use fewer fixations to monitor how exactly we place our feet. Thus, in a study using mobile eye tracking, Patla and Vickers (1997) found that when we find an obstacle in our way, we tend to look at it before we reach it, but not while we step over. During object avoidance preparation, we direct our gaze mainly to the outer edges of the obstacle (Rothkopf et al., 2016). If we are confronted with stairs, our gaze patterns also adapt. For example, the number of saccades when walking down stairs compared to walking on a descending surface is clearly increased ('t Hart & Einhäuser, 2012). Moreover, our gaze distance seems to increase slightly when a walking task becomes more difficult and risky. For example, we look further ahead when descending stairs than when ascending them (Ghiani et al., 2023). During stair walking, we tend to fixate targets about three steps ahead (Zietz & Hollands, 2009) and our gaze often skips the first step. In addition, D. Marigold and Patla (2007) noted that, when we change from one surface to another, a particularly large number of fixations are directed towards this transition. They concluded that in this way, fixations maximise the amount of available information to enable safe foot placement.

Interestingly, there is some evidence to suggest that peripheral vision rather than foveal vision, is important for passing obstacles. This fits with previous observations in athletes: Without central vision, slalom skiers were easily able to complete a 150 m course, javelin throwers hurled their javelins far and figure skaters performed clean spiral patterns on the ice (Craybiel et al., 1955; Franchak & Adolph, 2010). Even patients with central vision loss navigate through their environment without major impairment

(Hassan et al., 2002). Without peripheral vision, however, patients have great difficulty finding their way in the world and often stumble over obstacles (Geruschat et al., 1998). Skiers without peripheral vision go off course, javelin throws become shorter and figure skating patterns become unpredictable (Craybiel et al., 1955; Franchak & Adolph, 2010). The results of a field test with mobile eye trackers, in which participants ran through an obstacle course while searching for stickers, point in a similar direction. In 41% of the obstacle encounters of children and 68% of those of adults, participants controlled their locomotion adaptively without ever fixating the obstacle (Franchak & Adolph, 2010).

On uneven terrain, where different surfaces alternate and we have to overcome gaps and small differences in height, we also adjust our walking. For example, we increase the stride variability and the height of the swing foot (Kowalsky et al., 2021). Complex terrain seems to influence our sensorimotor decision-making and path planning based on depth information since we consistently tend to choose indirect routes to reach flatter paths (Muller et al., 2024). We typically reduce our speed and simultaneously lower our gaze so that our fixations are closer to us (Matthis et al., 2018; 't Hart & Einhäuser, 2012; Thomas et al., 2020a). These shorter fixations could also serve to gather more information about where we set foot. When comparing gaze patterns across different terrains, Matthis et al. (2018) found that gaze was almost exclusively focused on the upcoming path and was closely associated with the upcoming footfalls, in medium and rough terrain. In rougher terrain especially, gaze was more evenly distributed between the upcoming footprints 2 and 3.

This finding fits well with a previous study in which Matthis and Fajen (2014) found that if participants could not see what is coming in 2.5 steps, walking speed was reduced and the likelihood of colliding with some of the objects increased. They explain this finding based on the different phases of walking. If we would only plan ahead one step, visual information about an obstacle is perceived while we are already in the one-legged phase of the opposite leg. In this situation the foot position and the alignment of the centre of mass are already partially predetermined since the toes of the previous step have been pushed away from the ground already. Thus, at this point, only less efficient adjustments in the flight phase of the step are possible to adjust the foot landing position. However, if visual information about the location of obstacles is available before the start of the single support phase (at least two full stride lengths before the obstacle), walkers can adapt the initial velocity of their centre of mass and the location of the touchdown foot to available footfalls by applying an appropriate push-off force. Similarly Patla (1998) also found that it is sufficient to have the necessary information available only in a critical time window before overcoming an obstacle, and found that walking movements are not compromised if visual information about an obstacle is withheld during the overcoming of the obstacle or up to two steps before overcoming it. In a study, in which the participants had to step on visual targets that could be hidden or visible, Matthis et al. (2017) found that the latest possible time at which the participants needed the visual information was 1.5 steps before reaching the target. So if visual information was available just in time, participants were able to walk over the visual markers fairly accurately. Interestingly, they also found that visual information that was only available further in advance, but was not visible 1.5 steps before reaching the target, led to less accurate foot placements. This suggests that we process visual information for walking, similar to when grasping an object, just before the actual movement. When comparing the accuracy of foot placement with one or both eyes, Bonnen et al. (2021) concluded that, although it was possible to place the feet fairly accurately in both conditions, we probably use depth information from both eyes for our walking plans. Furthermore, under certain conditions of their experiment, artificially restricting vision to one eye led to the slightly more frequent fixations on nearby footholds.

Interestingly, despite the different viewing behaviours in the different environments and despite the varying difficulty of the terrain, participants maintained a constant look-ahead time of approximately 1.5 seconds under all terrain conditions and appeared to use their eye movements in such a way that they knew what would happen in the next 1.5 to 2 seconds (Matthis et al., 2018). This could explain why we make specific, anticipatory walking speed adjustments on difficult terrain: The slower speeds may represent the maximum speed at which we are able to process the information needed to support locomotion in the face of higher uncertainty in complex terrains. This would also fit to the results of a walking study by Darici and Kuo (2023). Here, variations in walking speed could be reproduced over several repetitions of the same trail and began about 6 to 8 steps before reaching a terrain feature, while closer features were weighted higher. Usually these speed adjustments are accompanied by a downward movement of the gaze, which consists of both head and eye movements (Thomas et al., 2020b).

Gaze During Turning

If we want to change direction while already moving, we can also continuously align ourselves with the new target by walking in a curve. When we go around curves, our head and eyes normally point to a location about 1 s ahead of us in the direction of the apex of the curve (Grasso et al., 1998). Initially, our eyes move in saccades in the direction of the rotation, interrupted by slower compensatory movements that compensate for the simultaneous head movement (Imai et al., 2001). The gaze thus guides the rotation of the head so that the eyes are initially directed further in the direction of the rotation than the head. Once the rotation is complete, the eye position relative to the head returns to zero (Imai et al., 2001). Interestingly, eye movements towards the apex do not appear to be based on our visual input, as they occur in both light and darkness (Grasso et al., 1998). Therefore, Grasso et al. (1998) argue that instead these type of eye movements are a necessary part of our behavioural repertoire to prepare a stable frame of reference.

When we decide to start walking in a certain direction, we usually orientate ourselves towards the target first, rather than approaching it in a curve. Such turns also occur, when we suddenly change our walking target. Hollands et al. (2002) found that as soon as participants do such turns, they make a synchronised eye and head movement towards the new target. They argued that the close temporal relationship between the onset of eye, head and turning movements suggests that they are all generated as part of a single reorientation process. This idea is also supported by another experiment in which the head was immobilised during guided turning, resulting in a visible change in the timing of the body's realignment (Hollands et al., 2001). Interestingly, participants typically spent less than a third of their time looking at aspects of their future route before turning. However, while turning, they focused their gaze strictly on the updated destination of their walking route until the slower head movement, compared to saccades, had reached the new walking direction. Thus, Hollands et al. (2002) concluded that the visual information about a new target is used to orient the body towards the new direction by creating a gaze-centred frame of reference by first completing the saccade, then a head-centred frame of reference after completing the head movement and finally realigning the body (Hollands et al., 2001).

A similar study by Hollands et al. (2004) observed the orientation behaviour of standing participants when they were asked to align themselves with a visual target and then walk towards it. In this experiment, it took between 0.3 and 0.4 s to make the first eye movement after the target was visible. After about 0.6 s, the head and upper body followed and after 1.2 s, the feet also moved. This order was independent of how far the participants had to turn and also remained the same for turns to targets that were not initially visible in the FOV. The way in which the foot moved to align with the target was similar to the eye movement that was directed at the target about 1 s earlier. Interestingly, larger turns led to a slightly increased delay in the onset of the initial eye movement of about 0.1 s. Moreover, Hollands et al. (2004) found a positive correlation between the latencies of eye and foot movements. As an explanation Hollands et al. (2004) suggest that the long latency of the saccades indicates the presence of underlying coordination networks. This would mean that the central nervous system delays the onset of the eye movement until it is ready to initiate a coordinated whole-body movement, or when an eye movement is part of a coordinated whole-body movement, saccade programming incorporates additional information about other body segments, delaying the onset of the movement.

Walking Alters Eye Movements & Perception

In addition to the eye movements that help us orient ourselves, there is also some experimental evidence suggesting that walking itself influences some of our gaze movements. For example, Cao et al. (2020) found that blinking and saccades during walking occur preferentially during the phase in which both feet are on the ground (double support phase). They found that increasing the walking speed went along with an increased blink rate. Interestingly, blink rate increased regardless of whether the task was performed in the dark or in the light, suggesting that the change in blink rate was related to the walking rhythm rather than the visual input. This was not the case for saccade frequency, which only increased in light, which fits well with the idea that saccades during walking are mainly used to focus on the next waypoint. Barnes et al. (2025) observed that the probability of eye movements matched the rhythm of the steps, with saccade probability peaking in the approximate swing phase of each step, just after the midpoint of the stance. At the same time, electroencephalography (EEG) power increased during the swing phase and decreased during the approach to heel strike, which was mainly observed in the theta and alpha bands, producing an oscillation that also corresponded to the stride frequency of about 2 Hz. This effect was stronger when walking at natural speed than at slow speed.

Some walking studies with simultaneous EEG measurements provide evidence of coordination between saccades, walking behaviour and associated cortical activity. This suggests that locomotion might not only influence our gaze movements but also visual perception. For example, Benjamin et al. (2018) found that walking leads to an increased effect of masking around low contrast stimuli in a contrast recognition task. Cao and Händel (2019) showed that walking can alter contrast perception in the periphery. At the same time, they observed alpha oscillations, which indicate that cortical activity changes depending on walking behaviour. Generally, alpha power is also considered a valuable marker for the inhibition of sensory processing and have been shown to correlate negatively with the neural firing rates in monkeys (Haegens et al., 2011). In a follow-up study, Cao et al. (2020) showed that walking led to a decrease in alpha activity in the occipital cortex in both light and darkness. Moreover, they found that alpha activity was lower in the swing phase than in the double support phase.

Finally, there are two studies that showed varying performance of tasks performed during walking dependent on the walking speed and phase: Davidson et al. (2023) found that the performance of tracking a virtual object with a controller was improved when walking at a natural pace as opposed to walking slowly. Later, Davidson et al. (2024) found phasic modulations and optimal periods of sensorimotor precision across the step cycle. A follow-up study in which participants had to respond to visual targets with varying contrasts showed that the participants' recognition rate fluctuated at a rate of approximately 2 cycles per step (Davidson et al., 2024). In this study, accuracy, reaction time and reaction probability also showed clear oscillations that were systematically linked to the phase of each step.

Our eye movements play an essential role in locomotion. We use our gaze to plan our future path, avoid obstacles, navigate difficult terrain and change direction. Thus, we can adjust our steps at short notice and react to other people walking near us. However, we usually do not monitor each of our steps closely. Instead, we make many task relevant fixations on prominent objects in our FOV mostly related to just-in-time information needed for the next steps and long-term planning. Especially when avoiding obstacles, we also seem to rely on information from the periphery. It also appears that our walking rhythm is closely linked to the rhythm of our eye movements and other processes relevant to visual perception in our brain. As a result, eye movements during walking serve several purposes and are influenced not only by our visual input but also our own movement patterns. This usually makes it difficult to analyse eye movement data recorded during long sequences of natural tasks that include locomotion.

With HMDs equipped with eye tracking sensors, it is possible to record walking and eye movement data simultaneously while maintaining complete control over visual input. This could make it possible to untangle the multi-causal stream of eye movements, to get a better understanding of our behaviour. This setup might enable the extraction of navigation information contained in our eye movements, as these are an essential part of planning walking behaviour. Due to the large amount of movement data that can be measured during a walking task, this could even be automated using machine learning algorithms. It might also possible to use this information as an early indicator of imminent or just initiated behaviour. In particular, fixations associated with future waypoints and changes in direction should precede our actions by several steps. To further investigate this approach, we conducted the study II, in which we recorded eye and walking movements during various tasks in order to predict future waypoints based on these data.

Eye Tracking-based LSTM for Locomotion Prediction in VR

Niklas Stein, Gianni Bremer[*] & Markus Lappe

This chapter was published at the International Conference on Virtual Reality and 3D User Interfaces (IEEE VR) in 2022

Abstract

Virtual Reality (VR) allows users to perform natural movements such as hand movements, turning the head and natural walking in virtual environments. While such movements enable seamless natural interaction, they come with the need for a large tracking space, particularly in the case of walking. To optimise use of the available physical space, prediction models for upcoming behaviour are helpful. In this study, we examined whether a user's eye movements tracked by current VR hardware can improve such predictions. Eighteen participants walked through a virtual environment while performing different tasks, including walking in curved paths, avoiding or approaching objects and conducting a search. The recorded position, orientation and eye tracking features from 2.5 s segments of the data were used to train a LSTM model to predict the user's position 2.5 s into the future. We found that future positions can be predicted with an average error of 66 cm. The benefit of eye movement data depended on the task and environment. In particular, situations with changes in walking speed benefited from the inclusion of eye data. We conclude that a model utilizing eye tracking data can improve VR applications in which path predictions are helpful.

---

[*] Shared co-first authorship

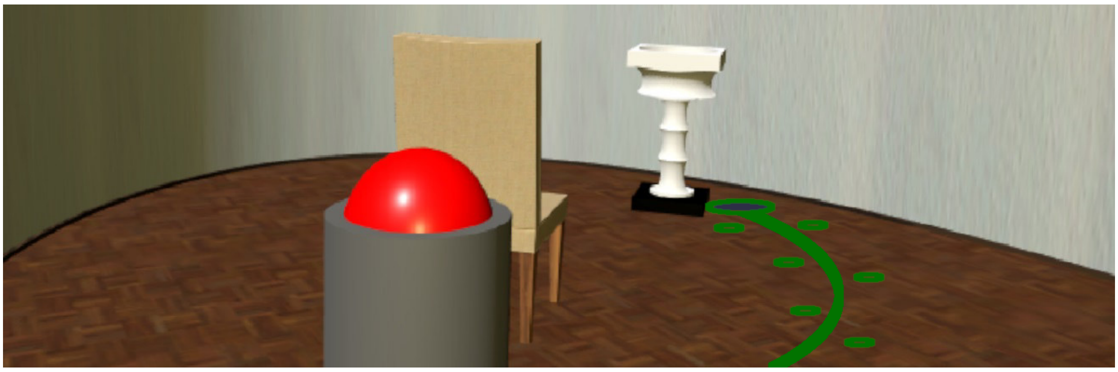


Figure II.1: User perspective of the experiment while walking to a target and avoiding an obstacle.

Introduction

Walking is the most natural and immersive way of moving through a VE (Steinicke et al., 2013). It enhances presence and allows users to unconsciously acquire spatial knowledge (Langbehn, Lubos, & Steinicke, 2018; Usoh et al., 1999). However, walking creates a set of physical problems for VR applications. Collisions with physical walls, objects or other users in real space need to be always prevented. Therefore, large VEs either require large tracking areas or methods such as redirected walking (RDW). With methods that estimate intentions or predict future actions of walking users, RDW and other applications such as programmable interaction patterns of avatars, automatic, user behaviour based level design could become more effective.

RDW is the subtle and unbeknownst redirection of the user to change her trajectory in real space while keeping her perception of movement in virtual space the same. RDW is limited by the user's manipulation threshold (Steinicke et al., 2009). If visually presented and physical walking paths diverge too much, the user notices the manipulation. Although the detection threshold can be slowly adapted over time through learning, the general limitation remains (Bölling et al., 2019).

The implementation of RDW requires applying these manipulations automatically using RDW controllers. Scripted steering controllers have a predefined set of rules based on given information about the VE and the physical space (Razzaque et al., 2001). As a result, RDW manipulations are applied whenever a participant reaches a predefined position. Accordingly, this type of controller needs to be carefully readjusted whenever the VE or the available physical space changes. In the past, several methods to automatize this readjustment process have been developed (Zmuda et al., 2013). The available virtual walking paths can be generated automatically from the environmental data (Zank & Kunz, 2017). Equipped with predefined probability scores for each path and a given skeleton map of the virtual and physical space, the controller then chooses the best manipulation from a predefined set during walking. Generalised controllers follow a different approach. Instead of using information about the VE, algorithms such as steer-to-center, steer-to-orbit, steer to way-points, or steer in figure-eight patterns are designed to work in any VE by steering users along specified physical paths (Nilsson et al., 2018).

RDW controllers can greatly benefit from predictions or assumptions about the upcoming behaviour of the user. Given predictions of the user's virtual path, manipulations can be applied earlier and with a smaller divergence between real and virtual paths. One source for the prediction in unknown environments can be recurring behaviour patterns. For example, assumptions such as typical walking paths (C. Hutton & Suma, 2016) have proven useful in this context. Besides general assumptions, predictions of upcoming locomotor activities may also be gained from observing the user's prior behaviour and actions, such as the trajectory of the user's path over the immediate past and the orientation of her head and body during that time.

Another potential source of information is the user's gaze pattern. When walking, humans typically use specific gaze strategies to inspect their future path, monitor their next target and avoid obstacles (Hayhoe & Ballard, 2005; Hollands et al., 2002; 't Hart & Einhäuser, 2012). Therefore, the pattern of gaze during walking contains valuable information about the user's locomotor intentions. Gaze tracking has recently been incorporated into commercially available HMDs. Although their current quality does not match that of research-grade eye trackers (Lohr et al., 2019; Sipatchin et al., 2021; Stein et al., 2021) it appears still possible to use gaze information to improve locomotor predictions of VR users (Bremer et al., 2021; Gandrud & Interrante, 2016; Zank & Kunz, 2016a, 2016b). Since gaze behaviour depends to some degree on the task (for example obstacle avoidance is different from walking straight to a target (Rothkopf et al., 2016)) the usefulness of gaze data for walking prediction might likewise depend on the situation.

In the present study, we explore how combinations of position, orientation and gaze data can be used to predict a user's future position during walking in VR, and how the prediction quality depends on the different data features and task demands. Our prediction model uses an artificial neural network for time series prediction, the long short-term memory (LSTM) architecture (Hochreiter & Schmidhuber, 1997), which can be fitted to the relevant data without prior assumptions and has been used successfully, for example, for prediction of public pedestrian traffic (Becker et al., 2018) as well as in RDW (Cho et al., 2018).

Related Previous Work

Gaze Behaviour During Walking

Eye movements have a tight correlation with other motor action because we need to move our eye to targets of interest to collect the visual information we need for good action control (Land & Tatler, 2009b). Usually, eye movements to a behavioural target precede any other motor action (Hayhoe & Ballard, 2005; Land & Hayhoe, 2001). Therefore, they present a rich signal for the estimation of action intention (Belardinelli et al., 2016; Bremer et al., 2021; Gandrud & Interrante, 2016; Zank & Kunz, 2016a, 2016b). However, during walking this does not mean that people lock their gaze on a future target, for example a door, at all times. Instead, during walking, eye movements serve the identification of both targets and obstacles. For example, walkers often look at the ground in front a few steps ahead for safe placements of the feet, particularly in uneven terrain (Calow & Lappe, 2008; Hollands & Marple-Horvat, 1996; Hollands et al., 1995; Matthis et al., 2018; ’t Hart & Einhäuser, 2012). This involves not only a direction of gaze towards the ground but also a pitch of the head (D. S. Marigold & Patla, 2008). Before approaching a goal, however, walkers typically direct their gaze towards the goal (Durant & Zanker, 2020; Hollands et al., 2002).

Eye movements are also linked to changes of direction during walking (Hollands et al., 2002). When walking in a curve, for example, walkers typically direct their gaze inward from the curve (Grasso et al., 1998; Imai et al., 2001). Furthermore, eye movements are involved in deciding between alternative targets (Wiener et al., 2011; Zank & Kunz, 2016a) and in searching of targets between distractors (Kit et al., 2014). Thus, eye movements during walking depend on task demands (Tatler & Tatler, 2013).

In summary, eye movements during natural behaviour may be a useful, though not straightforward, signal for predicting a user’s intention and future action which could be extracted in deep learning approaches for action prediction. Their usefulness is potentially different for different tasks. For the present investigation, we chose three tasks to cover a set of general locomotor scenarios in which eye movements likely play a role: searching for a target amongst distractors in a room, walking along a curved path and avoiding an obstacle.

Locomotor Prediction

Different methods for the prediction of future trajectories of users in VR have been proposed in the past. One approach relies on the analysis of possible or probable paths in the current environment (Zank & Kunz, 2017; Zmuda et al., 2013). Available virtual walking paths are generated automatically from the environmental data and probability scores for each path along with a skeleton map of the virtual and physical space allow, for example, to choose an optimal manipulation for RDW. This approach, however, is tied to the specific environment and needs environmental information.

A different approach focuses on observation of the user's prior actions to predict the future trajectory. Zank and Kunz (2016a) used eye tracking data to predict the next locomotion target. In their experiment, two predefined target positions were presented and the user was either instructed to go to one of the targets or freely choose one of them. Based on the chosen targets, they also compared different models that use previous movements to calculate probabilities for the two targets based on assumptions about human walking behaviour (Arechavaleta et al., 2008; Fink et al., 2007; Zank & Kunz, 2016b) and graph representations of the environment (Nescher et al., 2014). In particular, in a narrow T-shaped corridor with little open space, models using eye data were able to provide accurate predictions earlier than models without eye data. Later in a trial and in cases with open space, the overall prediction accuracy was higher and including eye data had no benefit for the prediction at those times.

Gandrud and Interrante (2016) also implemented a binary choice between two walking targets: Users walked along a virtual hallway, at the end of which they had to decide between two targets to approach. In the experiment they measured head direction, gaze direction and the position relative to the virtual hallway midline and compared these three measures for predicting the chosen target. They found that both head orientation and gaze orientation had the potential to be useful in predicting a person's future direction of locomotion.

The use of eye data in these approaches assumes that the gaze position directly precedes the direction of human walking. Indeed, there is evidence supporting this notion (Brument et al., 2019; Hollands et al., 2002; Land & Tatler, 2009b; Tuhkanen et al., 2019). However, since eye movements during walking also depend on task demands (Tatler & Tatler, 2013) a more complex processing procedure of gaze data could allow an even better prediction. Moreover, these pioneering previous studies only distinguished between discrete walking decisions. Further scenarios with fewer restrictions need to be evaluated to advance the use of behavioural measures for locomotor predictions.

A promising current approach to predict future positions is the use of deep learning models. These models can be fitted to the relevant data without prior assumptions. They have been successful, for example, with respect to public pedestrian traffic (Becker et al., 2018; Yu et al., 2020) or for the prediction of future gaze directions (Cornia et al., 2018; Feng et al., 2020; Hu et al., 2020, 2021; Xu et al., 2018). With respect to walking prediction, Cho et al. (2018) presented a preliminary study in which they implemented a deep learning model for locomotion prediction in the context of RDW. They used head position and orientation to train an LSTM model (Hochreiter & Schmidhuber, 1997) to predict the user's position 100 frames (about 1 second) into the future while the user navigated a maze. They report that the prediction worked well for two example users.

However, their model is limited to the pre-defined maze map they used and did not include gaze data.

Scope of this Study

In the present work, we created a scenario in which participants fulfilled a range of typical VR tasks, including walking in curved paths, avoiding or approaching objects and conducting a search. Based on the walking and gaze data, we trained and evaluated different indoor path prediction models with an LSTM architecture that had no information about the used VE. We were especially interested in whether a user's eye movements during walking could be a useful addition to the model. Therefore, we analysed the stability of the eye tracking over the duration of the study, described typical data patterns during different tasks and evaluated under which circumstances eye tracking data contributes to a better prediction of future position.

Methods

Participants

Eighteen participants (8 female) completed the experiment. The participants' age ranged from 20 to 47 years ($M = 27, SD = 6.34$). Participants gave written informed consent and the experimental procedures were approved by the Ethics Committee of the Department of Psychology and Sports Science of the University of Münster. Apart from the two authors who participated in the experiment (N.S. and G.B.) participants were naïve to the purpose of the experiment. Two additional participants we initially acquired had to be excluded due to failed recordings.

Materials

The virtual environment was displayed in an HTC Vive Pro Eye with a resolution of 1440×1600 pixels per eye at a frame rate of 90 Hz and a nominal FOV of 110°. The experiment was conducted at the VR laboratory of the Department of Psychology and Sports Science of the University of Münster. An area of 6 × 11 m was tracked using 6 Vive Lighthouses 2.0. During the experiment, all positional tracking data were Kalman filtered. The experiment was run using Unity3D on an MSI GE63VR 7RF Raider notebook equipped with an NVIDIA GTX 1070 graphics card. The notebook was carried in a backpack and supplied with power by a cable from the ceiling. A Vive tracker was attached to the backpack to record the body orientation and a Vive controller was used as input device. Eye tracking data were recorded using the integrated Tobii eye tracker of the Vive Pro Eye with a nominal accuracy of 0.5°–1.1° within a FOV of 20° at an output frequency of 120 Hz and a trackable FOV of 110°.

Eye Tracking Procedure and Questionnaires

At the beginning of the experiment, eye tracking calibration was done using the calibration method provided by Tobii Software.

In the last decade, eye tracking cameras have been included in several commercially available VR headsets (Clay et al., 2019). Different aspects like latency, accuracy and precision of eye tracking data in VR have been evaluated (Lohr et al., 2019). Under optimal conditions, the HTC Vive Pro Eye showed eye tracking data delays of around 50 ms and an accuracy below 1° in central, but up to 10° in very peripheral positions (27°). Eye tracking precision varied from 1.4° to 3.5° (Sipatchin et al., 2021; Stein et al., 2021). This makes the Vive Pro Eye suitable for a rough online estimation of gaze in VR, although it is not clear how side effects such as slippage through head movements during natural walking affect the measurements. To evaluate possible slippage of the HMD on the head and therefore the eye tracking calibration stability before and after the experiment, 16 participants did an additional, custom-made, simple calibration procedure with 5 fixation positions before and after the experiment.

The simulator sickness questionnaire (Kennedy et al., 1993) (SSQ) was completed before and after each session. Additionally, the participants completed the Slater-Usoh-Steed questionnaire for immersion (Slater et al., 1994) (SUS) after the session. Both questionnaires were translated into German. The authors did not participate in the questionnaires.

Tasks and Virtual Rooms

The virtual environment was divided into three rooms: search room (for an example see Figure II.4a in section II), transition corridor (example in Figure II.4b) and obstacle room (for an example see Figure II.1) in which all participants did ten trials including different tasks. To provoke a lot of natural walking in a limited physical space, search room and obstacle room were mapped onto the same physical space in an impossible room scenario (Suma et al., 2012). Whenever a participant moved through the transition corridor to the door on the other side, an entry to the room opened on the other side and the interior changed. Participants were asked to maintain a natural walking speed during the experiment while performing the following instructed tasks.

Search Room — Free Exploration

In the search room, participants had to look for a search object among six identical looking distractor objects. One object was placed in the centre. The others formed a hexagon around it (see Figure II.4a). All objects had a random yaw direction orientation and a distance of 2 m to their next neighbour objects. Whenever an object was reached, the participants could test whether this object was the search object by holding the controller close to it while pressing the trigger on the back of the controller. The result was signalled by a red or green light above the object. In addition, a sound was played if the search object had been found.

During the task, the participants were free to decide which object they wanted to go to next and did not know which object was the search object beforehand. The target position was individually pseudo-randomised for each participant. After completing the task, participants walked through the transition corridor to the obstacle room.

Transition Corridor - Curved Path

The transition corridor connected the two other rooms (see Figure II.4b) and followed a curved path with a radius of 5.5 m. Participants had to walk through the corridor to pass between the rooms. Data from the transition corridor was obtained to investigate walking along a curved path. Since the participants went back and forth between the rooms, ten right curves (search room to obstacle room) and nine left curves (obstacle room to search room) were recorded for each participant.

Obstacle Room — Straight Path and Obstacle Avoidance

In the obstacle room, participants were instructed to walk to a target object while avoiding an obstacle that might be positioned along the way. For each 4 m walk, the participants first positioned themselves on a white field in front of a pole with a red button. Pushing the button with the controller made the pole and button disappear and the target object and the obstacle (a chair) appear. The task had 4 different conditions: obstacle centred, obstacle 30 cm to the left, obstacle 30 cm to the right and no obstacle. During each visit to the obstacle room, these 4 conditions were run in pseudo-randomised order. Figure II.1 shows the button, the obstacle and the target at example positions and a typical walking path with footsteps. Note that in the real experiment, the three objects were never visible at the same time. The obstacle was placed in the middle between the button and the target. After selecting the target with the controller (by pressing the trigger at the back while holding it close to the target), the target and obstacle disappeared and a new red button appeared at a new location in the room. The participants then went to the new position and repeated the procedure. After four walks in each trial, the participants returned to the search room via the transition corridor.

Data

All raw data files are freely available from https://osf.io/b43uv/. For the analysis, data were sampled at intervals of 50 ms. Periods with missing data and periods in which the participant remained standing without any locomotion (threshold = 0.15 m/s) were removed. Such periods of prolonged standing occurred, for example, when the participants did not immediately start walking at the beginning of the experiment. Naturally, the eye tracker missed data whenever the participant briefly blinked. In this case, missing eye tracking values were linearly extrapolated from the data before the blink.

For the prediction model, we distinguished three categories of data:

1. Positional data from the Vive's infrared tracking system.
2. Orientation angles from the IMU.
3. eye tracking data from the Vive Pro Eye's eye tracking system.

Prediction Model

Features and Labels

The model aimed to predict the user's location in the VR environment 2.5 s into the future. Thus, the output of our prediction model was defined as the direction vector from the current head position (X_t^H, Y_t^H, Z_t^H) to the position 2.5 s into the future $(X_{t+2.5s}^H, Y_{t+2.5s}^H, Z_{t+2.5s}^H)$.

Since our prediction model should be based on user behaviour without information about the environment, it needed to use a coordinate system in a reference-frame attached to the user and not to the environment. For the present study, we used a head-fixed coordinate system, which has shown the best results previously (details in Bremer et al. (2021)).

To set up this coordinate system, we used the average head orientation in the horizontal plane of the input sequence $(\overline{\Phi_{t-i}^H}, \overline{\Theta_{t-i}^H}, \overline{\Psi_{t-i}^H})$ to create a fixed reference yaw angle that was used to describe each input-output-pair. This reference angle, along with its orthogonal directions in the horizontal and vertical planes, provided the axes of the head-fixed coordinate system. Thus, the label direction vectors $(\vec{F_t})$ were rotated using the reference yaw angle. Lower case letters are used to express variables in the new coordinate systems. (e.g., f, ψ, θ).

$$\begin{aligned}\vec{F_t} = (F_t^X, F_t^Z) = (X_{t+2.5s}^H - X_t^H, Z_{t+2.5s}^H - Z_t^H)\\ f_t^x = \cos(-\overline{\Psi_{t-i}^H})F_t^X - \sin(-\overline{\Psi_{t-i}^H})F_t^Z\\ f_t^z = \sin(-\overline{\Psi_{t-i}^H})F_t^X + \cos(-\overline{\Psi_{t-i}^H})F_t^Z\end{aligned} \tag{II.1}$$

The index i in these equations represents the different steps in the input sequence. A total of 7 input features were selected for our model. First, the current two-dimensional velocity of the head in the horizontal plane (V_{t-i}^X, V_{t-i}^Z) was calculated. Height was not used, as there is no evidence for the relevance of this information with regard to the direction of motion on a plane. These input velocities were rotated using the reference yaw angle to convert them to the head-fixed coordinate system.

$$\begin{aligned}v_{t-i}^x = \cos(-\overline{\Psi_{t-i}^H})V_{t-i}^X - \sin(-\overline{\Psi_{t-i}^H})V_{t-i}^Z\\ v_{t-i}^z = \sin(-\overline{\Psi_{t-i}^H})V_{t-i}^X + \cos(-\overline{\Psi_{t-i}^H})V_{t-i}^Z\end{aligned} \tag{II.2}$$

Second, we added the yaw and pitch angle of the head (Ψ_t^H, Θ_t^H) and the direction of eye gaze (Ψ_t^E, Θ_t^E) to the list of the features. Both features might be informative since humans usually orient their head to the target and direct their gaze to the floor along the future locomotor path. Third, we added the body tracker yaw (Ψ_t^B), which was captured by the additional Vive tracker in the subject's backpack. This provided the model with information about body orientation. The reference angles were subtracted from all angles. Before our final scaling, the angles were set between -180° and 180°, with 0° indicating the reference direction.

$$\begin{aligned} \psi^H_{t-i} &= \Psi^H_{t-i} - \overline{\Psi^H_{t-i}} \\ \theta^H_{t-i} &= \Theta^H_{t-i} - \overline{\Theta^H_{t-i}} \\ \psi^B_{t-i} &= \Psi^B_{t-i} - \overline{\Psi^H_{t-i}} \\ \psi^E_{t-i} &= \Psi^E_{t-i} - \overline{\Psi^H_{t-i}} \\ \theta^E_{t-i} &= \Theta^E_{t-i} - \overline{\Theta^H_{t-i}} \end{aligned} \tag{II.3}$$

The input length was set to 2.5 s (50 samples per input), i.e. the prediction was based on the present data and 2.5 s of past information.

To compensate for possible asymmetries resulting from the architecture of the VE (for example the positioning of the room entrance might have led the participants to favour right or left curves in the search room), every second data pair was mirrored (left-right).

Processing

The data went through two LSTM layers with 64 hidden units, a dropout layer ($p = 0.3$) and a dense output layer for the predicted coordinates.

We chose Adam as the optimiser with a learning rate of 0.003 (Kingma & Ba, 2014). A weight decay of 1×10^{-4} was added to prevent overfitting. The model was trained for 20 epochs using the mean squared error as the criterion and a batch size of 64.

To obtain a single comprehensible value for estimating the quality of the prediction, we computed the mean displacement error (MDE) as the distance between the true value and the prediction. To evaluate the influence of eye tracking and IMU data, we also created models without this data to compare them to the main model, which included all features. The model without IMU data also omitted the eye tracking features, because a tracking system equipped with an eye tracker but without IMU seemed to be an unlikely use case.

We implemented leave-3-out-cross-validation at the group level. One participant's data were used as validation data and two participants' data were used as test data. All remaining participants were used to train the model. The validation data were used for hyperparameter optimisation. All data were z-standardised with scalers fitted to the training data.

Results

Participants took on average 14 minutes to complete the experiment.

Evaluation of Eye Tracking Stability

The eye tracking calibration analysis revealed an average Euclidean eye tracking error of 1.99° before and 2.07° after the experiment (no significant difference, $p > .05, t(15) = -0.31$) across all tested participants. Thus, the overall error stayed small over the experiment. Within subject, tests of fixation positions before and after revealed a significant mean difference of 1.38° ($p < 0.001, t(15) = 5.37$). Thus, individually, the eye tracking changed non-systematically by a small amount, probably due to HMD slippage during the natural walking experiment. Centred fixation positions showed standard deviations up to 1.6°. The more peripheral fixation targets at ± 15° had standard deviations up to 2.2° (see Figure II.2).

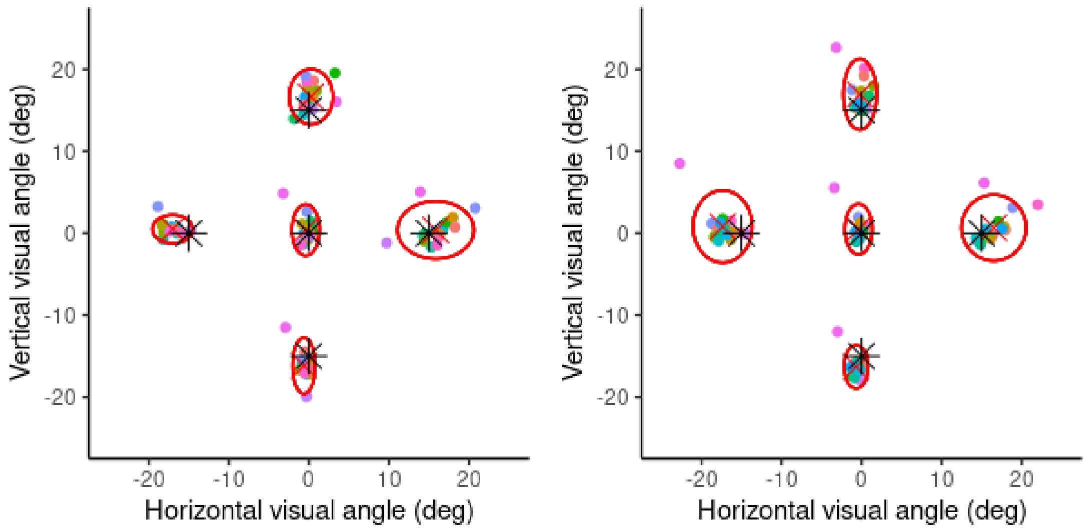


Figure II.2: Comparison of custom eye tracking calibration before (left) and after (right) the experiment. The five black asterisks represent the fixation targets. Red crosses show mean fixation positions across all participants (coloured dots) and two standard deviations (ellipses).

Overall Quality of Prediction

The full data set contained 156,076 input-output pairs to train the model. In the 2.5 s labels, the participants travelled an average distance of 165 cm with a mean walking speed of 0.72 m/s. The prediction model using all features (position, orientation and gaze) produced an MDE of 66 cm for a 2.5 s prediction. A model using only position and orientation performed slightly worse, with an MDE = 68 cm. A model that used only position data produced an MDE = 78 cm. For comparison, a linear regression model considering all features reached an MDE of 93 cm ($SD = 8$ cm) and an extrapolation benchmark based on the most recent positions reached an MDE of 131 cm ($SD = 16$ cm). Figure II.3 shows examples of predicted paths from the full model in the search room and the transition corridor.

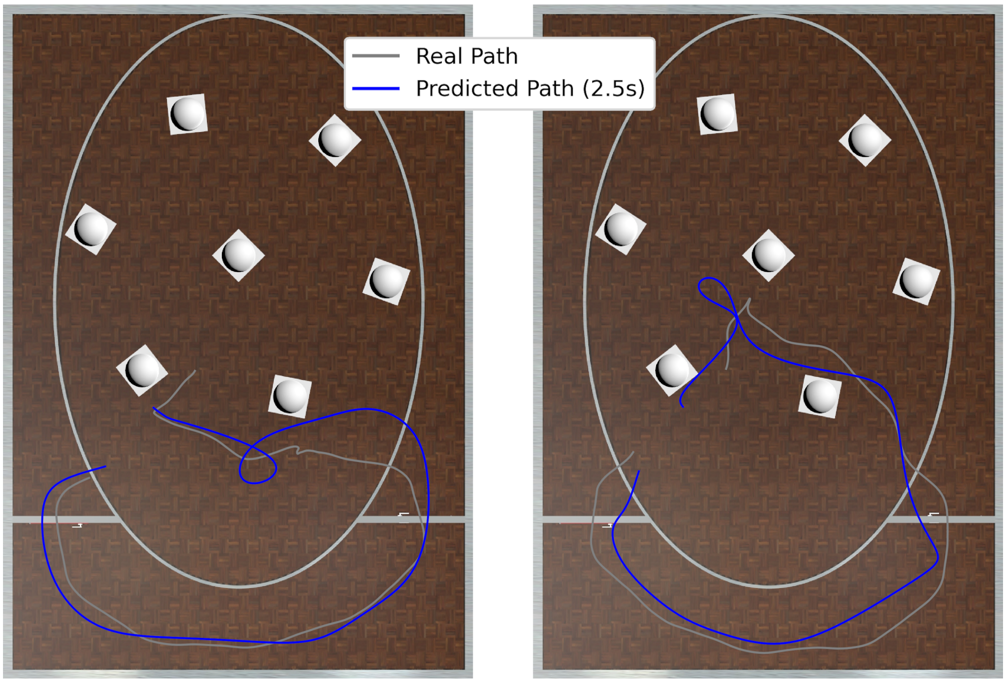


Figure II.3: Two examples of 15 seconds of sequential path predictions. The gray line depicts the real path that the participant walked. The blue line shows the path predicted by the model. Prediction at each data point was based on the preceding 2.5 seconds of movement.

Using the testing method proposed by Nadeau and Bengio (Nadeau & Bengio, 2003) for cross-validated data (alpha level = 0.05), the difference between the LSTM model and the linear model was significant ($t(5) = -11.08, p < 0.001$). The Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) was used to test the differences between the three LSTM models. The model using only position data produced significantly larger error than the model using all features ($t(5) = -6.99, p < 0.01$) and the model using only positional and IMU features ($t(5) = -4.92, p < 0.01$). The model using all features produces significantly smaller error than the model using only positional and IMU data ($t(5) = -3.01, p < 0.05$). Thus, eye tracking significantly improved prediction quality, even if only by a small amount (2.78%). Further analysis showed that using a gated recurrent unit (GRU) model did not lead to a significantly better prediction (details in Bremer et al. (2021)).

Analysis of the Different Rooms

Next, we compared the prediction performances of the model in the three rooms to test whether the improvement gained by eye tracking depends on the task and behaviour of the participants.

Search Room

In the search room, participants looked for a search object amongst six distractors. Each object had to be closely approached to check if it was the search object. On average, the participants walked with a speed of 0.6 m/s and found the search object in the 4th attempt ($SD = 0.6$). Typically, after entering the room, participants walked to the object closest to the entrance and approached one of the outer objects (either clockwise or anti-clockwise) next. In most trials, participants then tested the object in the centre and afterwards continued with the remaining objects (see Figure II.4a).

For the model trained with all data, the prediction error in the search room was higher than the average error across all rooms (see Table II.1). One reason could be quantity of training data since only 14.36 % of the data originated in the search room. However, training the model specifically on the *Search Room's* data improved the prediction only marginally (MDE of 84 cm). This suggests that the lower prediction quality in the search room might be related to the particular task performed, which might be more difficult to predict than the tasks in the other rooms.

The inclusion of eye data improved the prediction in the search room by a small margin, both in the model trained with all the data and in the model specifically trained on the search room data. Thus, eye movement data provided a benefit for the prediction in the search task.

Transition Corridor

In the transition corridor, participants had to walk along a curve to proceed from one room to the other. Participants passed the transition corridor at a mean speed of 0.9 m/s.

As the participants went around the curve, they looked towards the inside of the wall and eventually at the door (see Figure II.4b). Hence, gaze was to the inside of the curve at almost all times.

The full model trained on all data achieved a prediction error of 68 cm in the transition corridor (see Table II.1). Training a model specifically on the transition corridor data without including data from the other rooms reduces the error to 61 cm, showing a

Table II.1: 2.5 s prediction error for different rooms

Training Data	Eye Data	Test Data Search	 Corridor	 Obstacle
All	yes	88 cm	68 cm	59 cm
All	no	90 cm	68 cm	62 cm
Search	yes	84 cm	124 cm	103 cm
Search	no	86 cm	120 cm	101 cm
Corridor	yes	116 cm	61 cm	103 cm
Corridor	no	114 cm	60 cm	97 cm
Obstacle	yes	106 cm	106 cm	57 cm
Obstacle	no	105 cm	111 cm	59 cm
All*	yes	88 cm	75 cm	70 cm

*The number of samples in the transition corridor and the obstacle room was artificially lowered to match the number of samples in the search room.

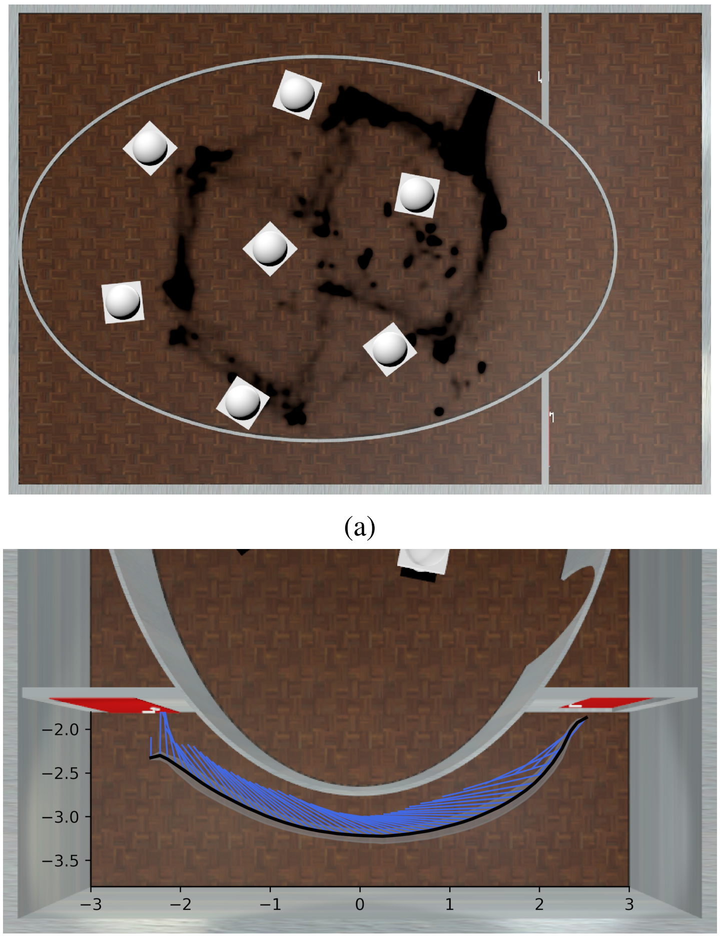
noticeable advantage. Note that 28.56 % of the full model data came from the transition corridor.

Eye tracking, on the other hand, did not prove useful in the transition corridor. When trained on all data and when trained on only the transition corridor data, the model with eye tracking data gave even slightly worse results than the model without eye tracking.

Obstacle Room

In the obstacle room, participants had to walk to a target object while avoiding an obstacle along the way (see Figure II.1). Participants moved at an average speed of 0.62 m/s through the obstacle room. They passed 90 % of all right obstacles on the right and 87 % of all left obstacles on the left. When the obstacle was placed in the centre, 68 % of all obstacles were passed on the right. For the model, this bias should not matter since half of the data were mirrored.

Every fourth walk contained no obstacle and participants could walk straight to the target. Comparing gaze data in these trials to those trials that contained an obstacle allowed us to look at differences in the eye movements depending on whether the obstacle was present or not. Figure II.5 shows the proportion of gazes (gaze prevalence) that were directed to the target (blue), obstacle (green), ground (orange), or walls (pink) as a function of time from the start of the walk. Conditions with an obstacle are plotted as lines, conditions without an obstacle as dashed lines. The figure shows that, at the start of the walk, within the first second, participants looked mostly at the ground. Then, when no obstacle was present, they looked at the target object and kept their gaze there most of



(a)

(b)

Figure II.4: Results of the experiments’ three parts (all participants). (a) Orthographic Projection of the search room with its 7 objects. Each black dot represents a position that one of the participants occupied in that room. The transparency represents path frequency (more frequently walked paths are shown darker). The black spot at the top right shows the entrance to the room. (b) Bird’s eye view of the transition corridor. The black line is the average walking path of all participants (standard deviation in gray). The blue lines show the mean viewing direction at the respective position.

the time. When an obstacle was present, participants looked less frequently to the target, partially because they looked at the obstacle, but also because they looked more often to the ground than when no obstacle was present. Overall, over the course of walking to the target, our participants looked progressively more often at the target and less often at the ground. Right after start, half of all gazes were directed to the obstacle when an obstacle was present. Gazes to the obstacle became less frequent after three seconds. At

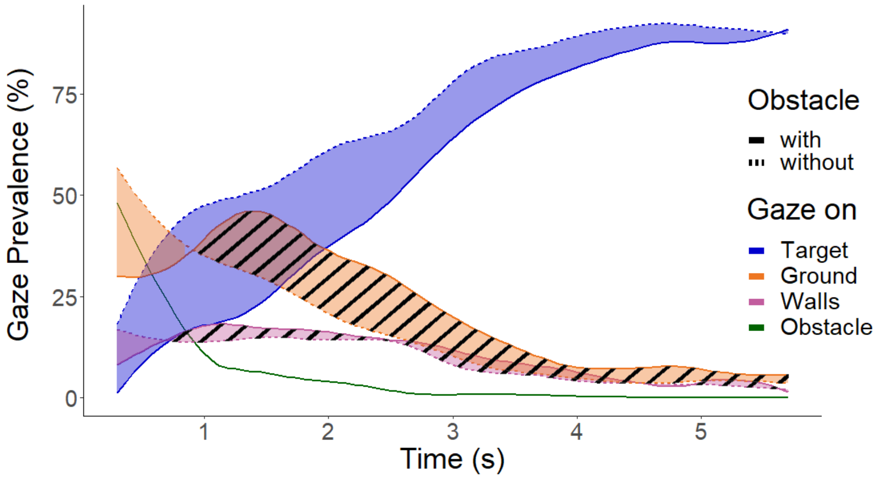


Figure II.5: Where did participants look in the obstacle room? The coloured lines represent the relative prevalence of looking at the target, ground, wall, or obstacle as participants walked from the buzzer to the target. Conditions with and without obstacle are indicated by continuous and dashed lines, respectively. Coloured areas between the lines identify phases in which participants looked to the respective object more often when the obstacle was absent than when it was present. Shaded areas indicate phases in which participants looked to the respective object more often when the obstacle was present than when it was absent.

that time, in many trials, the obstacle had been passed and was no longer visible.

The MDE in the obstacle room (59 cm) was smaller than the mean error over all rooms (see Table II.1). When the model was trained specifically on obstacle room data, the error was even smaller (57 cm). 57.08 % of the data originated in the obstacle room.

The inclusion of eye data provided a small advantage to predictions, both when data from all rooms was used for training and when only the obstacle room was used.

Comparison Between Rooms

When comparing the rooms, the best prediction result was achieved in the obstacle room, followed by the transition corridor and the search room. While there were differences in the amount of data collected in different rooms, the prediction results still hold when the amount of data was artificially reduced to the same number of observations in all rooms (see Table II.1). Training the model on only a single room's data slightly improved the predictions in that room at the expense of prediction accuracy in the other rooms.

Dependence of Eye Tracking Benefit on Locomotor and Gaze Behaviour

Eye movements provided the biggest benefits in the Search and Obstacle Room. This indicates that task and behaviour can influence the importance of eye tracking data for the model.

Figure II.6a shows the improvement of prediction error for a model with eye movements compared to a model without eye movements as a function of the mean acceleration in the 2.5 s input data segment. Colours indicate the different rooms, transparency shows the amount of data that was available. Prediction quality improved when acceleration was larger, most notably in the obstacle room. This shows that eye movements were particularly useful when the data segment contained high acceleration. Figure II.6b shows how the prediction error varies with the mean acceleration in the 2.5 s data segment that was to be predicted. Here, prediction quality was best when the user decelerated in the obstacle room. This suggests that eye movements are most useful for prediction in situations in which the user is likely to stop, presumably because she is close to the target. The search room and the transition corridor did not show such dependencies. However, in these rooms, large accelerations and decelerations were less likely to occur, because they were not part of the typical task-related behaviour in this room.

Besides the acceleration in the walking data, gaze data were also useful to indicate situations in which eye tracking reduces the prediction error. Figure II.6c shows the improvement of prediction error for a model with eye movements compared to a model without eye movements as a function of the gaze distance in the 2.5 s input data segment, i.e. the distance to the object at which the user currently looks. In the obstacle room, prediction quality benefits the most from eye tracking when the gaze target is close. This tendency was also present in the search room, but not in the transition corridor.

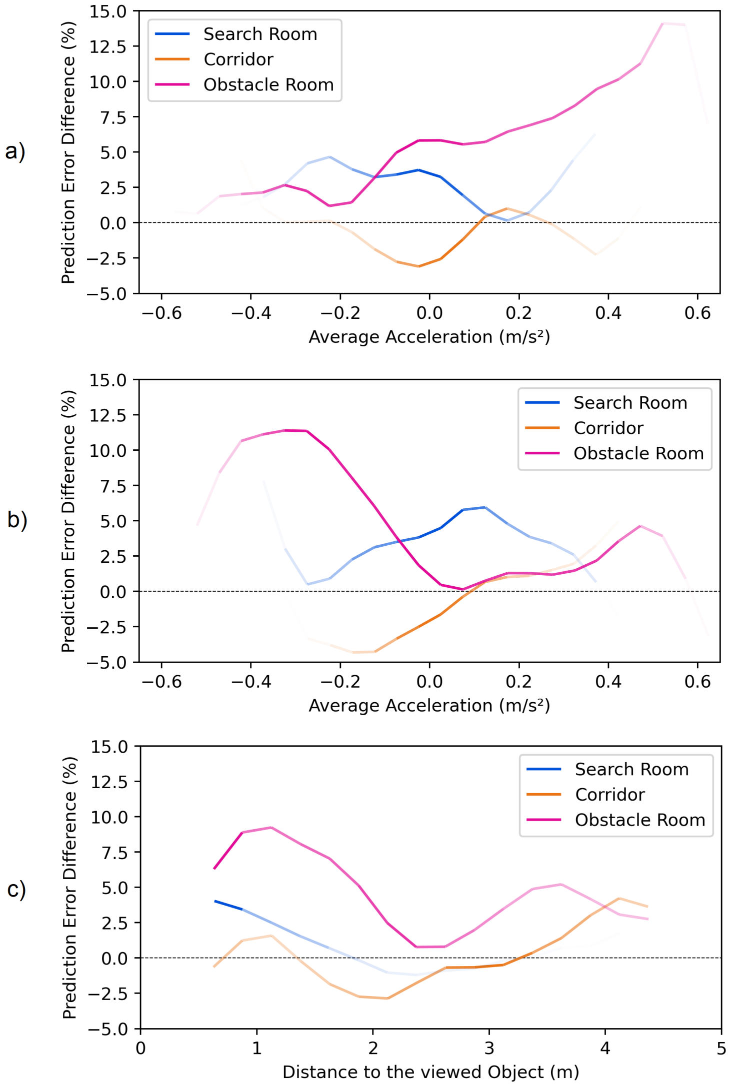


Figure II.6: Added value of eye tracking in the model. (a) shows the model advantage as a function of the acceleration in the input sequence position data. A 0.25 m/s² long Gaussian rolling window (standard normal distribution) was used to smooth the data. (b) shows the model advantage as a function of the acceleration in the true output paths. The same rolling window was used. (c) shows the model advantage as a function of the gaze length, i.e. the distance between the observer and the point where their gaze hits the world, at the moment of prediction. A 1.25 m long Gaussian rolling window was used. The transparency of the line colours indicates the number of observations that factored into each data point.

Questionnaires

The mean SSQ total simulator sickness increased from 7.48 (*SD* $= 7.70$) before the experiment to 20.36 (*SD* $= 34.55$) after the experiment. However, mean values around 20 should not be automatically attributed to a bad simulator in novice VR users (Bimberg et al., 2020). Therefore and because we did not notice a high occurrence of motion sickness, when talking to the participants after the experiment, we would interpret the value increase as a typical result from an experiment that included a 14 minutes long physical task (walking with a VR backpack) and was done by novice users. Moreover, such physical activity has led to increased sweating, which is also represented as a sickness-symptom in the SSQ.

The results from the SUS questionnaire indicated that the users perceived the presented VE as immersive. Participants scored an average of 4.99 (*SD* $= 1.28$, *min* $= 2.83$, *max* $= 7$).

Discussion

We investigated whether data obtained from eye tracking devices in current VR hardware can be used to enhance an LSTM model of locomotor path prediction for natural walking in VR. The model predicted future walking positions trained on position, orientation and eye tracking data from a free walking scenario, in which users performed different tasks in differently structured rooms. The full model produced good predictions that exceeded those of linear regression and simple extrapolation models.

To evaluate the use of eye tracking in this model, we compared it against models that used only position and orientation, or only position data for training. The model including eye tracking significantly improved prediction compared to those models, albeit only to a small amount overall. We evaluated the impact of eye tracking data in the model specifically for the different rooms, tasks, and walking behaviours. We found eye tracking benefits in search and obstacle avoidance tasks, and especially in situations in which users changed their walking velocity.

Quality of Eye Tracking in VR

A prerequisite for the predictive utility of eye data in VR is the quality and stability of the eye tracker. Different currently available eye trackers have different temporal resolutions, making them more or less usable for studies needing temporal accuracy in the millisecond range (Stein et al., 2021). For the present study, temporal accuracy was less critical since we averaged data over 50 ms intervals and based our prediction on data segments of 2.5 s. Since locomotor behaviour in walking is usually smooth, a small temporal lag of the eye tracker would not be detrimental in our prediction scenario. However, during extended periods of walking, on average 14 minutes in our study, spatial accuracy of the eye tracking might deteriorate, for example, if the HMD slips on the head during user movement.

To check eye tracking quality, we measured fixations to a standard set of five fixation targets before and after the experiment. We evaluated both the average error before and after the experiment and the within-subject change of the fixation positions before and after. There was an average error of 2° in the eye tracking data within our 30° FOV, which did not increase significantly after the experiment. The average within-subject change of the fixation positions was 1.38°. Therefore, we consider the stability of the eye tracking suitable for our purpose of measuring general directions of gaze in walking experiments. While this is encouraging for the use of eye tracking in VR, we believe that applications that require eye positions should include such a before-and-after calibration check.

Gaze and Locomotor Behaviour in VR

Gaze behaviour during locomotion in the real world shows certain characteristics, like looking on the ground in front in cluttered environments (Calow & Lappe, 2008; Matthis et al., 2018; Patla & Vickers, 1997; 't Hart & Einhäuser, 2012), looking at the target in approach (Durant & Zanker, 2020; Hollands et al., 2002), or looking towards the inside of a curve (Grasso et al., 1998; Imai et al., 2001). The observed behaviour in the different rooms is consistent with these specifics, suggesting that participants acted naturally in these virtual tasks.

The walking speed in the transition corridor was higher than in the other rooms. This might be explained by a lack of task-relevant object interaction, which allowed the participants to pass through quickly. In the search and obstacle rooms, participants walked more slowly and interacted with objects more deliberately. In the obstacle room, participants glanced more frequently to the ground in the first 3 s of their way to the target when an obstacle was present, consistent with looks to the ground in natural behaviour for negotiating a path containing obstacles (Calow & Lappe, 2008; Matthis et al., 2018; Patla & Vickers, 1997; 't Hart & Einhäuser, 2012). As the obstacle blocked the view towards the target, participants may have had to look sideways towards their path. Since gaze precedes path (Hollands et al., 1995; Matthis et al., 2018), it would also be possible that curves to avoid the obstacle necessitated smaller shifts of the gaze here. In the search room, the systematic search patterns shown in Figure II.4a revealed that short distances were preferred to arbitrary trial and error.

Use of Eye Tracking for Prediction

Our results show that eye tracking can help to estimate future actions. Gandrud and Interrante (2016) previously showed that gaze data are useful when predicting choice between two potential walking targets in a VR hallway. Likewise, Zank and Kunz (2016a) identified eye data as especially advantageous for binary walking path predictions in long, narrow environments. We expanded their results by following a more general approach, not limited to binary predictions. This provides the potential to be applicable to a broader range of environments and applications. Users were instructed to walk freely in the VE. Our aim was to predict their future position in space. Therefore, a direct comparison of the performance of our model to the prediction methods in those previous papers is not possible with the current data-set. However, it is possible to compare whether eye tracking data were a significantly useful addition for the prediction in the different studies. Eye movements provide useful information when approaching a target (Durant & Zanker, 2020; Hollands et al., 2002). This fits well with the results of the binary prediction studies. However, when walking in a curve with no task, the head orientation is likely to include the same information as the gaze (Grasso et al., 1998; Imai et al., 2001). Accordingly, gaze provided no benefit for the model using only data from the transition corridor.

In the obstacle room, an advantage from the inclusion of eye data was observed for decelerating paths and paths following accelerated motion. It seems possible that the presence of the obstacle influenced the motion and gaze patterns in a way that eye data improved model performance. However, since the obstacle was only looked at for a short time (see Figure II.5), it is not clear if this alone has caused this effect. The advantage could also be caused by a change in gaze behaviour when stopping in front of the target. However, the obstacle room was the only room where stopping was part of the task. Therefore, we cannot clearly differentiate if eye movements were most beneficial for predictions during stop-and-go, obstacle avoidance or a combination of both tasks.

The link between gaze and locomotor activity in natural behaviour and VR suggests that eye tracking might be useful for predicting upcoming user actions. The results from our LSTM prediction model confirm that the addition of eye tracking data can provide a significant benefit in predicting future walking paths. Overall, the MDE of the best model including gaze data was 66 cm for a 2.5 s prediction. The average distance walked during 2.5 s was 165 cm. It remains to be evaluated in future studies how effective this prediction is for different applications.

Possible Application Scenarios

For walking in VR, a particular application scenario is RDW, in which users are steered along a physical trajectory that differs from the virtual trajectory to optimally use the physical space available for tracking (Razzaque et al., 2001). In dynamic RDW, one has to decide in which direction and by how much the user should be redirected at any given point in time. Knowledge of where the user most likely intends to go can be advantageous for implementing RDW fast and effectively (Nilsson et al., 2018). Since our prediction model worked well in different rooms and tasks and since it only uses data directly from the user, it can be considered a useful tool for a generalised RDW controller. Future user studies are needed to evaluate if an RDW controller using eye tracking data for redirection outperforms previous approaches regarding the needed number of resets and the user experience in different physical and VR scenarios.

A very simple way to include our prediction in RDW is to focus on whether a user, at a particular point in time, should be redirected leftward or rightward. In such a case, a simple left/right prediction of the user's intent could be sufficient. To quantify the quality of such a prediction in our model, we took the predicted future positions at each point in time and subjected them to a simple left/right discrimination. The results showed that the model could predict whether the user will turn left or turn right with an accuracy of 84.8 %. A model explicitly trained to distinguish between left curves and right curves would likely achieve even better accuracy. Additionally, there might be some potential for synergistic effects with other methods, such as following behaviour (Nguyen et al., 2020) to improve performance even further.

Our method of locomotion prediction might also be useful in other scenarios. For example, control of non-player characters could benefit from user prediction to avoid collisions. Moreover, valid predictions of upcoming behaviour could be helpful in determining the objects and locations where computational resources should be focused to for example increase their responsiveness or level of detail.

An important feature of our method is that it relies exclusively on egocentric data and does not need information about the environment. This feature makes the method very versatile. By using inside-out tracking, head worn IMUs and eye trackers, the data for prediction can be gathered by sensors worn by the user, not only in VR. For example, similar models could likely be trained to provide rough predictions of locomotion intentions in augmented reality scenarios. Moreover, the anticipation of human actions, such as walking, can also play a key role in the development of assistive robots (Koppula & Saxena, 2015).

Limitations

Although we tried to create an experiment with a representative sample of typical VR walking tasks, our data set is limited to our selection of rooms and tasks. This could compromise the performance of our model in new contexts, as movements associated with our tasks could be disproportionately represented in the data set. This effect is already observable in our results since the models that limited access to some rooms during training produced worse results in the other rooms, compared to the full model (see Table II.1). While the overall smaller amount of data might have decreased prediction accuracy, the discrepancy between known and unknown environments indicates that the transfer from one of our rooms to the others has only been successful to a limited extent. This suggests that performance of our model is likely to drop when used with tasks that were not included in the training phase. However, even these higher errors are still in a range far below the distances travelled. Therefore, we think that our model's path prediction can be used in new situations as long as an error of about 1 m in comparison to the real path is tolerable for the application.

Another limitation is concerned with the predicted variable we used in our model. Minimizing the error between the real and the predicted end position of a 2.5 s subset of data gave us valuable information about the range of MDE and the incremental value of gaze data for the prediction. However, it does not include information on the deviation between the predicted and the true trajectory during the walk. Prediction of the full trajectory could be valuable for some application scenarios and therefore should be considered in extended path prediction models in the future.

Conclusion

eye tracking data from current VR hardware improves deep learning path prediction for natural walking in VR. Eye tracking benefits appear especially in situations in which the user interacts with the virtual environment during locomotion. Including eye tracking data while considering the user's task and behaviour is a useful tool for deep learning path prediction.

Declaration of Conflicting Interests

The author(s) declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

This work was supported by the German Research Foundation (DFG La 952-4-3, La 952-7) and has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951910.

How Does Visual Input Guide Our Gaze?

Many eye movement patterns can be mapped to different task demands. Some task-related gaze patterns, such as turning, are very similar in different scenes and seem to be part of our natural motor programme for the task. Thus, it is possible to predict future actions based on our eye movements to a certain extent.

However, this does not apply to all eye movements and does not fully explain how we decide which objects to inspect next. How do we, for example, decide which object to look at first when there are multiple task-relevant options available? Where does our gaze wander when we look around without any particular intention?

Some of these gaze movements can be explained by general eye movement patterns: For example, when looking at images we seem to have a bias for making horizontal eye movements (Foulsham et al., 2008). This seems to also be the case, while we are searching for something (Gilchrist & Harvey, 2006). Furthermore, we seem to choose locations close to our current fixation as our next fixations when we look around in a natural environment (Gajewski et al., 2005; Tatler et al., 2006, 2011). One possible explanation for this pattern is that we adjust our fixation targets based on the available information and thus on the basis of the distribution of the resolution of the retina (Najemnik & Geisler, 2008).

Even though these biases can explain some of our eye movements, they are not sufficient to fully explain where we intuitively look and which objects are particularly eye-catching. In some situations, this seems to happen almost automatically. When a visual stimulus suddenly appears in our FOV or starts to move, we tend to look at it (Irwin et al., 2000; Tatler et al., 2011; Theeuwes et al., 1998, 1999). Similarly, visible optical flow intuitively triggers saccades and subsequent smooth pursuit movements towards the focus of expansion (Chow et al., 2021). To prepare such a reactive saccade towards an appearing object, we only need between 150 and 200 ms (Smit et al., 1987). In contrast, the preparation time for a voluntary saccade towards a memorised or continuously visible target while viewing a scene is likely to be somewhat longer. However, the exact preparation time is difficult to quantify, as we do not know the exact moment when the decision to move the eyes is made (Gremmler & Lappe, 2017). If we consciously alter a reactive saccade, for example, by making an eye movement to a memorised location in the opposite direction of an appearing stimulus (anti-saccade), our saccade latency increases even further, reaching values between 300 and 500 ms (Smit et al., 1987).

Whenever we make a saccade towards an interesting object, our visual system must somehow select a future fixation point and plan an eye movement towards it, even before it can be seen with the high resolution of the fovea. One theory describing this process is the principle of attentional shift. According to this principle, before each eye movement, our focus first shifts to the location of the next fixation, even before our eyes move. Objects at the location of attention are then processed in a targeted manner. Visual attention could therefore be interpreted as a process that highlights part of our sensory impressions in order to make better use of the information available at that location.

In so-called cueing experiments, participants have to press a button as soon as they perceive a visual target that could either be at an expected position (cued) or somewhere else. If the target appears at a location that was previously indicated by a cue so that the participants could shift their attention before answering, the reaction time is shortened (Posner et al., 1980). In the first experiments of this kind, participants were instructed not to move their eyes. In later experiments, in which participants were asked to fixate on the target when selecting it, it turned out that cueing also had an influence on the timing of the onset of saccades towards the target (Shepherd et al., 1986). In other words, if our attention is already focused on a location that will later be the target of a saccade, the eye movement can be performed more quickly. First, we seem to make a voluntary shift of attention towards the cue. As instructed, our eyes follow this shift to the same location, once the target appears. In contrast, if we first receive an attention cue on the right side, while a saccade target subsequently appears on the left side, our saccade preparation time increases. In this case, we seem to make a reactive shift of attention towards the non-target followed by a second voluntary shift of attention towards the target in preparation for the saccade that follows. This additional shift of attention to a different location just before the saccade increases the latency.

Interestingly, this also means that shifts in attention can occur even when we direct our gaze to another point (covert attention) and when no saccade is ever executed towards the point of focus. Thus, it is possible that our attention has already shifted while our eyes are still fixating the previous target. While we focus our attention on a new object, it is also possible that, at least for a short period of time, two objects are simultaneously in the focus of our attention (Coral Gabbay & Lamy, 2019). Our eye movements therefore do not always reveal where we are focusing our attention and lag slightly behind any ongoing shifts in attention. Accordingly, the terms ’fixation’ and ’focus of attention’ should not be used interchangeably in the context of attention research.

When we focus our attention on a specific location, we can better and more rapidly process targets at that position (Boynton, 2005; Posner et al., 1980; van Ede et al., 2012). However, there is also plenty of evidence that our representation of the visual world is built from a combination of the content we are currently paying attention to and a non-selective global view of what is visible in our FOV (Wolfe et al., 2011). For example, when looking at a forest, we are able to perceive and process the scene as a whole without looking at the details of any individual tree using the so-called non-selective pathway of our visual system (Greene & Oliva, 2009). However, if we want to focus our attention on a tree to inspect individual branches or leaves, it would require us to use the selective pathway (Evans & Treisman, 2005). As a result, shifting attention is an important mechanism to process the details of objects in our FOV. This in turn leads us to a series of new questions: why do we focus our attention on certain stimuli and not others? What are the characteristics of objects that we intuitively focus our attention on? And finally: how are these processes implemented in our brains?

Visual Search and Salience

In many situations, even before we enter a room, we already have a preconception of what we want to do there and which objects we want to interact with. To do this, it is important to be able to locate these target objects in the FOV. The psychophysical paradigm of visual search provides some insights in the basic functions of this process.

Visual search was originally developed to investigate the capacity limits of the attention system (Treisman & Gelade, 1980). During these experiments, participants are asked to search for a target object among a set of distractors as quickly as possible. Usually, a search ends either immediately because the search target pops out quickly to the participant, or after a longer search period. One explanation for this pattern could be that objects in the FOV are either processed serially one after the other, or in parallel.

In their theory of guided search, Wolfe and Horowitz (2017) assume that the type of search (parallel or serial) depends on the salience of the target. They postulate that salience increases with the visible difference between target and distractors and also with the homogeneity of distractors. In other words, if a target is particularly unique and stands out from very similar distractors, it pops out and attracts our attention. Thus, we can solve this search task in parallel. The most important property dimensions in which targets and distractors can be unique or similar are referred to as the basic (or guiding) features. These special features are capable of drawing attention in a way that other factors cannot (Higuchi et al., 2019; Wolfe & Horowitz, 2017): The list of guiding features includes colour, sudden appearance, motion onset and target shape. Other factors that can create such a pop-out effect in search tasks are object size, stereoscopic depth and tilt, optic flow, luminance, line termination and curvature. Indeed, individual targets that differ from the distractors in one of their basic characteristics, usually catch our attention immediately (Treisman & Gelade, 1980; Treisman & Sato, 1990; Wolfe & Horowitz, 2017) and can be found without or within only a single saccade (Binello et al., 1995). Moreover, these stimuli seem to automatically draw our attention towards them since we intuitively execute saccades towards their location when we detect them in a search display (Hulleman & Olivers, 2017).

If no salient target is available because the distractions do not have at least one clear basic guiding feature that differs from the target itself, individual stimuli or groups of stimuli are examined one after the other. In these conjunction searches, the length of the search, during which we typically perform multiple saccades and fixations, is influenced by the time required to examine and reject each distractor (Wolfe & Horowitz, 2017).

The Visual Pathway

To understand how attention shifting, target perception and, to a certain extent, eye movement planning work, it is also necessary to examine the neural structures that perform these processing steps. This section provides an overview of the various elements of the visual pathway in our brain and how these structures are linked to some of the behavioural findings presented in the previous sections.

After light is detected by the photoreceptors, the biochemical signal, which from then on represents our visual input, is transmitted first via bipolar and then via ganglion cells, while horizontal and amacrine cells form lateral connections between the retinal layers (Prasad & Galetta, 2011). To maintain high spatial acuity, the input from the central areas of the FOV is passed on to individual bipolar cells, while in the periphery one bipolar cell has a receptive field that combines several photoreceptors (Prasad & Galetta, 2011). Next, visual information is carried from the eyes towards the brain via the optic nerve which is formed by the axons of the ganglion cells (Prasad & Galetta, 2011). At the optic chiasm, the information from both eyes partially crosses sides, so that from here on a combined image from the perception of both eyes is available.

About 10 % of retinal axons from the optic nerve end in the tectum of the midbrain and in the hypothalamus (where they most likely contribute to the regulation of our circadian rhythm (Joukal, 2017)) the pretectal nuclei and the superior colliculus (SC). Together with input from other sensory organs, such as the vestibular and somatosensory systems and the visual cortex, these connections are important for various ocular functions. For example, they regulate the pupillary reflex and the VOR (De Moraes, 2013; Joukal, 2017). Of these structures, the SC in particular has been shown to be a likely candidate for coordinating our head and eye movements and may even be involved in the preparation of saccades. Indeed, a number of experiments with monkeys suggest that the brain cells in the SC contain a representation of the FOV and are involved in planning and executing our eye movements (Bollimunta et al., 2018; Groner & Groner, 1989; Krauzlis et al., 2013). Already in 1972, Goldberg and Wurtz observed an increase in the activity of individual cells in the SC as soon as the eyes of a monkey began to move. Artificial lesions of the same area led to an increase in saccade latency of 150-300 ms when a saccade was directed to the receptive field of the lesioned brain cells. At the same time, saccade velocity decreased slightly after the lesion (Wurtz & Goldberg, 1972). Interestingly, however, targeting accuracy remained unchanged. In addition, further experiments showed that when a monkey performed hand movements instead of eye movements towards a target, the firing rate of cells in the SC did not increase. Accordingly, the previously observed cell discharges were linked to the preparation of saccades, but not attention shifting (Groner & Groner, 1989; Wurtz & Mohler, 1976).

Further down our visual pathway, the majority of all retinal axons (90 %) form the optic tract and send signals to the lateral geniculate nucleus (LGN) of the thalamus (Joukal, 2017). Although neighbouring cells can inhibit each other, the visual pathway still appears to run mainly unidirectionally from the eyes towards higher-level areas of the brain. Thus, up to this point, the activity of the cells involved is mainly driven by the incoming light. In the LGN itself, however, the retinal axons account for only 5-10 % of the synapses and most of them originate from upward-downward modulating connections from other brain areas such as the reticular nucleus of the thalamus, the pulvinar nucleus and the visual cortex (De Moraes, 2013). Retinal signals in the LGN travel to the visual cortex via three pathways. The parvocellular pathway that includes the information for high spatial acuity and red-green colour vision, the magnocellular pathway that holds the signal for achromatic visual sensitivity and motion vision and a third pathway via heterogeneous koniocellular tracts (Solomon, 2021). We still know very little about the koniocellular pathways, but the lower sampling density suggests that they could provide only a rough sketch of the retinal image, which could still be used to regulate the other pathways and perhaps emphasise particularly important areas of the retinal image (Solomon, 2021).

Processing in the Visual Cortex

From the LGN visual information is transmitted to the two occipital lobes, where the more retinotopic areas are arranged in parallel mirror-symmetrical bands (Grill-Spector & Malach, 2004). Current research suggests that the organisation of visual information in the cortex follows two different basic principles: hierarchical processing and functional specialisation (Grill-Spector & Malach, 2004). Thus, a particular brain function, such as recognizing a particular shape, is typically achieved through a stepwise process. The information is first reflected locally and retinotopically in the brain and then transformed into more abstract and even multimodal representations through a sequence of processes (DeYoe & Van Essen, 1988; Grill-Spector & Malach, 2004). Therefore, many of the early cortical areas of our visual system are defined by their retinotopic maps, whereas others are defined by their function, like their preference for certain classes of visual images such as faces or scenes (Kanwisher, 2010; Wandell & Winawer, 2011; Winawer & Witthoft, 2015).

In general, the average size of the receptive fields in the retinotopic regions is smallest in primary visual cortex (V1). From secondary visual cortex (V2) to third visual cortex (V3), third visual complex accessory (V3a) and visual area V4 (V4), it then increases continuously (Grill-Spector & Malach, 2004; Kastner et al., 2001; Press et al., 2001; Smith et al., 2001). Even in the visual cortex, the representation of the centre of the FOV is magnified. The connections of 1 mm² around the fovea (representing the central 10° or about 2 % of the total FOV) represent about 60 % of cells in V1 (De Moraes, 2013). Interestingly, while V1-V3 have representations of the full FOV, the receptive fields of cells in V4 are mostly located within 3-4° of the fovea, which suggests that this area is specialised for functions that depend heavily on foveal and para-foveal vision (Winawer & Witthoft, 2015). Anatomically, the foveal representations in V1, V2, V3 and V4 are very close to each other at the occipital pole, which suggests they are part of one common map cluster (Winawer & Witthoft, 2015).

The visual cortex is classified not only on the basis of its anatomical structure but also on the basis of its functions. Broadly speaking, the visual pathway in the brain is often divided into two pathways (Joukal, 2017): The ventral 'what' pathway extends from the occipital lobe downwards to the temporal lobe and is responsible for functions related to the recognition of object information such as shape, contrast and colour. The dorsal 'where' pathway extends from the occipital lobe upwards to the parietal lobe and is responsible for functions such as spatial features and movement. Grill-Spector and Malach (2004) investigated hierarchical processing based on retinotopy, motion sensitivity and object selectivity: They found that early retinotopic areas V1, V2 and V3 showed a high degree of retinotopy, but only low specificity for stimulus motion and form. Thus, in principle, V1 and V2 are active in every type of visual task and activation in V1 increases, for example when the contrast of a visual stimulus increases.

In V1 itself, there are three different systems that fulfil separate functions (Joukal, 2017). The first system is formed by three cortical columns and is specific for binocular vision and is thus fundamental for depth perception. The second system consists of cells with identical retinal positions and preferred orientation axes and enables the perception of movement. The third system is responsible for the perception of colours and shapes. As we move from one area to the next, features to which the neurons respond become more complex. Some functions are also represented in several areas: For example, in both macaque monkeys and the human brain, there is a relationship between the areas V1 /V2 and higher areas including V4 for colour processing and the areas V1 /V2 and middle temporal visual area (V5), medial superior temporal (MST) and V3a for motion perception (Brewer et al., 2005; Grill-Spector & Malach, 2004; Winawer & Witthoft, 2015; Zeki et al., 1991). In general, intermediate visual areas such as V3a and V4 show a lower degree of retinotopy and, to some extent, stronger responses to objects and moving low-contrast gratings (Grill-Spector & Malach, 2004). In subsequent studies, V4 neurons were associated with a range of other functions (Roe et al., 2012), such as selectivity for binocular disparity (Hinkle & Connor, 2001), disparity-defined shapes in random-dot stereograms (Hegdé & Van Essen, 2005) and three-dimensional orientation of bars (Hinkle & Connor, 2002). A similar area was found in humans (Hansen et al., 2007) and could also be associated with the encoding of shapes (Dumoulin & Hess, 2007), surfaces (Bouvier et al., 2008) and object-selectivity (Konen & Kastner, 2008).

For higher-order areas further down the processing stream, Grill-Spector and Malach (2004) found that retinotopy continues to decrease, while a higher degree of specialisation develops. Neurons in V5 showed a strong preference for moving over stationary stimuli, while having no object selectivity. Neurons in lateral occipital complex (LO) respond more strongly to objects compared to scenes and textures, but show little response to moving compared to stationary gratings (Grill-Spector & Malach, 2004).

Neural Basis of Attention

With regard to the results of the visual search, we can now ask ourselves which areas of our brain recognise the features of, for example, a salient search target. It seems plausible that neurons that enable basic feature recognition rely on visual input, but not necessarily multimodal input. The fact that we are able to use these features in a parallel search to quickly identify the location of the target anywhere in our FOV suggests that these features are represented in a brain area that is retinotopically organised. In addition, many guiding features require at least the detection of contrast, colour and motion. These considerations suggest that the neurons that can control our attention may be located in one of the intermediate areas of the visual cortex.

In line with these thoughts, Bichot et al. (2005) conducted experiments in which macaque monkeys searched for a target defined by colour and shape, and found interesting neurons in V4. Throughout the search phase, the neurons showed enhanced responses as soon as a stimulus in their receptive field matched a feature of the target. Thus, this detection mechanism worked in parallel across the entire visual field display, even before the target was localised. This could potentially represent the parallel processing of all stimuli in our FOV while we search. The neural response of these attention neurons encodes position and specific target features. Signals are sent from V4 to other areas, such as the inferotemporal and frontal cortex. This signal could therefore ultimately direct spatial attention to a potential target and trigger an eye movement to the same location. This mechanism could explain how, for example, when attention is directed to a stimulus with lower contrast, the influence of this stimulus can outweigh the influence of a competing stimulus with higher contrast (Reynolds & Desimone, 2003).

Interestingly, Bichot et al. (2005) also found that the same neurons responded more strongly to potential targets selected for saccades or foveation, which could potentially represent shifts of attention during serial search when we move our eyes from one potential target to the next. This is consistent with results by Theeuwes et al. which suggest that detecting the presence of a coloured target is always associated with a shift of spatial attention to the location of the object (Theeuwes, 2010; Theeuwes et al., 2008).

Therefore, V4 is likely an important area for mediating attentional effects that recognises or is indirectly involved in detecting important features and then uses them for attentional mechanisms in a spatially and functionally specific manner (Roe et al., 2012). Interestingly, this list of selective features that have been found in V4 could potentially also explain some of the results in visual search. In particular, it might shed light on why some unique features of a given target create a pop-out effect while other features lack the capability to make a target salient.

The neural activity found in V4 in monkeys could explain findings of conjunction searches in which the target does not pop-out and a serial search is necessary. In these searches, the distractors, which shared some but not all characteristics of the target, led to neuronal activity similar to that of the actual target (Bichot et al., 2005). The positions that the animals then worked through serially corresponded to the activity map resulting from the receptive fields of these neurons. In other words, the serial search was the result of several target candidates based on a basic feature-guided parallel pre-selection.

In summary, the neurons found in V4 seem to enable an important part of our attention processes and have a receptive field that encompasses all potential stimuli of a search display in the FOV. Thus, one could conclude that our attention system also operates in a retinal coordinate system.

However, when we search for something in the real world, our target is often not yet visible in our FOV. Instead, we often use a combination of eye and head movements to inspect different parts of our environment. Therefore, we might ask ourselves what guides our head movements? How does the attentional system guide our eye movements while we move our head to extend our FOV? From behavioural observations, we already know that both types of movements are closely linked and that we can intuitively plan ahead and perform our comparatively slow head movements while inspecting our surroundings. Studies on walking have shown that we can gather knowledge about the structure of our environment and even about other people in it, and use this knowledge to adjust our gaze distribution accordingly. However, how exactly our attention system deals with head movements and how these influence our eyes during a search is still an open question.

In Study III, we used an HMD with eye tracking capabilities to closely monitor eye and head movements while participants performed a visual search in a controlled and simplified VE.

Eye and Head Movements in Visual Search in the Extended Field of View

NIKLAS STEIN, TAMARA WATSON, MARKUS LAPPE, MAREN WESTENDORF & SZONYA DURANT

This chapter was published in Scientific Reports (vol. 14) in 2024

Abstract

In natural environments, head movements are required to search for objects outside the FOV. Here, we investigate the power of a salient target in an extended visual search array to facilitate faster detection once this item comes into the FOV by a head movement. We conducted two VR experiments using spatially clustered sets of stimuli to observe target detection and head and eye movements during visual search. Participants completed search tasks with three conditions: (1) target in the initial FOV, (2) head movement needed to bring the target into the FOV, (3) same as condition 2, but the periphery was initially hidden and appeared after the head movement had brought the location of the target set into the FOV. We measured search time until participants found a more salient (O) or less salient (T) target among distractors (L). On average, O's were found faster than T's. Gaze analysis showed that saliency facilitation occurred due to the target guiding the search only if it was within the initial FOV. When targets required a head movement to enter the FOV, participants followed the same search strategy as in trials without a visible target in the periphery. Moreover, faster search times for salient targets were only caused by the time required to find the target once the target set was reached. This suggests that the effect of stimulus saliency differs between visual search on fixed displays and when we are actively searching through an extended visual field.

Introduction

Detailed knowledge of how we search an array of visual stimuli for a unique target has mostly been acquired using stimuli that onset in front of a participant on a computer monitor. Participants are often instructed to keep their heads still and initially focus their eyes on a central point on the screen. Outside controlled experimental conditions like these, we often search for items beyond our current FOV. The work presented here compares visual search when targets are presented within the visual field at the start of a search, to when they appear due to the act of moving the head and eyes to bring them into view. We compare the effect of salience on visual search under these circumstances, anticipating that a more salient stimulus entering the FOV after a head or eye movement would assist the participant in orienting to this stimulus.

Experiments limiting search to a fixed FOV have provided a detailed understanding of how target stimuli are located amongst distractor stimuli. Some targets are found with very little delay after onset within the visual field, regardless of the number of distractors also presented. These are said to ‘pop out’ relative to surrounding distractor objects, which means they are found without the need to make any exploratory saccadic eye movements, or within only a single saccade (Binello et al., 1995). Targets that pop out are differentiated from distractors by some basic attribute (for example opposing colours) (Treisman & Gelade, 1980; Treisman & Sato, 1990; Wolfe & Horowitz, 2017). Otherwise, an effortful search must take place. This involves serially selecting/inspecting each stimulus or ‘clumps’ of stimuli (Hulleman & Olivers, 2017). This kind of ‘serial search’ is suggested to occur by establishing a priority map (Fecteau & Munoz, 2006; Serences & Yantis, 2006), which is a planned sequence of locations to allocate attention toward and thereby explore ordered by the weighted average of the bottom-up salience, top-down feature guidance, scene guidance, historical performance and value according to past reward (Wolfe & Horowitz, 2017). Serial search can be undertaken by shifting covert attention (i.e. by making no overt gaze change) via a priority map of stimuli that can be ascertained within the current functional FOV (Wolfe, 2020). The functional FOV can be considered to encompass the region of the visual field that is resolved well enough that stimuli can be identified for response or rejection without making a gaze shift and is therefore a smaller region than the entire FOV. Outside the functional FOV, the priority map may also take into account what we know about the likely location of any particular target object within that scene. In this case, previous knowledge in the form of spatial memory becomes an input to the search plan. This is referred to as contextual cueing (Tatler & Land, 2011).

Visual search involving a search array that extends beyond the FOV of the participant and therefore requiring head and eye movements, has been less often investigated. To achieve an extensive search array like this, studies must be carried out beyond a 2D computer monitor. Stimuli need to be situated in the real world or presented via a simulation; either through immersive simulators or VR. In this study, we use VR to present a search array extending beyond the participant's initial FoV. Previous studies have investigated visual search in a VR environment and found effects transferred from 2D screens in terms of the overall effect on reaction times of target–distractor discriminability (Botch et al., 2023; Olk et al., 2018). Shioiri et al. (2018) have shown that contextual cueing builds up across trials even when the search display covers 360° and the participant must look around themselves to find the targets. As yet, whether more salient targets are able to support a pop-out-like search when the target comes into view due to the participant's own movement during an ongoing search has not yet been systematically investigated. This is despite the fact that it is of great importance for guiding the attention of users in virtual and natural environments.

On the one hand, Lukashova-Sanz and Wahl (2021) showed that blurring salient regions of a visual scene can guide the observer's attention away from those regions and enable better search performance. On the other hand, studies using naturalistic search stimuli or tasks suggest eye movements are not guided by salience as predicted via image statistics (Haskins et al., 2020; Henderson et al., 2007; Vogel & de Freitas, 2008). In addition, contextual cues and memory of the likely location of targets have been shown to play an important role in search strategy in naturalistic virtual environments (Li et al., 2016, 2018). This opens up the possibility that stimuli found to 'pop-out' when appearing in the FOV at the start of a visual search trial may not 'pop-out' when they appear in peripheral vision due to a head turn or eye movement. However, naturalistic environments present a complex salience landscape, whereas traditional visual search experiments tightly control the appearance of the visual scene. As such, pop-out may not be changed in this context, despite extensive FOV shifts.

In the following experiments, we investigate visual search beyond the initial FOV by asking participants to find a target stimulus presented within one of four search 'sets' presented in VR. The 'sets' are arranged such that two were presented in the periphery at the start of the trial (inner panels) but the remaining two are only viewable after participants make a large eye or a head movement (outer panels). Therefore, this is a controlled environment within which participants are instructed to freely move their head and eyes to find more or less salient targets as quickly as possible among distractors that are placed around them.

We report two experiments that differ in the number of distractor stimuli presented on each panel. The first experiment shows eight stimuli per set, while the second experiment presents a reduced set of two stimuli (see Figure III.1). Additionally, in Experiment 2, we introduce a condition in which the stimuli on the outer panels are hidden until the participant's gaze is directed almost directly at the panel. Both experiments aim to answer what the effect of salience of a peripheral stimulus on saccade and head movement planning during visual search in an extended FOV is.

Experiment 1

Method

Participants

35 participants (18–25 years old, M = 20.1, SD = 1.3, 17 female, 18 male) were recruited via a combination of personal approach and online advertising to gain course credit or a £5 reimbursement. 5 participants were excluded according to data quality criteria (see below in Data analysis), leaving 30, enough for 80% power and a medium effect size ($dz = 0.4$) for a 2 × 2 repeated measures ANOVA (27 participants required) (Brysbaert, 2019). All had normal or corrected-to-normal vision and gave written informed consent to participation on the study. The study adhered to the tenets of the Helsinki Declaration (2013). Experimental procedures were approved by the Royal Holloway, University of London Ethics Committee.

Equipment

Experiment 1 was conducted in the VR lab of the Psychology Department of Royal Holloway, University of London. Stimuli were presented on an HTC Vive headset retrofitted with Tobii eye trackers with a resolution of 1080 × 1200 pixels per eye, a frame rate of 90 Hz and a nominal FOV of 110°, which was connected to a Lenovo ThinkStation P510 desktop with an NVIDIA Quadro M5000 graphics card. The virtual display was created by drawing 2D 3840 × 1920 pixel images which were presented by Tobii Pro Lab version 1.130 software as a 360° image using equirectangular projection in the HTC Vive headset display. Two Vive Lighthouses 1.0 were used to track the head movements and the position of the headset in the room. Head and eye movements were recorded using Tobii Pro Lab software, gaze output frequency was 120 Hz. During typical usage, we expected eye tracking accuracy of 1.08° — 2.74° for the Vive Pro Eye (Schuetz & Fiehler, 2022).

Procedure & Stimuli

All participants were seated on a standard office swivel chair that could be rotated 360° and the headset was fitted. They were told they could swivel the chair round during the experiment. The cabling was laid out in such a way that the participant could move their heads freely and without resistance. Participants were given a practice session in which four trials were shown, one each O/T, inner and outer.

At the start of the experiment, after a 5 point eye tracker calibration, four crosses were shown, arranged in a square around the fixation cross location and the participant was asked to look at each in turn (in no particular order). This was partially to make

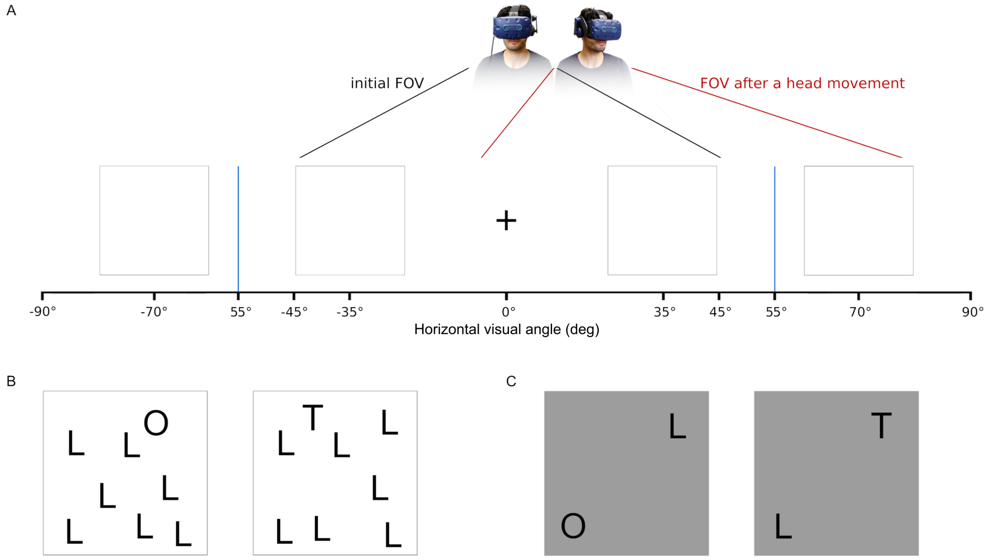


Figure III.1: Illustration of the VR search experiments with head movements. (**A**) Schematic illustration of the experiments. Participants were instructed to find a target, either a ‘more salient’ letter O or ’less salient’ letter T amongst distractors (L) that were arranged in four sets, two of which were visible in the initial FOV while the other two were outside of it. There was only one target in each trial. In the hidden set condition in Experiment 2, the letters in outer locations appeared after gaze passed the corresponding invisible threshold at 55° (blue line). (**B**) In Experiment 1, each set contained 8 letters. Sets were presented in a uniform grey environment, the bounding box around each set (19.6° x 19.6°) was not visible and 2D images were projected on a sphere using equirectangular projection. (**C**) In Experiment 2, each set was presented as a grey 2D panel (19.6° x 19.6°) with visible borders containing two letters spaced 4.9° and -4.9° from the centre of the panel. The sets were positioned in a uniform grey environment facing the participant.

sure the participant was facing in the correct direction, so that the fixation cross would appear in front of them, and was also used as a calibration check.

The virtual search environment was arranged in a fixed layout that was consistent throughout the whole experiment (see Figure III.1). A fixation cross was presented at 0° of horizontal and vertical angle straight ahead. At 35° (inner) and 70° (outer) left and right from the fixation cross were sets of letters, each containing 8 letters randomly spaced out within a defined area. Inner sets were visible peripherally from the start, outer sets could only be reached through a head movement. In each trial, participants had to start at the fixation cross and search as fast as possible for a more salient (O) or less salient (T) target in an environment with 31 distractors (L). They were asked to fixate it and press any button on the keypad in their hand. The keypad was not visually represented in the virtual environment. Participants were not aware whether the target would be an ’O’ or ’T’ on each trial. After the button response, the stimuli for that trial disappeared and only the central fixation cross remained on screen in the intervals between each trial for 3s. At the end of the experiment the four crosses appeared again with the same instruction as at the start. To further ensure eye tracking data quality, the experimenter remained present and monitored how the stimuli appeared to the participant, their real-time gaze location and the Tobii Pro Lab visual indicator of how well the eyes were being tracked.

Within one experimental run, the target appeared in each of the sets 10 times, yielding 2 (O/T) x 4 (set location) x 10 = 80 trials. This resulted in 20 measurements per condition (O/T and inner/outer). Five different full experimental stimuli were pre-generated (randomizing the locations of the letters within sets each time) and each participant saw one of these. The order of stimuli was randomly shuffled for each participant.

Data analysis

To stabilise an object on the retina, for example when the head catches up with an eye movement after a saccade, the eyes then move into the opposite direction of the head (Gresty, 1974; Morasso et al., 1973). Therefore, in contrast to head fixed experiments, a fixation can not be identified by just using the eye tracking data and a standard velocity threshold. Instead, the sum of eye and head movement (gaze) needs to be analysed.

The raw gaze data was filtered using the default Tobii Pro Lab I-VT attention filter with 60 ms minimum fixation duration and the velocity threshold parameter set to 100°/s. The reported time of a fixation was the first time point at which a gaze fixation was identified. Gaze-contingent stimulus presentation was not possible, so trials were excluded post hoc if the fixation at the start of the trial was further than 9.4° (100 pixels) away from the fixation cross. This arbitrary threshold was chosen based on inspection of trials containing obvious non-compliance with task instructions. Each fixation was assigned a ’set’ label according to the set the nearest letter belonged to (having moved more than 9.4° from the fixation). We use this label to report the time of the start and the distance from target of the first fixation within a set and the last fixation within a set. If none of the final three fixations before responding were in the set containing the target, or the participant never fixated within 9.4° of any letter, these trials were also excluded (as incorrect trials). Participants with more than 10% of trials excluded were fully excluded.

Results

Altogether from the 30 participants included, 3.6% of the trials were excluded, and we used 2313 trials for further analysis. On average, over all trials, there was a median search time (to final fixation on target set) of 1.36 s (1. Qu 0.83 3. Qu 2.22) and it took a median time of 2.09 s (1. Qu 1.50 3. Qu 2.87) until the button response. Median button response search times averaged across participants for the inner target locations were 1.48 s (more salient targets) and 2.11 (less salient targets). For the outer target locations, search times until button response were 2.52 s (more salient targets) and 2.77 s (less salient targets).

Based on final fixation time on the target set, targets in inner sets were found faster (median = 0.84 s) than targets in outer sets (median = 1.97 s) and more salient targets (median = 1.10 s) were found quicker than less salient targets (median = 1.84 s).

First Search Direction

As a first step, we analysed whether the direction of the first gaze movement went in the correct direction in all searches (Figure III.2 A). Only when a salient target was in an inner set were participants able to reliably initiate their first gaze direction change in the correct direction. When the target was in an inner set, a paired t-test revealed a significant difference between the more salient (M = 84.1 %) and less salient (M = 52.9 %) condition (t = 12.78, df = 29, p <.0001), while the confidence interval of the less salient condition overlapped chance level (50 %). As expected, in searches with outer targets (that were not initially in the FoV) the proportion of correct first gaze movements was at chance level and there was no significant difference between the two salience conditions (M = 50% and M = 47.2%) (t = 0.94, df = 29, p = .35).

Timing of Head-, Eye- and Gaze Movements

To be able to summarise search times from different trials in a sensible way, we sorted all trials based on the order in which participants looked at the different sets. Participants reliably searched in one of the inner sets first (98.0% of trials). If the target was not in the inner set, they then typically moved their gaze to the outer set on the same side (72.6%) (based on the location of the first gaze movement on a set). In 90 % of trials, the target was found after 4 sets were inspected. In 73% of trials, the target was found after 3 sets were inspected. In 61% of trials, the target was found after two sets were inspected. In 34% of trials, the target was found after 1 set was inspected.

We defined two common types of searches that we analysed further. The first category were search trials in which participants started their search in the correct direction and found the target in the first set they looked at (minimal search with target in the inner location). The second category included trials in which participants started their search in the correct direction, did not find a target at the inner set, and continued their search in the same direction to find a target there (minimal search with target in the outer location). In these trials we then defined the timing of the first saccade offset (i.e. fixation onset) on the inner set, the start of final fixation in the set containing the target, and the button response. For the minimal outer searches, we also defined the timing of the

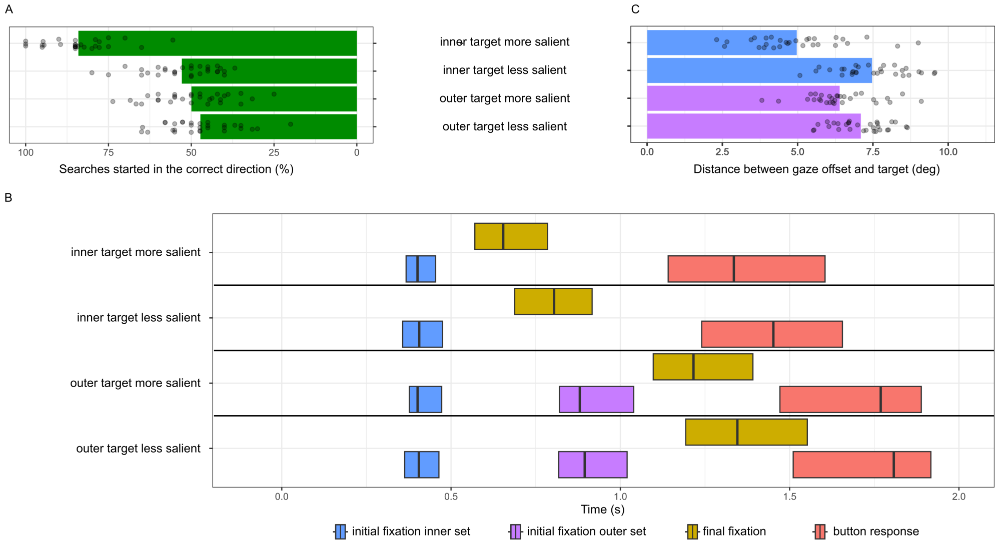


Figure III.2: Gaze during searching in experiment 1. (**A**) Percentage of trials in which participants started the search in the target direction based on the position of the first fixations in Experiment 1. (**B**) Averaged timing of key events in searches across all participants in Experiment 1. Vertical lines show median times, box thresholds indicate lower and upper quartiles. (**C**) Distance between the location of the initial fixations on the target set and the target in Experiment 1 (for all trials).

saccade offset on the outer set (i.e. beginning of first fixation on the outer set). The three gaze related timings were defined based on the fixation filter from the Tobii software (see Experiment 1 Methods — Data analysis).

Figure III.2 B shows the results of the gaze timing analysis for minimal searches. Since search times are naturally skewed, we used non-parametric tests to compare the individual means of the two salience conditions.

For the first category, participants initiated a gaze change with the same reaction time no matter the stimulus but took less time to complete their search to a salient target. A paired Wilcoxon signed rank test for the initial fixation time on the inner set showed no significant difference between the more salient (median = 0.4 s) and less salient (median = 0.41 s) condition ($W = 450$, $p = .99$). There was a significant difference between final fixations on the target set in trials with more salient (median = 0.65 s) than less salient (median = 0.8 s) targets in the inner set ($W = 240$, $p = 0.00161$).

For the minimal searches with an outer target, there was no effect of salience on search timing. A paired Wilcoxon signed rank test indicated no significant difference between the salient (median = 0.88 s) and less salient (median = 0.89 s) condition ($W = 444$, $p = .93$) for the initial fixation time on the outer set. There was also no significant difference between final fixations in trials with more salient (median = 1.22 s) and less salient (median = 1.35 s) targets ($W = 352$, $p = .15$).

Gaze Landing Position

As a final step, we analysed distances of the landing points of the gaze movements from the target of all trials from the experiment (Figure III.2 C). Participants were able to use the more salient stimulus to better guide gaze behaviour toward the target in both the inner and outer sets. When the target was in an inner set, a paired t-test revealed a significant difference between the more salient ($M = 5°$) and less salient ($M = 7.5°$) condition ($t = 7.64$, $df = 29$, $p < .0001$). In minimal searches with outer targets a paired t-test also indicated a small significant difference between the more salient ($M = 6.4°$) and less salient ($M = 7.1°$) condition ($t = 3.37$, $df = 29$, $p = 0.002127$).

Discussion

We find that for the inner targets, the results are as expected. In terms of search times, the more salient targets are detected from the fixation cross more often and are found more quickly. Analysis of landing positions and timing of the final fixation from minimal searches (where gaze moved in the correct initial direction) confirms that the gaze lands more accurately on target for more salient inner targets.

For outer targets however, although in the minimal searches (where gaze continues in the correct direction to outer targets) the gaze landing position for the more salient targets is still a little more accurate, a more salient target does not significantly decrease the time it takes to move the eyes towards the target set. Although it seems clear that peripheral information could influence gaze planning, we found no evidence for this in search time.

We considered whether our results might be due to the difficulty of finding either type of target, perhaps not allowing for the effects of salience to be fully seen. The lack of ability to extract real time fixation measures and provide feedback led to participants not always fixating accurately on target, meaning we had to use a measure of 'final fixation on target set' rather than 'final fixation on target', which is a less accurate measure of gaze based search time. These considerations led to Experiment 2.

Experiment 2

In Experiment 2, we reduced the number of distractors per set to make the peripheral target even more visible, potentially enhancing the effect of saliency. We also wanted to find out whether the effects of salience weakened when stimuli were not visible in the periphery after the first change in eye position. Therefore, we added a new *hidden* condition in which outer sets became visible after gaze crossed 55° on the corresponding side.

Method

Participants

Thirty-three participants took part in the study (18–35 years old, M = 22.3, SD = 3.6, 23 female, 10 male). Participants received student credits or 4€ for taking part in the experiment. All participants had normal or corrected-to-normal vision. They gave written informed consent to the participation on the study. The study adhered to the tenets of the Helsinki Declaration (2013). Experimental procedures were approved by the Ethics Committee of the Department of Psychology and Sports Science of the University of Münster.

Equipment

The experiment was conducted at the Department of Psychology and Sport Science at the University of Münster. Stimuli were presented on an HTC Vive Pro Eye headset with a resolution of 1440 × 1600 pixels per eye, a frame rate of 90 Hz and a nominal FOV of 110°, which was connected to an MSI GE63VR 7RF Raider laptop with an NVIDIA GTX 1070 graphics card. The virtual environment was created using Unity3D and the Unity Experiment Framework UXF (Brookes et al., 2020). Two Vive Lighthouses 2.0 and the gyroscope of the HTC Vive Pro Eye were used to track the head movements and the position of the headset in the room. During the entire experiment, head and eye tracking data were collected with a target frequency of 90 Hz.

Procedure & Stimuli

All participants were instructed to face towards a desk with a keyboard on it and sit in a steady, straight position at the beginning of each trial. They sat on a chair that could be rotated 360° and whose height they were allowed to adjust. The cabling was laid out in such a way that the participants could move their heads freely and without resistance. Experiment 2 was a re-implementation of the previous experiment as an interactive three-dimensional environment instead of a 360° image. Moreover, the custom software enabled more control over the head and eye tracking data to produce further insight into the search behaviour. At the start of the experiment, participants followed the eye tracking default calibration procedure implemented in the SRAnipal Software (version 1.3.6.8) of the head mounted display.

Again, participants were asked to search for a more salient (O) or less salient (T) target as fast as possible. The environment contained a total of 7 distractors (L) evenly distributed across 4 panels. Participants were asked to fixate the target and press the space bar of an external keyboard in front of them. The keyboard was not visually represented in the virtual environment, therefore participants were instructed to keep one hand on the space bar during the whole experiment. The visual environment was arranged similarly to the first experiment (Figure III.1 A). At the beginning of a search, a fixation cross was presented at 0° of horizontal angle and participants were asked to fixate it while keeping the head steady and facing forward. At 35° (inner) and 70° (outer) left and right from the fixation cross were two grey panels each. After the fixation cross was fixated for 0.5 s the trial started and the fixation cross disappeared. Now, each panel showed two letters in fixed positions (see Figure III.1 C). Again, the inner sets were visible peripherally from the start, the outer sets could only be seen after a head movement. In the *hidden set* condition, objects in the outer locations only became visible after their gaze passed a threshold of 55° on the correspondent side.

Each participant performed 320 trials in which an ’O’ or a ’T’ was presented at one of the 8 possible positions. Each type of trial was repeated 10 times. Target positions were balanced and presented in a pseudo-random order. Trials in which the participant did not gaze at the correct target when pressing the space bar were counted as error trials and were repeated at the end of the session. Errors could also result from participants not fixating the correct target or inaccurate eye tracking of a fixation of the correct target. As before, we expected an accuracy of 1.08° — 2.74° for the Vive Pro Eye during typical use (Schuetz & Fiehler, 2022), which was expected to be accurate enough to fixate the targets used in this study. However, because potential headset slippage could have resulted in invalid data, the eye tracker was automatically calibrated every time 10 error trials were reached before the next trial was started.

Data analysis

For Experiment 2 we used the raw eye movement and head movement data to define gaze movements and use the offset of these on the letter sets for the search timings. To get a valid gaze and gaze velocity signal, eye- and head tracking data need to be perfectly synchronised to each other. Otherwise, eye and head movements in opposite directions would result in a noisy gaze signal and overly large velocities during gaze fixations. Based on previous studies (Stein et al., 2021) and pilot studies of participants fixating a target while moving the head using the HTC Vive Pro Eye, we assumed that the eye tracking data measured with Vive Pro Eye headsets is delivered with some delay.

Therefore, we shifted the eye tracking data 4 frames (~44 ms) back and calculated gaze as the sum of the head- and shifted eye tracking data.

Results

Data were collected at an average frame rate of 89.3 Hz. Overall, participants had to repeat 14.76% error trials, because they did not fixate the correct target when pressing the space bar. These trials were excluded from further analysis, resulting in 320 trials per participant. In 9.53% of trials, participants moved their head faster than 15°/s at the start of the trial (-100ms to the start of the search). During the same period, in 5.49% of the trials, the head was not aligned with the centre (more than 10° away from 0°). In 9.83% of the trials, participants did not fixate the fixation cross stably enough at the start of the trial. 3.11% of trials included missing eye tracking data for more than 100ms at the start of the trial. Therefore, we excluded 22.085% of trials and used 8227 search trials for the further analysis.

On average, participants needed a median search time (to final fixation on target) of 0.94 s (1. Qu 0.53 s 3. Qu 1.76 s) and took a median time of 1.56 s (1. Qu 1.14 s 3. Qu 2.31 s) until the button response. Median search times averaged across participants for the inner more salient targets were 0.48 s (1. Qu 0.37 s 3. Qu 0.62 s) and 0.6 (1. Qu 0.46 s 3. Qu 0.83 s) for less salient targets. For the outer target locations, the median search time was 1.63 s (1. Qu 1.0 s 3. Qu 1.65 s) for more salient targets and 1.66 s (1. Qu 1.12 s 3. Qu 2.13 s) for less salient targets. In the hidden condition at outer target locations, the median search time was 1.75 s (1. Qu 1.12 s 3. Qu 2.21 s) for more salient targets and 1.67 s (1. Qu 1.16 s 3. Qu 2.18 s) for less salient targets.

Typically, the search was initiated with a synchronous head and eye movement. However, in some rare cases, participants made one saccade towards one of the inner sets at the start of the search while keeping the head still and then continued the search with a combined head- and eye movement to the opposite side.

Participants reliably searched on one of the inner sets first (91%). If the target was not in the inner set, they then typically moved their gaze to the outer set on the same side (61%). In 92 % of trials, the target was found after 4 sets were inspected. In 76% of trials, the target was found after 3 sets were inspected. In 64% of trials, the target was found after two sets were inspected. In 37% of trials, the target was found after 1 set was inspected.

First Search Direction

As a first step, we analysed whether the direction of the first gaze movement went in the correct direction in all searches. The first search direction was defined based on a 10° threshold around the fixation cross. When the target was in an inner set, a paired t-test revealed a significant difference between initially moving gaze in the correct di-

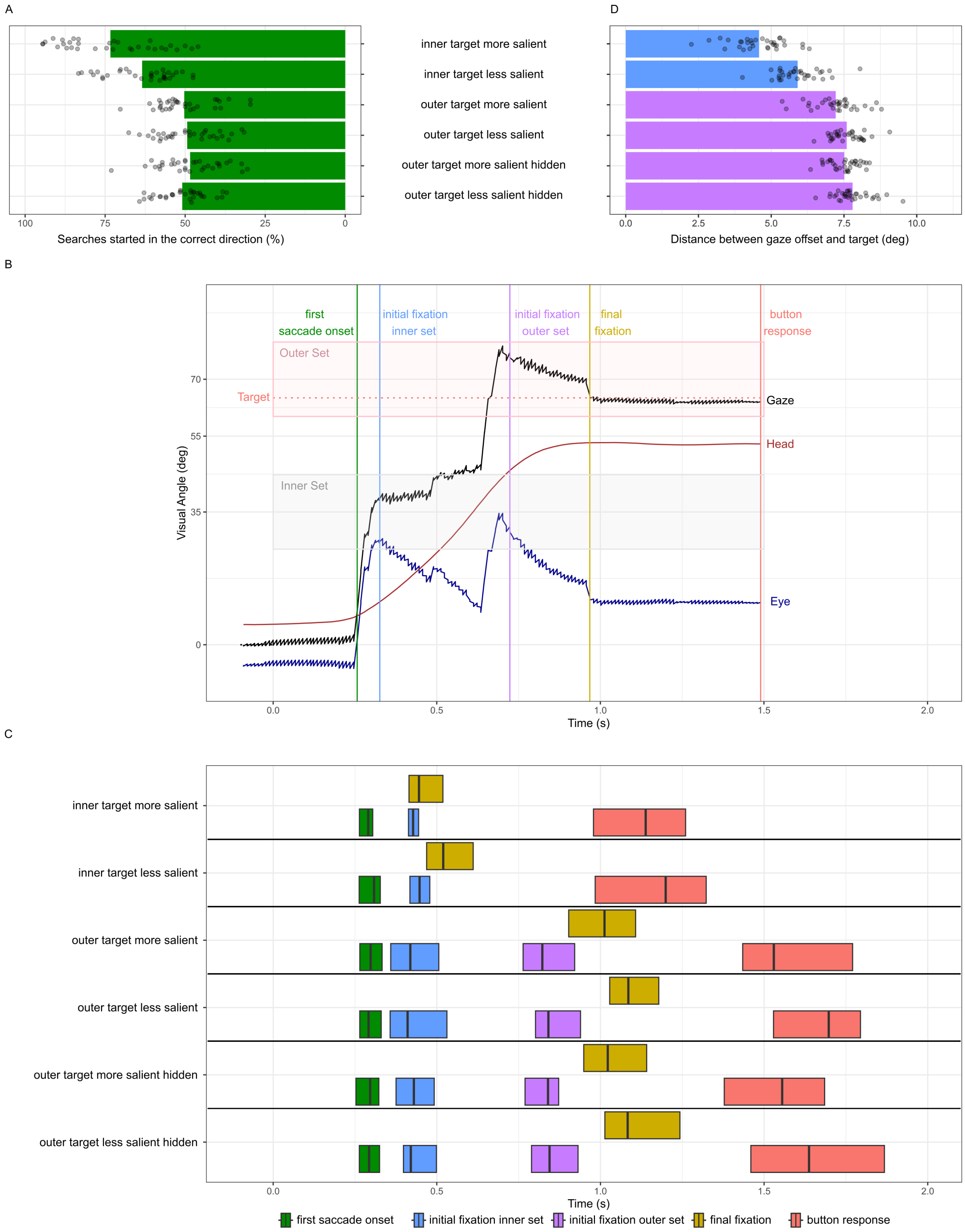


Figure III.3: Gaze during searching in experiment 2. (**A**) Percentage of trials in which participants started the search in the target direction in Experiment 2. (**B**) Sample trial of gaze-, head and eye movements during a minimal search for a salient target in an outer set. Key properties are represented as vertical lines. (**C**) Averaged timing of key events in minimal searches across all participants in Experiment 2. Vertical lines show median times, box thresholds indicate lower and upper quartiles. (**D**) Distance between initial fixations on the target set and the target position in Experiment 2 (for all trials).

rection in the more salient (*M* = 73.3 %) and less salient (*M* = 63.6 %) condition (*t* = 3.07, *df* = 56.362, *p* = 0.003), while neither confidence interval overlapped chance level (50 %). In searches with outer targets the proportion of correct first gaze movements was at chance level for all conditions as expected (means (confidence interval from one sample t-test, *df* = 32): more salient = 50.1% (46.8%, 53.3%), less salient = 49.5% (46.1%, 52.9%), more salient hidden = 48.5% (45%, 51.9%), less salient hidden = 50.9% (48.3%, 53.5%), all *p* >.05) and therefore there was no difference between any of the conditions. Figure III.2 A shows an overview.

Timing of Head-, Eye- and Gaze Movements

Again, we defined two common types of searches that we analysed further. In these, participants either inspected an inner set and found a target, or inspected an inner and outer set on the same side and found a target at the outer location (see Figure III.3 B for a sample trial).

Differentiating clearly between fixations, saccades and eyes that move opposite to an ongoing head movement in free headed experiments is challenging (Lappi, 2015). To detect fixations, every eye tracking sample was labelled based on collisions of gaze ray and objects in the virtual environment. We then calculated a gaze velocity signal and smoothed it with a Gaussian filter with a width of 5 samples. Fixations were then defined based on a velocity threshold of 50°/s. Only gaze sequences with the same label for more than 3 samples were kept as eye movement offsets/fixations.

Across all searches from the first and second category, the first gaze movement was detected at 0.3 s. The following fixation on the inner cluster was 33.4° away from the centre and was detected at on average 0.442 s. In minimal searches with an outer target, the participants' gaze then remained on the first inner set for a mean time of 0.656 s at an average horizontal position of 36.4°. Then, participants typically initiated a big gaze movement to reach the outer panel.

To compare trials from both categories, we defined the following key properties in each trial (see Figure III.3 B): 1) First saccade onset: the first sample in which the standard deviation of the horizontal eye angle on a moving window of 100 ms starting from trial start reached a threshold of 1°. 2) Initial fixation on inner panel: the first sample labelled as fixation in which gaze was on one of the inner target sets. 3) Initial fixation on outer panel: the first sample labelled as fixation in which gaze was on one of the outer target sets. 4) Final fixation: the first sample of the final fixation on the target object, before the button response was given. 5) Button response: the first sample with a button response while fixating the target.

Since search times are naturally skewed, we used non-parametric tests to compare these key properties. Participants were able to use the salience of the target, as seen from the periphery, to undertake a faster search. For the initial fixations on the inner set, there was a significant difference between the more salient (median = 0.43 s) and less salient (median = 0.44 s) condition ($W = 416$, $p = 0.014$). For inner searches a paired Wilcoxon signed rank test revealed also a significant difference between the timing of final fixations in trials with more salient (median = 0.45 s) and less salient (median = 0.52 s) targets in the inner set ($W = 532$, $p < 0.001$).

Participants were not able to speed up the time it took to undertake gaze changes to bring a salient peripheral target into the direct FoV. For the initial fixations on the outer set with outer targets, a Friedman test for paired rank sum group comparisons indicated no significant difference ($\chi^2 = 3.58$, $p = .31$, $df = 3$) between any of the conditions (medians: more salient = 0.82 s, less salient = 0.84, more salient hidden = 0.84 s, less salient hidden = 0.84 s).

Once gaze was directed at the set containing a target, the more salient target did reduce search time. For the final fixation in searches with outer targets a Friedman test for paired rank sum group comparisons indicated significant differences ($\chi^2 = 24.75$, $p < .001$, $df = 3$) between the conditions (medians: more salient = 1.01 s, less salient = 1.09 s, more salient hidden = 1.02, less salient hidden = 1.08). Bonferroni corrected post hoc paired pairwise Wilcoxon signed rank tests revealed significant differences between more salient vs less salient ($p <.001$).

For this to occur, the target did not need to be visible in the periphery prior to directing gaze toward the target containing set. Bonferroni corrected post hoc paired pairwise Wilcoxon signed rank tests revealed significant differences between more salient hidden vs less salient hidden ($p = 0.011$), less salient vs more salient hidden ($p = 0.019$) and more salient vs less salient hidden ($p <.001$).

Participants performed the same whether or not they had the opportunity to 'see' the target in the periphery while gazing at the inner set (see Figure III.3 C). There were no significant differences between more salient vs more salient hidden ($p = .16$) and less salient vs less salient hidden ($p = .99$).

Gaze Landing Position

As a final step, we analysed the distance of the landing points of the initial fixation from the target of all trials from the experiment. When the target was in an inner set, a paired t-test revealed a significant difference between the more salient (4.6°) and less salient (5.9°) condition ($t = 8.44$, $df = 32$, $p <.0001$). In searches with outer targets, a paired t-test also indicated a small difference between more salient (7.2°) and less salient (7.6°) condition ($t = 3.03$, $df = 32$, $p = 0.005$). When outer targets were hidden, there was no significant difference between more salient and less salient targets ($t = 1.93$, $df = 32$, $p = .06202$). Figure III.3 D shows an overview.

Therefore, the peripherally visible, more salient target was better able to guide accurate fixations within the set when it appeared under traditional search conditions. Like in Experiment 1, we also found a small increase of accuracy when the target was in an outer set. As expected, salience did not influence the gaze landing point within the outer set in the hidden set condition.

Discussion

The results are similar to Experiment 1. With a reduced number of stimuli, gaze moves initially in the correct direction a significant proportion of times even when there is a less salient target in an inner set in this case, but still an even higher proportion of times for salient targets. The gaze landing point within the set is also closer to the target for salient targets. Thus, participants were able to use the available peripheral information of the target to guide their initial search behaviour. Participants' gaze change behaviour also indicated that they were able to identify the more salient target faster when it is in an inner set than less salient targets. This is expected from typical visual search findings on 2D screens.

A timing difference between more salient and less salient targets was also observed for targets in the outer sets. This was the case whether or not the target was available to enter the FOV from the start of the trial or required gaze to be beyond 55° before the outer set's onset (*hidden set*). Despite an overall search time advantage for more salient targets, the gaze behaviour across both conditions showing targets in the outer sets confirm this advantage mostly emerges after the participants are directly inspecting the set containing the target. This means that salience does not contribute to a faster search time at the time the target set appears in the periphery during a search.

Like in Experiment 1, the participants' initial gaze did land closer to more salient targets in outer sets. However, when comparing the landing distance between the inner and outer sets, the overall accuracy was still worse for the outer sets, even though the number of stimuli was reduced compared to the previous experiment. This suggests again that peripheral information might have been available before an outer set was approached. To investigate whether this information was used by the participants for their search strategies, we set up an ideal observer model.

Ideal Observer Model

To analyse the participants' search strategy, we set up an ideal observer model that derives decisions based on two latent parameters. The first, γ, describes the proportion of times that a participant saccades in the direction of a target in the periphery when there is no information from the periphery to guide their decision. Thus, γ could be considered a guess rate, but it could also model a bias, if participants tend to prefer one direction over another, for example continuing on the same side for the second decision as for the first. Specifically, γ may vary across different target locations, due to the layout of the stimulus. The second parameter, σ, describes the proportion of time that the participant makes a saccade to a peripheral target when the target is present, over and above what could be expected based on γ. Thus, σ is a measure of the behavioural effects of the target, i.e., how often the target guides the decision. Salient targets should result in high σ values, as they make it more likely that a target is seen in the periphery. Less salient targets should lead to a smaller σ-value. The combination of σ and γ determines the percentage of correct decisions P_c:

$$P_c = \gamma + (100 - \gamma) * \frac{\sigma}{100}. \tag{III.1}$$

The two parameters can be estimated from the data of the first decision, i.e., when the participant sees the two inner panels, which may or may not contain a target. If the panels do not contain a target, σ is 0 and choice is driven by γ. Based on the setup of the stimuli, we expect γ at the start of the search to be $\approx 50\%$ as the two locations to chose from are equidistant, and the target is equally likely to occur either side. Indeed, in both experiments, when there was no target in the inner locations the proportion of correct first gaze movements was at chance (Experiment 1: $\gamma = 48.5\%$ (46.0%, 51.1%), for single conditions see Experiment 1: First Direction and Figure III.2 A; Experiment 2: $\gamma = 49.7\%$ (48.9%, 50.6%), for single conditions see Experiment 2: First Direction and Figure III.3 A).

Looking at the percentage correct (P_c) data from trials that did contain a target at an inner location, we can then calculate σ for the two salience conditions (more salient and less salient) by inverting Equation 1:

$$\sigma = \frac{P_c - \gamma}{100 - \gamma} * 100 \tag{III.2}$$

In Experiment 1, participants initially went to a location in the correct direction in P_c = 52.9% (48.8%, 56.9%) of trials when a less salient target was present in one of the inner locations. This equates to a value of σ = 5.5% (-4.9%, 15.9%). In trials with a more salient target in inner locations, participants chose the correct location in P_c = 84.1% of trials (σ = 67.9% (59.8%, 75.9%)). Therefore, they were able to use the peripheral information at the start of the trial when a more salient target was present.

In Experiment 2, participants initially went in the correct direction in P_c = 63.6% (60%, 67.2%) of trials when a less salient target was present, which results in σ = 30.6% (23.7%, 37.6%). When a more salient target was present, participants initially went in the correct direction in P_c = 73.3% (67.9%,78.7%) of trials, which results in σ = 49.1% (38.8%, 59.4%). Therefore, they were (as expected) able to use the peripheral information at the start of the trial in both salience conditions.

We can now use the ideal observer model to consider the next step of a search, when the target is not present in the inner locations. We consider what should happen after the gaze has moved to one of the inner locations in the correct direction, i.e. the target is now present in the peripheral outer location, for example the target is in a set at 74.9° and the observer has moved their gaze to 35°. This, hence, is the second decision to be taken by the participant. Here, the value of γ is less clear a priori, as the layout is less balanced. For example, the observer might have a bias to carry on towards the outer location rather than changing direction. This would be reflected in a biased value of γ for the second decision. We can estimate γ for the second decision from the data in the hidden condition in Experiment 2, in which no information about the outer target was present before the saccade to the outer panel was initiated. In this condition, participants on average continued their search in the same direction (i.e. with a saccade to the outer panel) in γ = 73.3% (69.7%, 76.9%) of trials (Figure III.4). Hence, our participants were strongly biased towards continuing their search in the same direction as in the first decision.

How much should target salience then affect choice probability in the second decision? If salience is used in the same way as in the first decision at the start, we can use the values for σ estimated from the first decision to calculate an expected percentage correct for the second decision from Equation III.1. Using the previous σ values of each participant to predict P_c, the model predicts an average value of 84.7% (80.4%, 89%) in the more salient condition and 80.6% (76.1%, 85%) in the less salient condition. These predictions are clearly at odds with the data: the measured P_c in Experiment 2 remain very similar to the 73.3% mark both for more salient (73.9% (68.3%, 79.4%)) and less salient (71.8% (66.5%, 77%)) targets present in the periphery. Thus, our observers did not follow an ideal observer strategy, or they could not use saliency in the second decision in the same way as in the first decision (for an overview of all decisions and all search times in all conditions see Figure III.5).

Experiment 1 did not include a hidden condition, and we cannot compute γ for the second decision in the same way as in Experiment 2. However, we may use data from trials in which no target was present on the corresponding side (because the first decision was wrong) to estimate γ for the second decision and then predict P_c based on the σ-values obtained from the first decision. When no target was present in the outer location, participants continued on the same side in 71.2% (62.6%, 79.8%) of trials. This is close to the γ calculated from the hidden condition in Experiment 2 (73.3%). Then, from the

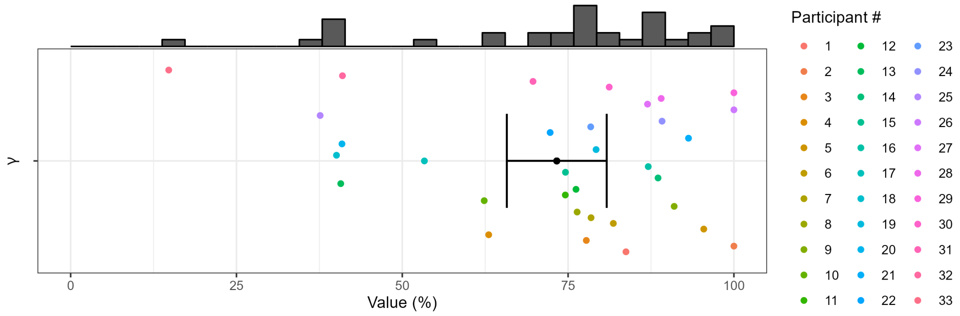


Figure III.4: Distribution of γ of the second decision in experiment 2. Individual values of γ were estimated using the results from the hidden condition. They represent how often a participant continued their search on the same side as in the first decision when no information about targets in the outer set was available. The black dot represents the mean across participants, the black error bar represents the confidence interval.

σ estimated in the first decision on the more salient target (67.9%) the model predicts P_c = 89.8% (83.3%, 96.2%) for the second decision with the more salient target. As in Experiment 2 these predictions are not consistent with the observed data (P_c= 77.7% (69.1%, 86.3%)). For the less salient target, σ estimated in the first decision was already very small (5.5%), indicating little effect of salience already at the start of the search in this experiment. Consequently, the model prediction for P_c was determined mostly by γ. The model predicted P_c = 73.4% (64.1%, 82.7%), which is both in the range of γ (71.2%) and in the range of the observed data (P_c = 77.8% (68.6%, 87%)). In summary, therefore, the values for σ obtained from the first decisions in either experiment provided poor predictions for P_c of the second decision, indicating little effect of salience when the target was in an outer set.

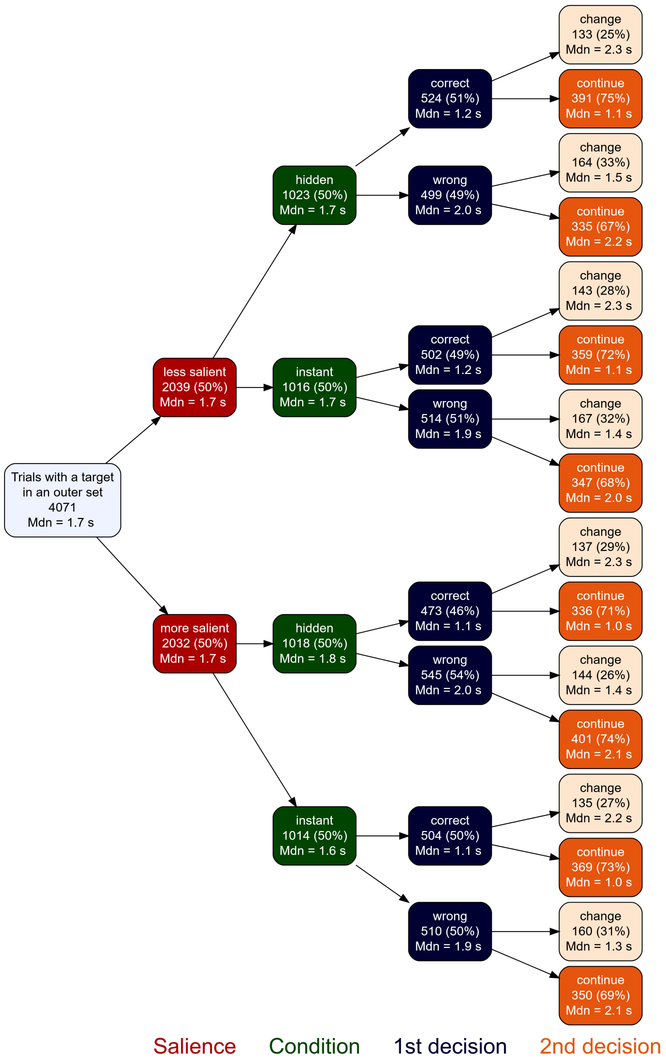


Figure III.5: Distribution of decisions during searching. Trials with a target on an outer set, including median search times (Mdn) and the absolute and relative amount of trials in each condition and decision. While the first decision (blue) was made at random, on the second decision (orange) participants preferred to continue their search on the same side. However, in 25 — 33% of trials with a target in an outer set, they switched sides after having previously inspected only an inner set.

General Discussion

We tested whether there was a difference in visual search for stimuli that became visible in the periphery by appearing at the start of a trial versus those locations that only became visible in the periphery as the participants moved their eyes and heads to inspect the scene. We designed two experiments in which participants were required to choose between and inspect up to four sets of letters across different locations, to detect a target. These sets were arranged such that two were visible to participants when the trial began, whereas two required participants to move their head to bring them into view. We manipulated the salience of the targets, the number of distractors and how the stimuli become visible to investigate the way in which the location and salience of peripheral targets influences the planning of search strategy and the associated landing positions. Analysis of gaze dynamics highlights key differences in the effect of salience on visual search when targets were in the initial FOV versus when they appear in the FOV due to head and eye movements.

When targets appeared in the FOV at the start of the trial, both experiments showed faster identification of the more salient target. Participants were requested (Experiment 1) or required (Experiment 2) to fixate the target when responding, which involved making a large change of gaze direction to complete the task. Accordingly, both experiments showed participants reliably made an initial gaze movement in the direction of the salient target. When there were two stimuli per set, participants still reliably made an initial gaze direction change in the direction of the less salient target. With 8 stimuli per set, participants were unable to reliably make their initial gaze in the correct direction of the less salient target. The ability to better identify the more salient target from the periphery was accompanied by a gaze movement that landed closer to the target.

We anticipated that a stimulus capable of supporting 'pop out'-like search, as found when the target was in the inner sets, would also facilitate faster target detection when the stimulus appeared in the FOV on the basis of the participant's own head and eye movements as part of the ongoing search task. We expected participants would first fixate an inner set and from there a more salient target would allow participants to accurately and immediately continue on to the outer target set, facilitating faster target fixation in these trials. This was not the case. Both experiments showed that it took equally long for the participant's gaze to arrive at the correct outer set no matter the target saliency, even when we reduced the stimuli to two per set. Experiment 2 also introduced a *hidden set* condition whereby the outer sets only appeared after gaze direction crossed 55° of visual angle, to further test if peripheral information had any influence on gaze movements. This was not the case. There was no difference in search performance whether the outer sets were available in the environment throughout the whole trial or appeared under the hidden set condition, suggesting that the process to reach the target cluster differed between inner and outer sets.

In addition, we found that participants' gaze lands closer to the target on salient trials than on less salient trials (Experiment 1 = -2.5°, Experiment 2 = -1.3°). However, this difference becomes smaller for the outer target sets (Experiment 1 = -0.7°, Experiment 2 = -0.4°), while the absolute distance between initial gaze landing point and target is also higher in these sets. This also indicates that the participants approach sets that are brought into the FOV during an ongoing search differently.

The more refined gaze direction analysis in Experiment 2 showed that once participants had fixated within the correct outer set, more salient stimuli were responded to faster, which is in line with previous studies on visual search without head movements in a fixed FOV. Thus, our results show a difference between visual search when stimuli appear in the periphery as in a traditional visual search experiment, relative to when they are brought into the FOV by a head movement as part of the ongoing search task. In other words, a highly salient stimulus shown to 'pop out' from the periphery under the same conditions as typical visual search, does not show this property when it enters the FOV due to the participants' own head and eye movements as part of an ongoing search task. The fact that there is no difference between the two outer target conditions in Experiment 2 (hidden vs always visible), further confirms that peripheral information was not used in this experiment during the search task after the initial gaze change has occurred.

There are two possible explanations for participants apparently not using the peripheral information about the stimulus to reach the target faster after they have started to actively look around the stimulus sets. The first is that they are unable to use the peripheral visual information in the same way as they could at the start of the trial. The second is that participants were able to gather the presence of the more salient stimulus from the periphery but did not use this information to guide their search behaviour.

Participants may have been unable to see the peripheral stimuli when they entered the FOV because they were actively shifting their gaze at this time. For example, it could be that the gaze on the inner set was never stable enough or at the equivalent distance from the target-containing set relative to fixation. When examining the more detailed eye tracking data in Experiment 2, we find that participants look at an inner set for a sufficiently long time and at a sufficient proximity to the outer set. Figure III.3 B exemplifies the typical gaze on the inner set prior to moving to the outer set containing the target. It also shows a pattern we observed that suggests that despite gazing at the inner set for a period, the participant had decided to inspect the outer set at the point of making the first gaze change. The example trial shows one large head movement, with the eyes acting to stabilise the gaze on the two sets around the single large head movement. This behaviour was common across participants and indicates participants may have decided on a course of action already after the initial view of the stimuli.

Although saliency alone has been shown not to predict gaze direction when freely viewing 360° environments in VR (Xu et al., 2018), in our experiments the salient target was relevant to the task and the visual scene was controlled as in traditional visual search experiments. Therefore, we expected the reduced visual search environment and the task relevance of the salient stimulus would allow salience to guide search. This did not occur consistently across stimulus locations, suggesting that salience even when relevant to the task is not the key predictor of gaze direction when active exploration is occurring around a scene with strong spatial layout cues. In studies with more natural environments, other factors like memory (Li et al., 2018) and scene context (Boettcher et al., 2018; Lukashova-Sanz et al., 2022) have been shown to influence searching for objects in a simulated real world context in VR. Other research has also suggested that encoding of stimulus salience and its utility in driving gaze behaviour in a naturalistic scene is time-limited and diminishes throughout a trial (Anderson et al., 2015; Donk & van Zoest, 2008; Schütt et al., 2019). In addition, it has been shown that saccades with a short latency are guided by salience, while saccades with a long latency are more likely to take value information into account (Schütz et al., 2012). Visual search in the extended FOV could therefore be influenced by the use of strategies. However, these search patterns could also make use of salience.

To find out whether target salience affects search strategy, we then analysed two key decisions during each search trial and compared them to an ideal observer model: first, the initial search direction and, second, the decision whether to keep on searching on the initially taken side vs changing sides. The ideal observer model was based on two parameters: γ, an indicator of uninformed choice bias and σ an indicator of choice based on target salience. The results indicated that σ influenced the first decision as expected, providing a measure of the salience of each target. However, σ had little predictive power for the second decision, indicating that the second decision was not affected by salience. Instead, the second decision was dominated by γ, a bias to continue search on the same side in 73% of trials, no matter whether a salient target appeared in the periphery or not.

This indicates that the participants followed a predetermined strategy in the second decision, rather than using peripheral vision to guide their search. If no peripheral information is gathered for the second condition, an ideal observer should decide between continuing on the same side and changing sides based on minimal effort and time loss. We can estimate effort based on the median search times of Experiment 2. The search time advantage of continuing the search on the same side, vs changing sides after inspecting the first set, was 0.18 s. Searches with an outer target in which the search was continued at the second decision took a median search time of 1.61 s, searches with an outer target in which the search direction was changed at the second decision took a median search time of 1.79 s. Shifting gaze from an inner panel to an outer panel (35°) took on average 0.17 s while shifting gaze from one inner panel to the other inner panel (70°) took on average 0.22 s. Hence, continuing on the same side leads to a slightly faster search on average. Therefore, an ideal observer should always follow this strategy when no peripheral information is used. However, our participants continued to search on the same side only 73% of the time. This is not ideal behaviour. Yet, similar patterns, following non-ideal strategies in two- and threefold decisions have been described in humans (Gardner, 1957; Lo et al., 2021) and animals (Bullock & Bitterman, 1962; Saldana et al., 2022) before. While an ideal observer would always minimise the effort, distance and trial time and should therefore choose the ideal move in every trial, humans often deviate from this behaviour and follow a strategy that can be described as probability matching. In pattern matching, participants show a discrete behavioural pattern that asymptotes at the probability of success, instead of always picking the most probable choice. The occurrence of pattern matching is quite common and has shown to be influenced by the effort needed to implement a strategy (Schulze & Newell, 2016), reward (increased task reward makes pattern matching less likely) (Siegel & Goldstein, 1959) and task expectations (for example setting a local focus to look at each trial separately makes the occurrence of pattern matching less likely) (James & Koehler, 2011). Cognitive load during the decision does not change the likelihood that participants use pattern matching (Otto et al., 2011; Schulze et al., 2019). After numerous trials, participants typically overshoot the initially observed probability of success (Vulkan, 2000). Therefore, pattern matching can also be described as part of a learning trajectory that could at some point lead to the ideal strategy of maximizing the ideal choice and therefore going for the best option in every trial (Montag, 2021).

Looking at the distribution of γ in Figure III.4, we can see that although three participants reached an individual γ-value of 100% (and therefore always continued on the same side at the second decision) most participants show individual γ-values around 75%. In our experiments in particular, once the participant has determined that the target is not in any of the inner sets, the distance to the set on the opposite side is three times larger than the distance to the set on the same side (105° vs 35°). Therefore, a strategy that matches behaviour to the expected upcoming gaze movement distance could result in 75% staying on the same side vs 25% switching sides.

Overall, our results indicate that saliency plays a subordinate role in visual search in the extended FOV. In our two experiments, participants selected where to look based on information from the periphery only at the beginning of the search. All subsequent directional decisions appeared to be selected largely depending on the known arrangement of the stimuli. It is possible that the structure of the environment, alongside stored memory content, knowledge about the current task, the expected reward and cues from the environment are the more important information for the ongoing active exploration of our environment. This could also explain why salient objects are not always found at first sight when searching for them in the real world. However, this does not mean that salience plays no role at all. Once gaze is directed to a set through head movements, the mechanisms known from classic visual search paradigms take effect again and salient stimuli facilitate the last part of the search.

The use of VR technology has enabled us to bring experimental control to a less constrained FoV, however we cannot say for certain that viewing behaviour in this virtual space is not changed by the constraints imposed by the smaller FOV or the weight of the headset, for example. We also observed interesting differences between Experiment 1 that did not use a gaze-contingent design and Experiment 2 that measured whether participants were looking at the target when they responded via the keyboard. Experiment 1 presented a challenge for data analysis, as we observed participants were often eager to return their gaze to the fixation cross before responding, but were likely unaware of this. More detailed gaze data analysis in Experiment 2 was facilitated by the gaze contingency, however this controlled behaviour is a departure from the more naturalistic behaviour found in Experiment 1. While we feel these differences do not interfere with the findings in this research, it is important to keep in mind that fully virtual worlds introduce the necessity to explicitly make the trade-off between producing data structured via experimental control such that complex tasks can be aligned for analysis and allowing the participant to interact naturally with the environment. Our final observation is the importance of awareness of the temporal shift between the different sensors available in commercial headsets. The constraints on each sensor used to track a person in virtual space are that they produce a good enough experience for the user. Our experiences show that this may not be good enough for research purposes and temporal alignment cannot be taken at face value. With these caveats in mind, we feel there is great value in taking well established visual and cognitive tasks off two-dimensional screens and enacting them in virtual three-dimensional or immersive situations.

In summary, using VR to measure visual search in an extended FOV we found that salient targets do not produce ’pop-out like’ search when they enter the FOV after participants have started to move their gaze around the scene. This suggests that visual search using a traditional visual search array does not occur in the same way when active exploration outside the initial FOV is necessary. Instead, participants follow a pre-planned gaze sequence, likely informed by knowledge of the layout of the scene.

Data Availability

All data is available in an open science framework repository https://osf.io/ewsj5/

Acknowledgements

This work has received funding from the European Union’s Horizon Europe research and innovation programme under the Marie Skodowska-Curie grant agreement No. 101086206. Thanks to Doga Gulhan, Lillian Hawkins and Maia Passenda-Crichlow with help in running the experiments at Royal Holloway, University of London.

Author Contributions Statement

All authors conceived the experiments, N.S. analysed the results, N.S., T.W., M.L. and S.D. wrote and reviewed the manuscript, M.W. and S.D conducted the experiments.

Competing Interests

The authors declare no competing interests.

General Discussion

Today, HMDs have become powerful devices with impressive tracking capabilities, and to a certain extent, Sutherland's ideas of seamlessly experiencing and walking through virtual worlds have become reality. To top it off, HMDs with built-in eye tracking sensors enable VEs and user interfaces that automatically react to the user's eye movements. As a result, tracked eye movements can be transferred to virtual agents and VR buttons can be selected using fixations.

The research results presented in this paper show that we have reached a point where we can use VR as a research method to expand our knowledge of visual perception. In particular, when investigating the connections between vision and locomotion, VR eye tracking offers great potential for gaining a better understanding of how we perceive and explore the world around us. The investigation into how VR can be used to observe eye movements during natural behaviours was driven by two goals. We wanted to gain more insights into typical gaze patterns to increase our knowledge of human visual perception and hoped that the knowledge of these patterns might be used for VR applications.

As a first step, we developed and utilised a method for measuring the latency of eye tracking in HMDs. Our results make it possible to precisely match the eye tracking signal to the stimuli displayed on the screen. Our results also enable a more informed assessment of whether a specific VR setup can be used for gaze-contingent experiments. We followed this with Study II, in which gaze, movement and orientation data were recorded during several natural tasks. Subsequently, we tested how well future waypoints could be predicted based on this information. Here, we were particularly interested in whether eye tracking data improves prediction accuracy. In Study III, eye and head movements during visual search tasks in VR were recorded, to get a broader understanding of how our gaze is guided in this common scenario.

The following general discussion first examines the current methodological limitations of VR eye tracking and provides an outlook on possible directions for the future development of the method. Next, the prediction of our movement based on gaze data is revisited and new proposals for improving the method are summarised. In addition, the influence of salient objects that may obscure the gaze while walking on the prediction of movement is discussed. The discussion then reflects upon the findings on gaze behaviour during searching, specifically with regard to the factors that influence the interaction between bottom-up and top-down guidance of coordinated head and eye movements. This is followed by a summary of the current limitations of VR that should be taken into account when designing future psychophysical experiments, as well as a brief outlook on the future of the method. Finally, the discussion concludes by reflecting on the most important features of this work.

VR Eye Tracking Latency

To transfer human behaviour to virtual worlds, HMDs record data from a wide range of sensors. This means that the VR equipment simultaneously captures multiple facets of user behaviour in great detail, providing a range of opportunities to observe human behaviour. When recording eye movements for psychophysical studies, it is important to have accurate information about the latency of the measuring devices in order to be able to associate the tracked movements with the stimuli displayed in the VE. Moreover, experiments with gaze-contingent stimuli, require latencies that are small enough to go unnoticed by participants when moving their eyes across the screen.

In Study I, we developed a method that can be used to accurately measure the eye tracking delay and end-to-end latency in VR headsets with included eye tracking. By simultaneously recording eye movement data based on VOG, EOG and the light emitted by the HMD using a photodiode, we measured the latency of the sensors of different headsets. We measured eye tracking delay (time from the occurrence of an eye movement to the availability of the corresponding data) and latency (time from an eye movement to a corresponding change in the screen content). Delays ranged from 15 to 52 ms, latency ranged from 45 to 81 ms. Compared to the Varjo VR-1 and the HTC Vive Pro Eye, the Fove-0 was the fastest device and seemed best suited for time-critical psychophysical research.

The end-to-end latency (from an eye movement to a visible change on the screen) is the result of a chain of different processes (see Figure E). Initially, latency can be caused by the camera directly when the eyes are recorded. Then, the eye tracking algorithm that interprets the recorded image can add further delays (especially if temporal filters are applied to smoothen the signal by averaging the direction of gaze over several samples). The process of reading in the eye movement data can also lead to further time loss (e.g. due to the limited bandwidth of the connection). The time required to perform all these steps results in a specific eye tracking frequency. This frequency indicates how often this series of steps can be performed per second. At the same time, the virtual image is created at a certain rendering frequency. If both frequencies are not perfectly synchronised, this can lead to additional end-to-end latency. The total latency can also be increased by the rendering itself (e.g. when particularly complex scenes have to be generated, which slows down the time required to create a frame and thus reduces the frame rate). Finally, latency can increase by a few milliseconds due to the latency of the screen when displaying the image. Interestingly, in our comparison the Fove-0s display had the lowest refresh-rate (70 Hz vs 90 Hz in the Varjo VR-1 and the Vive Pro Eye), which should translate into a slightly shorter display latency of about 3 ms for both the Vive Pro Eye and the Varjo VR-1.

With this in mind, our results contribute to a better understanding of which step in this chain causes how many frames of latency. It is therefore possible to combine our latency measurements with the results of other studies to draw further conclusions. In an earlier study with the HTC Vive (the predecessor model of the Vive Pro Eye, which also uses an OLED display and the same position tracking system but is not equipped with eye tracking sensors) Niehorster et al. (2017) measured the end-to-end latency between the physical movement of the HMD and a corresponding change on the display. They found a latency of only 22 ms or 2 frames at 90 Hz. Combined with our measurements of the latency between receiving gaze data and updating gaze dependent screen content (see Table I.1), this suggests that motion tracking can be completed in under 1 frame, while rendering and displaying a frame based on this data can take up to 2 frames. The remaining latency is therefore probably caused by the eye tracking itself. Like us, Sipatchin et al. (2021) also measured the time lag of the Vive Pro Eye. In contrast to our experiment, they used an infrared light source that was used as an artificial eye instead of real eye movements. With this setup, they found an end-to-end latency of 5 to 6 frames at 90 Hz. In our study we measured an eye tracking delay of 4 to 5 frames from saccade onset to recording gaze data and an end-to-end latency of 7 frames at 90 Hz from saccade onset to measured changes on the display. As a result, it seems that recording an image of the users eyes, detecting the pupil and transforming these measures into eye tracking data, takes multiple frames. A direct comparison between the two studies shows that this latency varies depending on whether a real eye is tracked or a pupil reflection is simulated (see Figure E).

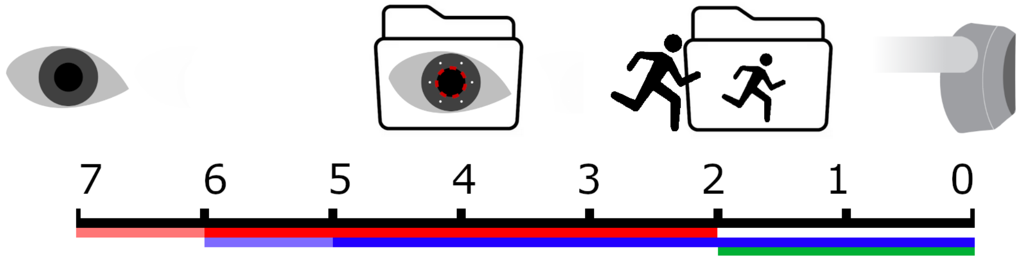


Figure E: Components of end-to-end latency in VR eye tracking in the HTC Vive Pro Eye. Latency from an eye movement to display change (black) takes 7 frames, while eye tracking delay from eye movement to data collection (red) takes 4 -5 frames (Stein et al., 2021). End-to-end latency when using an artificial eye (blue) led to a latency of 5 – 6 frames (Sipatchin et al., 2021). Latency from moving the HMD to a change on the display (green) takes 2 frames (Niehorster et al., 2017).

In a realistic gaze-dependent experiment, detecting the onset of saccades, could increase the overall latency even further. However, in the latency measurement studies presented here, each eye movement could be detected from one sample to the next. End-to-end latency is too high for gaze contingent rendering, if the user perceives stimuli that have not yet been adjusted to the current eye position after a saccade. To determine the maximum latency for a gaze-contingent setup, only a few empirical studies are available. Loschky and Wolverton (2007) found that user performance in a blur detection task did not change until up to 60 ms latency. In addition, Loschky and McConkie (2000) reported that users fixated slightly longer at a latency of 45 ms compared to a latency of 15 ms. Other recommendations for maximum end-to-end latency are based on unpublished results and expert opinions. Overall, the maximum latency for GCDs recommended in previous articles varies between 20 and 80 ms (Albert et al., 2017; Carmack, 2013; Loschky & McConkie, 2000; Loschky & Wolverton, 2007; Melnyk et al., 2025). As a result, a clearly defined maximum end-to-end latency is not available. At the same time this range of recommended values fits well to the expected length of saccadic suppression of saccades with typical amplitudes (Stevenson et al., 1986; Volkmann, 1986).

Our results suggest that among the HMDs we tested, the Fove-0 is best suited to present gaze-contingent stimuli in VR without the user perceiving this manipulation. To be able to measure gaze-contingent psychophysical experiments in VR, however, the exact parameters, such as a detection threshold as a function of saccade amplitude and parameters of stimulus manipulation detectability such as contrast, or degradation gradient would first have to be systematically determined for the specific task. Furthermore, even if previous measurements suggest that head movements can be detected relatively quickly, the latency of head position and orientation tracking also needs to be taken into account in order to display gaze-contingent stimuli in VR.

Eye tracking devices mainly used for research, such as the Eyelink 1000, achieve low latency times through a combination of very high recording frequencies, efficient algorithms for analysing the eye tracking camera images and fast forwarding of the eye tracking data via network interfaces. In VR eye tracking, recording frequency has also increased in some more recent devices, such as the Fove-480 (480 Hz) and the Varjo XR-4 (200 Hz). In both HMDs the recorded images of the eyes are forwarded to the computer, where they are analysed to generate gaze direction data. On the one hand, the improved temporal resolution can be used to recognise saccades earlier and better display their trajectory, on the other, the precision of eye tracking can be improved by applying a temporal filter to the eye tracking signal (Holmqvist et al., 2011).

At the same time, there are new approaches to further develop the recording frequency in mobile eye tracking using a combination of several recording methods. In addition to good data quality and new algorithms for merging data from different sources, energy consumption, waste heat and the weight of the additional sensors also play a role. For example, experiments with a combination of VOG and photo sensor oculography, which were combined with the help of machine learning algorithms, showed interesting results that could lead to VR eye tracking hardware with higher temporal resolution in the future (Katrychuk et al., 2019; Palmero et al., 2023; Rigas et al., 2017): As with limbal eye trackers, the photo sensor provides a measurement of the location of the white sclera of the eye, which reflects more infrared light than the iris. With today's sensors, this data has high spatial and temporal resolution. In addition, it is possible to only forward data from sensor input that changed since the previous sample. This event-based approach reduces the amount of data so that transmission can be faster. This data stream is then fused with the lower-resolution video recordings to create a data set with a high temporal and spatial resolution.

Another area of development is new algorithms that can predict future eye movements. Komogortsev et al. (2009) predicted saccade amplitudes based on the first two gaze samples using an eye tracker with 120 Hz. Although their model was good enough to predict the saccade direction with a high degree of certainty, it was not able to predict the exact landing position, as it gave an average prediction error of about 5°. Furthermore, Arabadzhiyska et al. (2017) predicted the saccade landing point during an eye movement using the data recorded at the start of the saccade with 300 Hz. Their prediction results in an average error of less than 4° if the prediction is made on the basis of data up to the middle of the saccade. At 80 % of the saccade, the error even drops to less than 1°. Interestingly, the best prediction according to their model is based exclusively on the last data point of the saccade. Based on their prediction, Arabadzhiyska et al. (2017) were able to reduce the detection rate in an experiment with foveated rendering.

One other approach is to use machine learning models trained on large eye movement datasets to predict the onset of a saccade based on a series of recent incoming eye movement samples (Rolff, Stein, et al., 2022; Rolff, Steinicke, & Frintrop, 2022; Rolff et al., 2023). So far, however, the models predict saccades with an average error of more than 0.12 s and are therefore not suitable for replacing conventional saccade detection algorithms. In the future, however, it may be possible to adaptively adjust the saccade threshold based on such predictions in order to reduce the time required for an eye tracking system to detect eye movement.

Since the publication of our study, in addition to updated eye tracking headsets connected to a computer, a number of stand-alone HMDs with inside-out positional tracking, controller and hand tracking capabilities and eye trackers have come to market (for example the HoloLens 2, the Apple Vision Pro, the Neo 2 Eye, Neo 3 Pro and Neo 4 Enterprise from Pico, the Varjo Aero, the HTC Vive Focus 3 and the Meta Quest Pro). For updated HMDs connected to a computer, such as the Fove-480 or the Varjo XR-4, our method of measuring latency by simultaneously recording eye movements using EOG can still be used. To measure the latency of stand-alone devices, however, a different synchronisation method is required to determine the latency between VOG and EOG.

Until today, researchers favoured the Quest Pro (Aziz et al., 2024; Wei et al., 2023). The Vive Pro Eye was also widely used to record fixation positions in many different scenarios (Moreno-Arjonilla et al., 2024). In contrast to the Fove-0, both headsets offer support for the use of controllers, additional trackers, large-area position tracking and a wireless mode. Even though end-to-end latency is comparably high, accuracy and precision of the eye tracking sensors in the Vive Pro Eye appear to be good enough to record fixation points in VR, at least in the central FOV (Schuetz & Fiehler, 2022; Sipatchin et al., 2021).

Within this context, our research showed that the latency of VR eye trackers can be reliably determined using a direct comparison with EOG. We found that the end-to-end latency between an eye movement and a display change varies depending on the device. In some cases, the latency even appears to be longer than the expected duration of saccade suppression for saccades with typical amplitudes. When conducting a gaze-dependent experiment with such latency, participants would see the visual stimuli jump from one position to another after each saccade. In principle, the latency measurement method we developed can also be applied to newly introduced devices, although a new type of signal synchronisation is required to determine the latency of stand-alone HMDs. It may therefore be possible to reduce latency by filtering gaze data less or by detecting pupils and reflections more quickly. Nevertheless, VR eye tracking can currently be used to transfer our eye movements to avatars, record fixation points in virtual space and even track travel-gaze-fixations while a user is walking. Moreover, VR eye tracking can be used for foveated rendering. In such cases, only the central part of the current FOV is rendered in high detail, allowing virtual worlds to be simulated more efficiently. However, as mentioned before, the latency of some current HMDs is in a range where it is seems likely that users can perceive that the method is applied. To reduce the detection rate, the resolution degradation factor can be carefully adjusted to increase the foveated area. In addition, algorithms for predicting the start or progression of the saccade can help to make the detection of the manipulation less likely.

The results from our study provide valuable information for analysing the behaviour being tracked in order to synchronise recorded eye movements and presented stimuli. In addition, our analysis provides an estimate of how much faster VR eye tracking would need to become in order to display stimuli in a gaze-contingent manner. Together with other studies that provide valuable insights into other parameters of data quality, such as the accuracy and precision of eye tracking and the latency of head tracking, our study contributes to a better understanding of VR eye tracking for applications in psychophysical research, which could ultimately help to enable gaze-contingent HMDs.

Next Steps in Gaze-based Locomotion Prediction

When walking, our gaze usually precedes our progress by about two steps (Hollands et al., 1995; Matthis & Fajen, 2014; Matthis et al., 2018). In contrast to grasping, for example, the execution phase between planning an action and reaching the visual target is quite long. The planning phase therefore potentially involves the selection of an action goal, which is reached comparatively far in the future.

Study II made it possible to observe how gaze precedes complex actions in a series of typical locomotor patterns. Typical gaze shifts during locomotion serve to capture visual information for the execution of upcoming movements and to monitor ongoing behaviour. We therefore assumed that eye behaviour would contain information about the selection of the next waypoint. Thus, we assumed it should be possible to predict future locomotion waypoints based on this data. Indeed, we were able to show that it was possible to combine this unsorted stream of gaze information with their corresponding position data to predict walking targets 2.5 s in the future with an error of 66 cm. We also found that eye movements enable a more accurate prediction of movement behaviour, especially in situations involving varying walking speeds. These could be decision-making situations, such as turning around or selecting an object with which to interact. Our results therefore suggest that our gaze plays an important role, especially when modifying an ongoing movement. Since a decision such as right or left or turning around has a strong impact on the selection of future walking destinations, it is precisely these phases (in which walking speed changes) that are particularly interesting for applications that use these locomotion predictions.

This result supports earlier theories that postulate gaze is an important part of planning our walking trajectories. Moreover, it appears that eye movements associated with planning our future locomotion make up a sufficiently large proportion of the total amount of gaze data to contribute to a significant improvement in prediction in LSTM models without further filtering. It seems that the previously proposed connection between our gaze and walking patterns during different tasks (that were observed in previous studies) are strong enough to systematically and automatically predict future walking actions using gaze and motion data. This means that this method of data processing could be used in various applications, for example, in improving redirected walking to reduce the number of user resets (Jeon et al., 2025). In addition, using a similar approach, it might even be possible to gain other information from gaze data to achieve such aims as decoding user preferences in product design (Palacios-Ibáñez et al., 2023).

Further Improvements

Since the publication of our research, two other studies have confirmed our main results. Bremer and Lappe (2024) found that gaze data improved predictions of the future pathway in two environments in which participants navigated using a joystick. Moreover, just like in our experiment, Y. Kim et al. (2024) found that gaze can improve the prediction of future waypoints in a walking task. However, in contrast to our model, which used 2.5 s of gaze and movement data to predict waypoints 2.5 s in the future, their model used 3 s of past data to predict waypoints 2 s in the future which resulted in overall higher prediction errors (between 113 and 126 cm) compared to our study (66 cm).

One reason for this could be the 0.5 s shorter input stream for each prediction. The association between waypoint fixation and future waypoint peaks approximately 1.5 s before a waypoint is reached (Matthis et al., 2017). Thus, the proportion of relevant information for the prediction of locomotion in the data stream from 2.5 s before reaching a waypoint could be greater than in a data stream from 3 s before reaching the waypoint. However, a post-hoc analysis of our data from Study II does not suggest this: a series of new LSTM models trained with input data of 0.5 s, 1 s, 1.5 s, 2 s and 2.5 s predicting the movement positions for the same time span into the future yielded average errors between 68 and 62 cm. Thus, the total error range did not change dramatically with changes of 0.5 s. Furthermore, in this analysis, the largest error occurred at 0.5 s and the smallest error at 2.5 s. It therefore seems more likely that the trained LSTM already reflects the typical information distribution during natural walking. Other reasons for the larger prediction error in the experiment by Y. Kim et al. (2024) could be the smaller amount of training data or differences in data quality resulting from different HMDs and the lower sampling frequency of the eye tracker used in this study.

Nevertheless, it might be interesting to systematically investigate in future experiments how the prediction error changes under identical conditions with even longer input lengths. Such an analysis was carried out by Bremer and Lappe (2024), who investigated prediction errors as a function of how far into the future their models predicted locomotion behaviour. They found that the prediction errors increased overall for longer predictions. However, based on an LSTM model, it was relatively constant between 1 s and 2 s, until it increased again after about approximately 2 s in the future. However, with Bremer and Lappe, 2024's prediction model based on a transformer architecture, the trend of increasing prediction errors for longer predictions appeared to be smaller.

Based on our results, some suggestions were made to further improve the prediction of locomotion: For example, Mayor et al. (2024) investigated the use of quaternions instead of directional angles with a range of 0° - 360°. They argued that jumps from 360° to 1° for a movement of only 1° would be especially difficult to learn for the machine learning algorithm. Quaternions, on the other hand, contain the same information without this problem. In their study, they found this to be an advantage, especially when predicting the movement position 2.5 seconds in the future. They go on to suggest that research into other machine learning algorithms could lead to further improvements. Indeed, in a direct comparison of different architectures trained on the same dataset, Bremer and Lappe (2024) showed that changing the network architecture to a transformer model can also reduce the prediction error.

Saliency-guided Eye Movements During Walking

As mentioned in the previous chapters, there are usually different types of fixations and eye movements in our gaze movements during locomotion. While visual salience is often described as an important factor in screen based visual search tasks, its influence often seems to be diminished when the ongoing tasks demand collecting additional visual information. Nevertheless, we kept the VE that we used to record the gaze data sparse and ensured that only objects related to one of the instructed tasks were present.

Thus, our recorded gaze data most likely consists of multiple types of fixations. First, there are travel-gaze-fixations that are associated with gathering just-in-time information to find the next waypoints (which are typically about 1.5 s in the future). Second, there are long-term planning travel-gaze-fixations that are on objects unrelated to the current destination, but are potentially helpful for future waypoints. Then there are other travel-gaze-fixations that are not directly related to locomotion and are used to remember properties of objects that are relevant for future tasks. Finally, there might be a small proportion of gaze movements related to monitoring grasping actions when interacting with the VE.

In walking experiments on flat terrain we usually spend slightly more than half of our fixations on the ground (Matthis et al., 2018; Patla & Vickers, 2003; Pelz & Rothkopf, 2007). In their study, Y. Kim et al. (2024) added a condition in which the amount of distracting and moving objects in the background was varied. They found that as visual distractions increased, participants performed more gaze fixations towards the background while solving the three different locomotion tasks. Normally, our travel-gaze-fixations are shorter than 0.6 s (Patla & Vickers, 2003) and tend to be directed in our current direction of motion (Hollands et al., 2002; Rothkopf et al., 2016). This facilitates getting environmental and self-motion information from the optic flow, which could be helpful for directing our feet to the planned landing positions (Patla & Vickers, 2003). In an environment with moving objects, our gaze must compensate for the visible movement of the objects to enable continuous fixation on the same target. When we walk and look at moving objects that are close to us, our gaze must compensate for both the movement of the objects and our own progress. However, in the study by Y. Kim et al. (2024), the objects were presented in the background. As a result, although our gaze continues to compensate for the movement of the target, it does not have to compensate for locomotion, as the objects remain in the background no matter how far we move. Therefore, it may be more difficult to extract information about future waypoints from the gaze, which could explain why the prediction error increased under conditions of a moving background.

Looking at salient objects that guide our eye movements as we walk raises new questions, such as where do we normally focus our attention while walking? Previous research suggests that particularly in direct comparison to situations without locomotion, the periphery contains relevant information (Craybiel et al., 1955; Franchak & Adolph, 2010; Geruschat et al., 1998; Hassan et al., 2002). Walking seems to be a highly automated action that we can perform without much cognitive effort, so it is likely that we have some reserve capacity to process other visual content as we move forward. Thus, even if our gaze is directed towards a salient non-task-relevant object, we may still be focusing covertly on walking-relevant elements of the scene. This is supported by previous experiments that suggest that our visual perception and eye movements are (at least to a certain extent) influenced by walking (Barnes et al., 2025; Cao et al., 2020; Davidson et al., 2023, 2024; Haegens et al., 2011). This process is not perfect, as shown by Matthis and Fajen (2014) who suggest that there is a certain time window in which we can absorb the visual information required for each step. If this information is not available, our steps tend to be less precise. It could therefore be interesting to investigate if shifting attention during this critical time-window has an influence on locomotion. Following the approach of Davidson et al. (2024), one could also investigate whether the direction of our attention can be assigned to three-dimensional positions relative to the head position and whether and how this relationship is modulated by our steps.

Although there are still several open questions regarding the influence of distracting objects during walking, this does not necessarily mean that predicting locomotion using gaze data is only possible in sparse environments in which the participants focus exclusively on walking tasks. Conversely, Bremer and Lappe (2024) found that gaze data can be valuable for predicting movement in various contexts, even when the link between gaze direction and future waypoints was particularly weak, as participants were solving a visual search task while moving forward in the VE.

As an alternative approach to improve locomotion prediction, Y. Kim et al. (2024) suggested adding additional gait sensors to collect data that are not influenced by salient stimuli and are therefore able to improve predictions. Another approach to improve predictions is to classify gaze data in terms of their expected information for the prediction. This could be done, for example, following the approach of Bremer et al. (2022), who trained an LSTM model that was able to distinguish gazes directed at future waypoints from gazes directed at other features of a scene. Similarly, Sharma et al. (2024) were also able to distinguish between targets and non-targets in fixations when searching for icons on desktop backgrounds and finding tools in a cluttered workshop, using a combination of eye tracking and EEG data.

In summary, although prediction accuracy varies slightly depending on the task and environment, gaze data can generally be used to improve prediction models for walking behaviour. By combining data from multiple locomotion tasks, we achieved an average prediction error of 66 cm.

Since the publication of our results, there have been well-founded suggestions as to how this accuracy could be further improved in the future. New findings about salient objects in the background that can distract our eye movements during locomotion may suggest that important information for movement prediction is encoded in gaze movements related to fixations on objects near us. However, further research is needed to confirm this hypothesis. More generally, the influence of salient background objects raises further questions about where we look while planning upcoming behaviour and how visual information and task requirements jointly control our eye movements. Our study not only shows that predicting walking goals based on gaze data is a potentially useful method for various VR applications, but also provides a clear example of how VR research can be used to better understand how we plan our behaviour.

Gaze Guidance

In Study III, we observed the trajectories of eye and head movements in a head-free search in a VE. Unsurprisingly, the results showed that some eye movements during searching are guided by the salience of visual stimuli. Initial saccades at the start of a search especially, seemed to be guided by information from the original FOV. Thus, salient targets visible in the periphery of the FOV were found quicker than less salient objects and initial eye movements landed closer to the target compared to less salient objects.

We additionally observed how top-down head movement strategies became an influential factor when the search space expanded naturally when the target was not found in the initial FOV and head movements were required to complete the task. Therefore, our results indicate that saliency plays a subordinate role in the second searching phase. Interestingly, the same stimuli which produced pop-out effects when they were visible in the FOV at the beginning of the search, did not have this capability when they were brought into the FOV by turning our head. Search behaviour was no longer influenced solely by the salience of visual stimuli and expected eye movement biases. Instead, the decision to make a particular head movement seemed to depend largely on the known arrangement of stimuli.

Peripheral targets that only became visible during the search through head movements were reached in similar search times, regardless of their saliency. In addition, initially hidden targets (that could only be seen when they were in the central FOV when the head movement towards them was almost complete) were found in a similar amount of time. This supports the conclusion that peripheral information was not used to guide the gaze. After head movements brought a set of potential targets close to the centre of the FOV, the search within this central set seemed to be again facilitated by salience.

Bottom-Up & Top-Down Factors

Our gaze can be guided at different stages by two different principles: by bottom-up visual features that are directly related to a visible salient object and by more top-down cognitive planning that is not always related to something currently in our FOV. The influence of salience often seems to be reduced when eye movements are driven by task demands. Schütz et al. (2011) assumed that bottom-up guidance is off duty whenever we perform a task. However, this does not mean that bottom-up guidance is always the most important factor when no instructions are given. Nyström and Holmqvist (2008) compared the fixations on images of three different categories and found that edge density and contrast at the fixated location did not influence initial fixation or the time course of viewing, even when no task was given. Thus, Wolfe and Horowitz (2017) hypothesised observers would internally intuitively follow a task, even without specific instructions, or when asked to view the presented scene freely.

Implicitly, these ideas imply some kind of modular structure with a bottom-up and a top-down pathway that can replace each other in different situations. This aligns with the results of Lin et al. (2008), who found that suddenly appearing (and therefore salient) objects during an ongoing search strongly attract attention, especially if they approach from the periphery or threaten to collide with us. In other words, switching between bottom-up and top-down control can take place during a task, based on visual input. Naturally, this means that information from the bottom-up path is continuously available to our control system, even if it is not currently the main guiding factor. This is consistent with the results of a few other experiments in which bottom-up mechanisms seem to continuously have the capability to distract us during free headed searches in natural VEs. For example, blurring of salient regions in a visual scene enable an overall better search performance for less salient targets (Lukashova-Sanz & Wahl, 2021). Conversely, Bektaş et al. (2019) reported that a gaze-contingent reduction in peripheral detail accuracy could improve visual search performance.

To better understand this, Einhäuser et al. (2008) set up a study on how both factors are weighted against each other. In their study, the influence of visual input as a guiding factor was first established using high- and low-contrast stimuli or flickering targets. Then the bottom-up influence was overridden by setting a simple search task for a clearly visible bulls-eye target. In the condition where participants initially preferred salient, high-contrast parts of the screen, they then developed a preference for low-contrast stimuli. This suggests that instead of completely overriding one channel of information, we are sometimes guided by a combined stream of bottom-up and top-down information. However, when using a less salient search target, contrast was not a relevant guiding factor. In the condition where flicker was the initially established as a guiding factor, the salient bullseye search target also neutralised the original influence of flicker. Therefore, Einhäuser et al. concluded that the extent to which one factor overrides the influence of another also depends on the overridden feature and the salience of the objects relevant to the new task, while the initial salience features in the environment appear to play a subordinate role.

The weighting of top-down and bottom-up factors can also be shifted by non-visual changes as a study by Foulsham and Underwood (2007) shows. They found that objects with higher salience were fixated earlier and more frequently than objects with lower saliency, while participants memorised different complex scenes. Once the task was changed to searching for a target defined by a category or example, salience was no longer the decisive guiding factor. A similar result was presented by Longstaffe et al. (2014), when participants searched in a room with a series of illuminated buttons on the floor to find a target that changed colour when pressed. Some buttons flashed while others remained lit. Participants chose the flashing buttons more often, even though they knew that doing so had no effect on the search results. At the same time, they were able to choose a sensible search sequence and avoid selecting the same switches twice. Nevertheless, in a condition where memorisation was not necessary because the already selected targets were visually marked, participants were less likely to explore flashing locations. Thus, in this task, top-down guidance was not able to completely override the influence of bottom-up information. Interestingly, in these experiments, salient targets were chosen even more often when a memorizing task was done simultaneously. This suggests that our memory capacity can also influence how highly we weight salient visual stimuli as a guiding factor.

In summary, it can be said that various guiding factors are integrated with each other in a common module. One possible outcome is that one factor neutralises all others. Tatler et al. (2011) assumed that this is particularly the case when non-visual factors such as time pressure or behavioural costs come into play. They assumed that this would allow particularly important fixations to be carried out without distraction and with particular precision in time-critical situations such as driving or walking. However, given the findings of Longstaffe et al. (2014), their assumption that salience prevails particularly in situations with low cognitive load appears overly simplistic. Nevertheless, it is possible that top-down factors such as instructions and working memory load modulate the effect of bottom-up factors.

Timing

The influence of different guiding factors on our eye movements changes depending on the visual content and over time. Schütz et al. (2012) observed that saccades with short latency were mainly guided by salience, while saccades with long latency could also be influenced by value information. Similar conclusions can also be drawn from a screen-based study by Wolf and Lappe (2021), in which participants were instructed to look at either the centre of a picture, a salient object, or a position between the two options. Their results showed that quickly executed saccades (triggered between 80 ms and 250 ms after the appearance of visual stimuli) are susceptible to involuntary distortions caused by salient objects or the centre of the image. However, saccades with a longer latency and without any instructions on where to look frequently landed on salient stimuli. This suggests an interaction between the guidance of the instructions and the timing of the execution of the eye movement.

This aligns with a range of findings from van Zoest et al. (2004), in which saccades with latencies below 250 ms were stimulus driven, whereas later eye movements were more goal driven. This suggests that visual input is likely used for bottom-up guidance as soon as it is available. However, some visual features are decoded relatively late in the visual pathway from the visual inputs and are therefore only available with a certain latency. For example, Palmer et al. (2019) found that it takes 200 – 300 ms to utilise colour information in a guided search. Then again, there are some search tasks that we can solve surprisingly quickly: it only takes 150 ms to recognise whether an animal is visible in an image or not (Thorpe et al., 1996). However, looking at the results of the influence of top-down processes after 250 ms, it seems that even with these highly efficient search processes, there is a very limited time frame in which bottom-up factors such as salience can have an effect in situations where an instruction is given.

One explanation of the potential functions leading to the reduction of the influence of bottom-up salience was given by Donk and van Zoest (2008). Based on two experiments, they concluded that salience could be represented in the visual system only for a limited period of time after it is received. In both experiments, participants were asked to find the most salient object among two salient objects and many distractors. Again, performance varied with timing. Initially, participants were able to detect the most salient object. However, in trials with saccade latencies longer than 300 ms, participants were not able to distinguish the most salient object from the other salient object. In a second experiment, the presentation time of the stimulus was systematically varied. If objects were visible longer than 83 ms the average proportion of correct answers fell by about 10%. From all of these results Donk and van Zoest concluded that the relative salience between objects is represented in neural response map via latency: Following this idea, neurons representing the salience map with receptive fields containing highly salient objects would fire earlier than neurons with receptive fields containing less salient stimuli. After the latency offset, the neurons corresponding to the location of the less salient object also fire. Then, the two may no longer be distinguishable in terms of their relative salience. Instead, the saliency map would then only contain information about the locations of all salient objects in the FOV.

Similarly, such temporal effects could potentially explain some of the typical observations made when shifting from bottom-up to top-down control. For example, Wolf and Lappe (2020) found that this shift did not simply depend on the duration of the task, but rather on how long salient objects that initially appeared distracting had already been visible. They concluded that the visually available bottom-up saliency must first be suppressed to allow the influence of top-down factors. In theory, this suppression mechanism could be related to the loss of relative salience information and a therefore more spread-out saliency signal.

In many situations, the first saccade could already be carried out while relative salience is still represented in the visual system. Thus, in many arrangements of objects in our FOV the first saccade is guided by bottom-up factors. This fits well to observations by Brouwer et al. (2009) , who found that when we look at an object that we want to grasp, the first saccade tends to go towards the centre of gravity of the target object. The following second fixation, however, is often more task-related and brings areas relevant for grasping into foveal vision. In line with the findings of Donk and van Zoest, van Heusden et al. (2021) found that relative salience appears to be represented long enough to guide the first saccade, even when it is directed towards peripheral targets, where the planning and execution phases are somewhat prolonged. Moreover, they found that in this scenario the whole process of switching from bottom-up to more top-down guidance seems to start up to 50 ms later. One could therefore argue that the weighting of the guiding factors is also influenced by the actual sequence of eye movements based on the structure of the scene.

Head Movements

Many processes in our visual system are organised retinotopically. This is also the case for the neurons processing bottom-up salience signals (Bichot et al., 2005). As a result, salient objects can pop-out from the FOV during searching and trigger eye movements towards them. In line with many experiments on guidance during searching in the FOV, also Study III showed that initial saccades at the start of a search are guided by salient visual stimuli.

At the start of a head movement no visual information about the head movement target in the extended FOV is available. Therefore, we need to rely on non-visual information when planning and initiating the head movement. In general, contextual guidance or the use memorized or learned content is an established principle for visual guidance. For example, there is experimental evidence that top-down information about scenes can be used to refine the search space in a scene (Hwang et al., 2011; Kanan et al., 2009; Neider & Zelinsky, 2006). In addition, many of the regions involved in the selection and generation of saccades are sensitive to the expectation of reward (Jovancevic-Misic & Hayhoe, 2009). For example, in monkeys saccade-related areas in the cortex (lateral intraparietal area, frontal eye field, supplementary eye field and dorsolateral prefrontal cortex) all show sensitivity to rewards (Dorris & Glimcher, 2004; Platt & Glimcher, 1999; Stuphorn et al., 2000; Sugrue et al., 2004). As a result, it makes sense to assume that we rely more heavily on memorised content to plan our head movements while the FOV is being updated.

The use of a non-visual signal to plan our head movements suggests that this movement plan is not represented retinotopically in the visual system. This raises the question of which coordinate system we use to plan our movements instead. Crowe et al. (2025) investigated search time differences for search displays that were occluded with a movable aperture. In separate conditions participants could use a mouse to either move the aperture or the presented stimuli behind it. They found that the search time was shorter when participants moved the aperture instead of the stimulus. This suggests that the non-visible search space is not organised in absolute world-coordinates, but relative to the hand movements. This was also the case when an additional visible reference was present and even when the aperture and the search display were constantly wobbling.

However, if the assignment between mouse movements and the movement on the screen was reversed, the search time pattern inverted. The search then became faster when the search display was moved instead of the aperture. This suggests that the congruence between the moving direction and the visual change was more important for the speed of the search than which object is moved.

Head movements result in a reliable shift of the current visual input. Therefore, the representation or organisation of the surrounding objects could also be organised using the movement signal. In our experiments, it seemed as if this head movement strategy was based mainly on the structure of the scene. This fits well with the results of Shioiri et al. (2018) who found that participants are able to quickly learn the spatial arrangements of displays placed around them, while doing a search task. As a result, the area in which we can move our eyes is determined by head movements that are guided by top-down factors. Consequently, the influence of top-down factors increases as soon as scenarios are considered in which head movements are necessary to reach a target. Using 360° scenes, Haskins et al. (2020) compared eye movements between two experimental conditions. In the first, simulated head movements were presented. In the second condition, participants were free to move their heads as they explored the scene. Haskins et al. found that during active viewing, participants used shorter, more exploratory fixations, while the influence of salience decreased and the effect of meaningful scene areas increased.

Generally, once a head movement has started, we could in theory use visual information that became available through the shift of the FOV to alter the ongoing head movement. In our experiments this did not seem to be the case. However, previous experiments found that individuals use extraretinal cues such as neck proprioception, information about the neck movement and vestibular information to perceive visual input while turning our heads (Crowell et al., 1998), which might explain behavioural differences between simulated and actual head turns. Then again, the image on the retina is not completely stabilised during head movements (Skavenski et al., 1979; Steinman & Collewijn, 1980), which suggests that guidance of head movements is more likely based on a combination of top-down factors, such as memory and previously learned structures. One could therefore argue that at the start of a search, bottom-up salience is guiding our first saccade towards the first object to inspect. If the search target is not in the initial FOV, we need to expand the search space by moving our head to obtain more information in the extended FOV. During this process, we typically follow a saccade and catch-up strategy in which first a rapid eye movement towards potential targets in the periphery is made and simultaneously a head movement is initiated. Once the head movement has caught up and the previous saccade target is close to central vision, another saccade follows to inspect new objects located in peripheral regions of the newly established FOV. This means that no bottom-up visual information is available when planning a head movement. Although head movements are slow and visual stimuli can be stabilised to a certain extent on the retina even during continuous head rotation, it appears, at least in our experiments, that when searching for a scene with a known structure, we do not change our planned head movement trajectory based on the information received during the shift of the FOV. Once a new FOV was established and the head movement ended, bottom-up information seemed to be able to guide gaze again.

This leads us to conclude that our attentional system has both bottom-up and top-down information available at the same time, so that the weight given to different guiding factors can change rapidly. It seems that both, bottom-up and top-down guidance, can modulate each other to a certain degree. For example, our eye movements may initially be strongly action-related, but can then be distracted by a salient stimulus. Conversely, we can also give ourselves instructions in an environment with a lot of salient factors to find relevant objects in our surroundings. Bottom-up guidance becomes available earlier than the influence of top-down factors, the latter usually becoming more important after a few hundred milliseconds, when specific instructions are given. After this delay, we often do a second, more top-down driven, saccade. The delay can be longer when salient objects are located in the periphery. Finally, it is clear that no visual information is available as a guiding factor for the initial planning of head movements towards the extended FOV. In line with this, our results suggest that these are planned based on the structure of the scene. In our experiments, it also appeared that head movements were not influenced by information that may have become visible during their execution. As soon as head movements are necessary to reach a target, top-down factors gain importance overall, as top-down-controlled head movements determine the area of our environment in which we can perform eye movements.

Limitations and Outlook for VR Research

The experiments in this thesis show that it is possible to collect behavioural data using VR that meets scientific standards. Nevertheless, it is reasonable to ask how VR measurements compare to data previously collected in simplified environments using desktop setups and whether there are systematic differences between the two. There are several experiments which suggest that eye movement patterns are similar when performing the same task in the real world and in realistic simulations (Drewes et al., 2021; Gulhan et al., 2021; S. Kim et al., 2018). Then again, studies have shown that VR alters the perception of walking speed (Banton et al., 2005) and can influence walking dynamics when walking on a treadmill (Sloot et al., 2014). Moreover, there are several other potential reasons why, even in ideal conditions, users may not act naturally in VR researchers should be aware of. First, wearing an HMD feels different from not wearing one. Therefore, although the devices are becoming lighter with every generation, right now the additional weight of more than 700 g on the head is still clearly noticeable. The extra weight on the head and joints can make many natural movements more strenuous, which can lead to unnatural compensatory movements. Second, on current devices, the FOV is more limited than in reality, which could potentially alter visual perception for example during locomotion.

In light of these limitations, it is important to distinguish between undesirable side effects of VR and differences that arise because the task is closer to natural human behaviour. For example T. C. Wu and Tsotsos (2025) conducted a real life search task, in which participants walked around a room to find objects that were presented on a screen. All movements of the head and eyes were tracked throughout the whole task. In their experiment, T. C. Wu and Tsotsos reported an average fixation duration of about 200 ms which is similar to results from other more simplistic studies. However, the reported average saccade amplitude was 10.4° which is higher than in 2D natural image search with average saccade lengths of 5-6°. They also found several environment fixations in which participants did not look at task relevant objects, but looked at other objects, for example, when navigating the room. Similar conclusions can be drawn from our data in Study II. In a direct comparison between VR and simple desktop studies, it would be easy to dismiss such differences as VR bias. In this situation, it is important to consider carefully whether the deviation in the behavioural data could also be a result of observing more natural behaviours in VR, which may describe typical human behaviour even better than the results of desktop studies.

Finally, some differences in measured behaviour may result from using different measuring equipment. Compared to some desktop-based eye trackers such as the Eyelink 1000, which has established itself as the gold standard in eye tracking research, VR eye trackers are still rather imprecise. Therefore, when conducting a VR experiment, more data is required to achieve a similar level of precision by averaging the information about eye movements across multiple trials. The overall lower accuracy means that it is not possible to distinguish eye movements with very short amplitudes, such as fixational eye movements. Our latency measurements in chapter I showed that measuring eye movements usually takes several frames. In addition, due to its significantly lower spatial and temporal resolution, the currently available VR eye tracking hardware cannot be readily used for online measurement of saccade dynamics. It also does not seem possible to use it to detect saccade onsets quickly enough to present VEs with gaze-contingent experimental stimuli. This would require the positions of the virtual objects to be adjusted to the gaze, which would require low latency eye and head tracking. Nevertheless, it is possible to use HMDs to observe eye and gaze paths of saccades and pursuit movements, as well as gaze points and fixations, if the data is analysed retrospectively and averaged across several trials.

When doing so, we should consider that the gaze signal resulting from combining eye and head tracking suffers from imperfections that must be taken into account when planning VR experiments and interpreting their results. Based on recordings during the pilot phase with the HTC Vive Pro Eye, we suspect that eye and head tracking are not recorded perfectly synchronised. For example, when horizontal head movements are recorded while we are fixating a point, our eye movements compensate for our head movements so that we can see the fixation target clearly while moving our head. If we calculate the difference between the speeds of eye and head movements, we often find rapid velocities of over 30 °/s when using current VR hardware. Then again, finding perfect stabilisation at speeds close to 0°/s seems to be unlikely based on previous results (Harris, 1994). Nevertheless, we cannot distinguish between actual retinal slip and additional errors caused by asynchrony between eye and head tracking with the current setup. This makes it difficult or even impossible to identify fixation objects based on an eye velocity threshold, as is often used in desktop eye tracking experiments. One method for determining gaze points and gaze patterns in VR is calculating intersections between the gaze ray and the positions of objects in the VE. When the same object is looked at for multiple frames, we can assume that it was fixated, even though the output velocity of the gaze signal wrongly suggests that the image on the retina was not perfectly stable.

In summary, as a scientific method, VR is no longer in its infancy and offers great potential for systematically observing complex movement sequences (such as searching for an object in a room). This means that although some previously established methods need to be adjusted due to flaws in the currently available hardware, based on our tests, the combined gaze signal provides a solid overview of where and how long a person looked in a sequence, especially when combined with information about the VE. All in all, this makes VR eye tracking particularly interesting as a method for observing eye movements during actions and enables experiments in which participants can explore their surroundings more naturally than in experiments in front of a desk. In situations like these, various processes simultaneously influence our behaviour. We can therefore use VR to gain new insights into our perception and behavioural control. High temporal resolution and the combination of detailed participant behaviour along with the exact description of the stimuli presented in the experiment make it possible to describe the timing and connection between perception and behaviour much more accurately than before. The extensive amount of data enables new evaluation methods, such as machine learning, to automatically classify the collected data.

Conclusions

The rapid development of VR technology in recent years has led to the combination and improvement of several different technologies for tracking human behaviour. Simultaneously, the possibilities for presenting visual stimuli that respond to our movements have improved.

In this thesis, we showed how observing eye, head and walking movements in VR can improve our understanding of how we perceive and explore our surroundings. As a first step, we evaluated VR eye tracking latency to be able to use it as a new method in psychophysics. We found that eye tracking latency ranged between 45 and 81 ms on different devices. Second, we illustrated that eye and head movements offer predictive added value for locomotion forecasting. This fits well with previous knowledge about the planning and execution of walking movements. In our study, we predicted future movement behaviour in VR. Such prediction models could be used to minimise errors in motion tracking by making plausibility assumptions based on such motion predictions. Moreover, similar prediction models could be used to compensate for any system latency with software by performing tracking based on a mixture of measured and predicted data. Our predictions can also be used to create immersive VEs that automatically respond to intended behaviour and are thus able to keep users away from obstacles. Since we relied exclusively on egocentric movement data, the same method could also be applied in many real-world contexts. In such contexts, it is helpful if controllable environmental parameters can react proactively to human movements. For example, when learning new movement sequences, this prevents errors at an early stage and avoids damage to work materials and the practice of incorrect movement patterns. Such predictions could be helpful in surveillance systems for hazardous areas. These safety systems would then no longer just react when a person enters a hazardous area, but could also issue an early warning or automatically eliminate potential hazards based on movement predictions using egocentric data. Finally, we used a combination of VR eye and head tracking to expand the classic visual search paradigm. In two experiments on search in the extended FOV, we observed how head and eye movements interact with each other to form our gaze trajectory. We found that head movements in particular tended to be guided by top-down features and gaze was not immediately influenced by stimuli that entered the FOV through a head movement. This description of search behaviour suggests that salience plays a rather subordinate role when searching in real and virtual environments.

On this basis, new experiments using VR eye tracking can be developed to describe the entire search process with its various subtasks in greater detail. To this day, there remain numerous unanswered questions about the extent to which the distribution of attention when walking or retrieving memories interacts with the control of search processes, especially when a target enters our FOV through self-motion.

VR and eye tracking technology has made significant advances over the past decade. Ivan Sutherland envisioned many of these developments when he contemplated the possibilities of the ultimate display. Although his vision of a space in which a computer can control the existence of matter has not (yet) been realised, the technology that has continuously evolved towards his ideas can already help us to observe more accurately how we process and perceive our environment. This is because the rapid development of VR hardware has been accompanied by increasingly detailed behavioural measurements. As a result, HMDs are useful tools for improving, refining and expanding dozens of experiments and existing theories. This makes VR a method with great potential for improving our understanding of how we see, perceive and act.

Bibliography

Abdlkarim, D., Di Luca, M., Aves, P., Maaroufi, M., Yeo, S.-H., Miall, R. C., Holland, P., & Galea, J. M. (2024). A methodological framework to assess the accuracy of virtual reality hand-tracking systems: A case study with the Meta Quest 2. *Behavior Research Methods*, *56*(2), 1052–1063. https://doi.org/10.3758/s13428-022-02051-8

Adhanom, I. B., Lee, S. C., Folmer, E., & MacNeilage, P. (2020). Gazemetrics: An open-source tool for measuring the data quality of HMD-based eye trackers. *ACM Symposium on Eye Tracking Research and Applications*, 1–5. https://doi.org/10.1145/3379156.3391374

Adhanom, I. B., MacNeilage, P., & Folmer, E. (2023). Eye tracking in virtual reality: A broad review of applications and challenges. *Virtual Reality*, *27*(2), 1481–1505. https://doi.org/10.1007/s10055-022-00738-z

Albert, R., Patney, A., Luebke, D., & Kim, J. (2017). Latency requirements for foveated rendering in virtual reality. *ACM Transaction of Applied Perception*, *14*(4). https://doi.org/10.1145/3127589

Alexander, R. G., & Martinez-Conde, S. (2019). Fixational eye movements. In C. Klein & U. Ettinger (Eds.), *Eye movement research: An introduction to its scientific foundations and applications* (pp. 73–115). Springer International Publishing. https://doi.org/10.1007/978-3-030-20085-5_3

Alsaeedi, N., & Wloka, D. (2019). Real-time eyeblink detector and eye state classifier for virtual reality (VR) headsets (head-mounted displays, HMDs). *Sensors*, *19*(5), 1121.

Anderson, N. C., Bischof, W. F., Foulsham, T., & Kingstone, A. (2020). Turning the (virtual) world around: Patterns in saccade direction vary with picture orientation and shape in virtual reality. *Journal of Vision*, *20*(8), 21–21. https://doi.org/10.1167/jov.20.8.21

Anderson, N. C., Ort, E., Kruijne, W., Meeter, M., & Donk, M. (2015). It depends on when you look at it: Salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision*, *15*(5), 9–9. https://doi.org/10.1167/15.5.9

Andersson, R., Nyström, M., & Holmqvist, K. (2010). Sampling frequency and eye-tracking measures: How speed affects durations, latencies, and more. *Journal of Eye Movement Research*, *3*(3). https://doi.org/10.16910/jemr.3.3.6

Andreu-Sánchez, C., Martín-Pascual, M. Á., Gruart, A., & Delgado-García, J. M. (2017). Eyeblink rate watching classical hollywood and post-classical MTV editing styles, in media and non-media professionals. *Scientific Reports*, *7*(1), 43267. https://doi.org/10.1038/srep43267

Andreu-Sánchez, C., Martín-Pascual, M. Á., Gruart, A., & Delgado-García, J. M. (2021). Viewers change eye-blink rate by predicting narrative content. *Brain Sciences*, *11*(4). https://doi.org/10.3390/brainsci11040422

Anthes, C., García-Hernández, R. J., Wiedemann, M., & Kranzlmüller, D. (2016). State of the art of virtual reality technology. *2016 IEEE Aerospace Conference*, 1–19. https://doi.org/10.1109/AERO.2016.7500674

Arabadzhiyska, E., Tursun, O. T., Myszkowski, K., Seidel, H.-P., & Didyk, P. (2017). Saccade landing position prediction for gaze-contingent rendering. *ACM Trans. Graph.*, *36*(4). https://doi.org/10.1145/3072959.3073642

Arechavaleta, G., Laumond, J.-P., Hicheur, H., & Berthoz, A. (2008). An optimality principle governing human walking. *IEEE Transactions on Robotics*, *24*(1), 5–14. https://doi.org/10.1109/TRO.2008.915449

Arefin, M. S., Swan II, J. E., Cohen Hoffing, R. A., & Thurman, S. M. (2022). Estimating perceptual depth changes with eye vergence and interpupillary distance using an eye tracker in virtual reality. *2022 Symposium on Eye Tracking Research and Applications*. https://doi.org/10.1145/3517031.3529632

Aytekin, M., Victor, J. D., & Rucci, M. (2014). The visual input to the retina during natural head-free fixation. *Journal of Neuroscience*, *34*(38), 12701–12715. https://doi.org/10.1523/JNEUROSCI.0229-14.2014

Aziz, S., Lohr, D. J., Friedman, L., & Komogortsev, O. (2024). Evaluation of eye tracking signal quality for virtual reality applications: A case study in the Meta Quest Pro. *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*. https://doi.org/10.1145/3649902.3653347

Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science*, *14*(6), 468–469.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, *24*(3), 191–204. https://doi.org/10.1016/0025-5564(75)90075-9

Baird, P. N., Saw, S.-M., Lanca, C., Guggenheim, J. A., Smith III, E. L., Zhou, X., Matsui, K.-O., Wu, P.-C., Sankaridurg, P., Chia, A., Rosman, M., Lamoureux, E. L., Man, R., & He, M. (2020). Myopia. *Nature Reviews Disease Primers*, *6*(1), 99. https://doi.org/10.1038/s41572-020-00231-4

Baker, S. (2022, April). Pulse width modulation (pwm). https://tftcentral.co.uk/articles/pulse_width_modulation

Ballard, D. H., Hayhoe, M. M., Li, F., Whitehead, S. D., Frisby, J., Taylor, J. G., Fisher, R. B., Barlow, H. B., Frisby, J. P., Horridge, G. A., & Jeeves, M. A. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions of* *the Royal Society of London. Series B: Biological Sciences*, *337*(1281), 331–339. https://doi.org/10.1098/rstb.1992.0111

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Bang, K., Jo, Y., Chae, M., & Lee, B. (2021). Lenslet VR: Thin, flat and wide-FOV virtual reality display using Fresnel lens and lenslet array. *IEEE Transactions on Visualization and Computer Graphics*, *27*(5), 2545–2554. https://doi.org/10.1109/TVCG.2021.3067758

Bansal, A., Weech, S., & Barnett-Cowan, M. (2019). Movement-contingent time flow in virtual reality causes temporal recalibration. *Scientific reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-40870-6

Banton, T., Stefanucci, J., Durgin, F., Fass, A., & Proffitt, D. (2005). The perception of walking speed in a virtual environment. *Presence*, *14*(4), 394–406. https://doi.org/10.1162/105474605774785262

Barnes, L., Davidson, M. J., & Alais, D. (2025). The speed and phase of locomotion dictate saccade probability and simultaneous low-frequency power spectra. *Attention, Perception, & Psychophysics*, *87*(1), 245–260. https://doi.org/10.3758/s13414-024-02932-4

Becker, S., Hug, R., Hübner, W., & Arens, M. (2018). An evaluation of trajectory prediction approaches and notes on the trajnet benchmark. *arXiv:1805.07663*. https://doi.org/10.48550/arXiv.1805.07663

Bektaş, K., Çöltekin, A., Krüger, J., Duchowski, A. T., & Fabrikant, S. I. (2019). GeoGCD: Improved visual search via gaze-contingent display. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. https://doi.org/10.1145/3317959.3321488

Belardinelli, A., Stepper, M. Y., & Butz, M. V. (2016). It's in the eyes: Planning precise manual actions before execution. *Journal of Vision*, *16*(1), 18–18. https://doi.org/10.1167/16.1.18

Benjamin, A. V., Wailes-Newson, K., Ma-Wyatt, A., Baker, D. H., & Wade, A. R. (2018). The effect of locomotion on early visual contrast processing in humans. *Journal of Neuroscience*, *38*(12), 3050–3059. https://doi.org/10.1523/JNEUROSCI.1428-17.2017

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berryhill, M. E., Chiu, T., & Hughes, H. C. (2006). Smooth pursuit of nonvisual motion. *Journal of Neurophysiology*, *96*(1), 461–465. https://doi.org/10.1152/jn.00152.2006

Bertera, J. H. (1988). The effect of simulated scotomas on visual search in normal subjects. *Investigative Ophthalmology & Visual Science*, *29*, 470–475.

Bertera, J. H., & Rayner, K. (2000). Eye movements and the span of the effective stimulus in visual search. *Perception & Psychophysics*, *62*(3), 576–585. https://doi.org/10.3758/BF03212109

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*(5721), 529–534. https://doi.org/10.1126/science.1109676

Bierman, A., Figueiro, M. G., & Rea, M. S. (2011). Measuring and predicting eyelid spectral transmittance. *Journal of Biomedical Optics*, *16*(6), 067011. https://doi.org/10.1117/1.3593151

Bimberg, P., Weissker, T., & Kulik, A. (2020). On the usage of the simulator sickness questionnaire for virtual reality research. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 464–467. https://doi.org/10.1109/VRW50115.2020.00098

Binello, A., Mannan, S., & Ruddock, K. (1995). The characteristics of eye movements made during visual search with multi-element stimuli. *Spatial Vision*. https://doi.org/10.1163/156856895x00043

Blignaut, P., & Beelders, T. (2012). The precision of eye-trackers: A case for a new measure. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 289–292. https://doi.org/10.1145/2168556.2168618

Boettcher, S. E. P., Draschkow, D., Dienhart, E., & Võ, M. L.-H. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects on eye movements during visual search. *Journal of Vision*, *18*(13), 11–11. https://doi.org/10.1167/18.13.11

Bollimunta, A., Bogadhi, A. R., & Krauzlis, R. J. (2018). Comparing frontal eye field and superior colliculus contributions to covert spatial attention. *Nature Communications*, *9*(1), 3553. https://doi.org/10.1038/s41467-018-06042-2

Bölling, L., Stein, N., Steinicke, F., & Lappe, M. (2019). Shrinking circles: Adaptation to increased curvature gain in redirected walking. *IEEE Transactions on Visualization and Computer Graphics*, *25*(5), 2032–2039. https://doi.org/10.1109/TVCG.2019.2899228

Bolte, B., & Lappe, M. (2015). Subliminal reorientation and reposition in immersive virtual environments using saccadic suppression. *IEEE Transactions on Visualization and Computer Graphics*, *21*(4), 545–552. https://doi.org/10.1109/TVCG.2015.2391851

Bonnen, K., Matthis, J. S., Gibaldi, A., Banks, M. S., Levi, D. M., & Hayhoe, M. (2021). Binocular vision and the control of foot placement during walking in natural terrain. *Scientific Reports*, *11*(1), 20881. https://doi.org/10.1038/s41598-021-99846-0

Borg, O., Casanova, R., & Bootsma, R. J. (2015). Reading from a head-fixed display during walking: Adverse effects of gaze stabilization mechanisms. *PLOS ONE*, *10*(6), 1–14. https://doi.org/10.1371/journal.pone.0129902

Bosco, A., Lappe, M., & Fattori, P. (2015). Adaptation of saccades and perceived size after trans-saccadic changes of object size. *Journal of Neuroscience*, *35*(43), 14448–14456. https://doi.org/10.1523/JNEUROSCI.0129-15.2015

Botch, T. L., Garcia, B. D., Choi, Y. B., Feffer, N., & Robertson, C. E. (2023). Active visual search in naturalistic environments reflects individual differences in classic visual search performance. *Scientific Reports*, *13*(1), 1–9. https://doi.org/10.1038/s41598-023-27896-7

Bouvier, S. E., Cardinal, K. S., & Engel, S. A. (2008). Activity in visual area V4 correlates with surface perception. *Journal of Vision*, *8*(7), 28–28. https://doi.org/10.1167/8.7.28

Boynton, G. M. (2005). Attention and visual perception. *Current Opinion in Neurobiology*, *15*(4), 465–469. https://doi.org/10.1016/j.conb.2005.06.009

Bremer, G., & Lappe, M. (2024). Predicting locomotion intention using eye movements and EEG with LSTM and transformers. *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 21–30. https://doi.org/10.1109/ISMAR62088.2024.00016

Bremer, G., Stein, N., & Lappe, M. (2021). Predicting future position from natural walking and eye movements with machine learning. *2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 19–28. https://doi.org/10.1109/AIVR52153.2021.00013

Bremer, G., Stein, N., & Lappe, M. (2022). Do they look where they go? Gaze classification during walking. *NeuRIPS 2022 Workshop on Gaze Meets ML*. https://openreview.net/forum?id=XP0k6ToFK7t

Brenner, E., Ghiani, A., Mann, D., & Smeets, J. B. (2024). Where do people look when they walk or run at different speeds? *Journal of Vision*, *24*(10), 271–271. https://doi.org/10.1167/jov.24.10.271

Brenner, E., Janssen, M., de Wit, N., Smeets, J. B., Mann, D. L., & Ghiani, A. (2024). Running together influences where you look. *Perception*, *53*(5-6), 397–400. https://doi.org/10.1177/03010066241235112

Brewer, A. A., Liu, J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience*, *8*(8), 1102–1109. https://doi.org/10.1038/nn1507

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*(6), 719–722. https://doi.org/10.1016/0042-6989(75)90290-4

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2020). Studying human behavior with virtual reality: The unity experiment framework.

Behavior research methods, *52*, 455–463. https://doi.org/10.3758/s13428-019-01242-0

Brouwer, A.-M., Franz, V. H., & Gegenfurtner, K. R. (2009). Differences in fixations between grasping and viewing objects. *Journal of Vision*, *9*(1), 18–18. https://doi.org/10.1167/9.1.18

Bruder, G., Steinicke, F., Bolte, B., Wieland, P., Frenz, H., & Lappe, M. (2013). Exploiting perceptual limitations and illusions to support walking through virtual environments in confined physical spaces. *Displays*, *34*(2), 132–141. https://doi.org/10.1016/j.displa.2012.10.007

Brument, H., Podkosova, I., Kaufmann, H., Olivier, A. H., & Argelaguet, F. (2019). Virtual vs. physical navigation in VR: Study of gaze and body segments temporal reorientation behaviour. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 680–689. https://doi.org/10.1109/VR.2019.8797721

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? a tutorial of power analysis with reference tables. *Journal of Cognition*. https://doi.org/10.5334/joc.72

Bullock, D. H., & Bitterman, M. E. (1962). Probability-matching in the pigeon. *The American Journal of Psychology*, *75*(4), 634–639. https://doi.org/10.2307/1420288

Burr, D. C., Ross, J., Binda, P., & Morrone, M. C. (2010). Saccades compress space, time and number. *Trends in cognitive sciences*, *14*(12), 528–533. https://doi.org/10.1016/j.tics.2010.09.005

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art.* Univ. Chicago Press.

Calow, D., & Lappe, M. (2008). Efficient encoding of natural optic flow. *Network: Computation in Neural Systems*, *19*(3), 183–212. https://doi.org/10.1080/09548980802368764

Caltrider, D., Gupta, A., & Tripathy, K. (2023). *Evaluation of visual acuity*. StatPearls Publishing. http://europepmc.org/books/NBK564307

Campbell, F. W., & Wurtz, R. H. (1978). Saccadic omission: Why we do not see a grey-out during a saccadic eye movement. *Vision Research*, *18*(10), 1297–1303. https://doi.org/10.1016/0042-6989(78)90219-5

Cao, L., Chen, X., & Haendel, B. F. (2020). Overground walking decreases alpha activity and entrains eye movements in humans. *Frontiers in Human Neuroscience*. https://www.proquest.com/scholarly-journals/overground-walking-decreases-alpha-activity/docview/2471936104/se-2

Cao, L., & Händel, B. (2019). Walking enhances peripheral visual processing in humans. *PLOS Biology*, *17*(10), 1–23. https://doi.org/10.1371/journal.pbio.3000511

Carmack, J. (2013). Latency mitigation strategies (by John Carmack) [Accessed 21-07-2025].

Carpenter, R. (2001). Express saccades: Is bimodality a result of the order of stimulus presentation? *Vision Research*, *41*(9), 1145–1151. https://doi.org/10.1016/S0042-6989(01)00007-4

Chaturvedi, V., & Van Gisbergen, J. A. (1997). Specificity of saccadic adaptation in three-dimensional space. *Vision research*, *37*(10), 1367–1382. https://doi.org/10.1016/S0042-6989(96)00266-0

Chen, H.-W., Lee, J.-H., Lin, B.-Y., Chen, S., & Wu, S.-T. (2018). Liquid crystal display and organic light-emitting diode display: Present status and future perspectives. *Light: Science & Applications*, *7*(3), 17168–17168. https://doi.org/10.1038/lsa.2017.168

Cho, Y.-H., Lee, D.-Y., & Lee, I.-K. (2018). Path prediction using LSTM network for redirected walking. *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 527–528. https://doi.org/10.1109/VR.2018.8446442

Chow, H. M., Knöll, J., Madsen, M., & Spering, M. (2021). Look where you go: Characterizing eye movements toward optic flow. *Journal of Vision*, *21*(3), 19–19. https://doi.org/10.1167/jov.21.3.19

Clay, V., König, P., & Koenig, S. (2019). Eye tracking in virtual reality. *Journal of Eye Movement Research*, *12*(1). https://doi.org/10.16910/jemr.12.1.3

Clemotte, A., Velasco, M. A., Torricelli, D., Raya, R., & Ruiz, R. C. (2014). Accuracy and precision of the tobii x2-30 eye-tracking under non ideal conditions. https://doi.org/10.5220/0005094201110116

Collins, T., Doré-Mazars, K., & Lappe, M. (2007). Motor space structures perceptual space: Evidence from human saccadic adaptation. *Brain Research*, *1172*, 32–39. https://doi.org/10.1016/j.brainres.2007.07.040

Coral Gabbay, A. Z., & Lamy, D. (2019). Splitting the attentional spotlight? Evidence from attentional capture by successive events. *Visual Cognition*, *27*(5-8), 518–536. https://doi.org/10.1080/13506285.2019.1617377

Cornell, E. D., MacDougall, H. G., Predbon, J., Curthoys, & S., I. (2003). Errors of binocular fixation are common in normal subjects during natural conditions. *Optometry and Vision Science*, *80*(11). https://doi.org/10.1097/00006324-200311000-00014

Cornia, M., Baraldi, L., Serra, G., & Cucchiara, R. (2018). Predicting human eye fixations via an lstm-based saliency attentive model. *IEEE Transactions on Image Processing*, *27*(10), 5142–5154. https://doi.org/10.1109/TIP.2018.2851672

Cornsweet, T. N., & Crane, H. D. (1973). Accurate two-dimensional eye tracker using first and fourth purkinje images. *J. Opt. Soc. Am.*, *63*(8), 921–928. https://doi.org/10.1364/JOSA.63.000921

Craybiel, A., Jokl, E., & Trapp, C. (1955). Notes: Russian studies of vision in relation to physical activity and sports. *Research Quarterly. American Association for Health, Physical Education and Recreation*, *26*(4), 480–485. https://doi.org/10.1080/10671188.1955.10612840

Crowe, E. M., Vorgia, D. T., & Brenner, E. (2025). Congruency between viewers' movements and the region of the display being sampled speeds up search through an aperture. *Perception*, *54*(4), 226–238. https://doi.org/10.1177/03010066251314181

Crowell, J. A., Banks, M. S., Shenoy, K. V., & Andersen, R. A. (1998). Visual self-motion perception during head turns. *Nature Neuroscience*, *1*(8), 732–737. https://doi.org/10.1038/3732

Cullen, K. E., & Roy, J. E. (2004). Signal processing in the vestibular system during active versus passive head movements. *Journal of Neurophysiology*, *91*(5), 1919–1933. https://doi.org/10.1152/jn.00988.2003

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of Comparative Neurology*, *292*(4), 497–523. https://doi.org/10.1002/cne.902920402

Darici, O., & Kuo, A. D. (2023). Humans plan for the near future to walk economically on uneven terrain. *Proceedings of the National Academy of Sciences*, *120*(19), e2211405120. https://doi.org/10.1073/pnas.2211405120

Davidson, M. J., Keys, R. T., Szekely, B., MacNeilage, P., Verstraten, F., & Alais, D. (2023). Continuous peripersonal tracking accuracy is limited by the speed and phase of locomotion. *Scientific Reports*, *13*(1), 14864. https://doi.org/10.1038/s41598-023-40655-y

Davidson, M. J., Verstraten, F. A. J., & Alais, D. (2024). Walking modulates visual detection performance according to stride cycle phase. *Nature Communications*, *15*(1), 2027. https://doi.org/10.1038/s41467-024-45780-4

Davis, A., & Kühnlenz, F. (2007). Optical design using fresnel lenses. *Optik & Photonik*, *2*(4), 52–55. https://doi.org/10.1002/opph.201190287

Davis, J., Hsieh, Y.-H., & Lee, H.-C. (2015). Humans perceive flicker artifacts at 500 hz. *Scientific Reports*, *5*(1), 7861. https://doi.org/10.1038/srep07861

Davson, H. (1990). *Physiology of the eye*. Bloomsbury Academic. https://doi.org/10.1007/978-1-349-09997-9

De Moraes, C. G. (2013). Anatomy of the visual pathways. *Journal of Glaucoma*, *22*. https://doi.org/10.1097/IJG.0b013e3182934978

Delabarre, E. B. (1898). A method of recording eye-movements. *The American Journal of Psychology*, *9*(4), 572–574. Retrieved September 19, 2024, from http://www.jstor.org/stable/1412191

Deubel, H., Schneider, W. X., & Bridgeman, B. (1996). Postsaccadic target blanking prevents saccadic suppression of image displacement. *Vision research*, *36*(7), 985–996.

DeYoe, E., & Van Essen, D. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, *11*(5), 219–226. https://doi.org/10.1016/0166-2236(88)90130-0

Dhawale, N., & Venkadesan, M. (2023). How human runners regulate footsteps on uneven terrain (M. A. Daley, T. Moore, & A. A. Biewener, Eds.). *eLife*, *12*, e67177. https://doi.org/10.7554/eLife.67177

Di Stasi, L. L., McCamy, M. B., Catena, A., Macknik, S. L., Cañas, J. J., & Martinez-Conde, S. (2013). Microsaccade and drift dynamics reflect mental fatigue. *European Journal of Neuroscience*, *38*(3), 2389–2398. https://doi.org/10.1111/ejn.12248

Diamond, M. R., Ross, J., & Morrone, M. C. (2000). Extraretinal control of saccadic suppression. *Journal of Neuroscience*, *20*(9), 3449–3455. https://doi.org/10.1523/JNEUROSCI.20-09-03449.2000

Ditchburn, R. W. (1955). Eye-movements in relation to retinal action. *Optica Acta: International Journal of Optics*, *1*(4), 171–176. https://doi.org/10.1080/713818684

Ditchburn, R. W., & Ginsborg, B. L. (1953). Involuntary eye movements during fixation. *The Journal of Physiology*, *119*(1), 1–17. https://doi.org/10.1113/jphysiol.1953.sp004824

Dodge, R., & Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review*, *8*(2), 145–157. https://doi.org/10.1037/h0076100

Donk, M., & van Zoest, W. (2008). Effects of salience are short-lived. *Psychological Science*, *19*(7), 733–739. https://doi.org/10.1111/j.1467-9280.2008.02149.x

Dorr, M., & Bex, P. J. (2011). A gaze-contingent display to study contrast sensitivity under natural viewing conditions. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human vision and electronic imaging xvi* (78650Y, Vol. 7865). Spie. https://doi.org/10.1117/12.872502

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(10), 28–28. https://doi.org/10.1167/10.10.28

Dorris, M. C., & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*, *44*(2), 365–378. https://doi.org/10.1016/j.neuron.2004.09.009

Doughty, M. J. (2001). Consideration of three types of spontaneous eyeblink activity in normal humans: During reading and video display terminal use, in primary gaze, and while in conversation. *Optometry and Vision Science*, *78*(10). https://j ournals.lww.com/optvissci/fulltext/2001/10000/consideration_of_three_types_of_spontaneous.11.aspx

Drewes, J., Feder, S., & Einhäuser, W. (2021). Gaze during locomotion in virtual reality and the real world. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.656913

Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of experimental psychology. Human perception and performance*, *31*(6), 1416–1438. https://doi.org/10.1037/0096-1523.31.6.1416

Duchowski, A. T. (2007). *Eye tracking methodology*. Springer London. https://doi.org/10.1007/978-1-84628-609-4

Duchowski, A. T., Cournia, N., & Murphy, H. (2004). Gaze-contingent displays: A review. *CyberPsychology & Behavior*, *7*(6), 621–634. https://doi.org/10.1089/cpb.2004.7.621

Dumoulin, S. O., & Hess, R. F. (2007). Cortical specialization for concentric shape processing. *Vision Research*, *47*(12), 1608–1613. https://doi.org/10.1016/j.visres.2007.01.031

Durant, S., & Zanker, J. M. (2020). The combined effect of eye movements improve head centred local motion information during walking. *PloS one*, *15*(1), e0228345. https://doi.org/10.1371/journal.pone.0228345

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2), 2–2. https://doi.org/10.1167/8.2.2

Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E., & König, P. (2007). Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, *18*(3), 267–297. https://doi.org/10.1080/09548980701671094

Epelboim, J., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research*, *35*(23), 3401–3422. https://doi.org/10.1016/0042-6989(95)00080-X

Epelboim, J., Steinman, R. M., Kowler, E., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1997). Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision Research*, *37*(18), 2597–2607. https://doi.org/10.1016/S0042-6989(97)00075-8

Evans, K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1476–92. https://doi.org/10.1037/0096-1523.31.6.1476

Evinger, C. (1995). A brain stem reflex in the blink of an eye. *Physiology*, *10*(4), 147–153. https://doi.org/10.1152/physiologyonline.1995.10.4.147

Ezenman, M., Hallett, P., & Frecker, R. (1985). Power spectra for ocular drift and tremor. *Vision Research*, *25*(11), 1635–1640. https://doi.org/10.1016/0042-6989(85)90134-8

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: A priority map for target selection. *Trends in cognitive sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Feng, X., Liu, Y., & Wei, S. (2020). Livedeep: Online viewport prediction for live virtual reality streaming using lifelong deep learning. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 800–808. https://doi.org/10.1109/VR46266.2020.00104

Ferree, C. E., Rand, G., & Harris, E. T. (1933). Intensity of light and area of illuminated field as interacting factors in size of pupil. *Journal of Experimental Psychology*, *16*(3), 408. https://doi.org/10.1037/h0072100

Fetter, M. (2007). Vestibulo-ocular reflex. *Developments in ophthalmology*, *40*, 35–51. https://doi.org/10.1159/000100348

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research*, *37*(5), 617–631. https://doi.org/10.1016/S0042-6989(96)00218-0

Fink, P. W., Foo, P. S., & Warren, W. H. (2007). Obstacle avoidance during walking in real and virtual environments. *ACM Transactions on Applied Perception (TAP)*, *4*(1), 2–es. https://doi.org/10.1145/1227134.1227136

Foulsham, T. (2015). Eye movements and their functions in everyday tasks. *Eye*, *29*(2), 196–199. https://doi.org/10.1038/eye.2014.275

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research*, *48*(17), 1777–1790. https://doi.org/10.1016/j.visres.2008.05.018

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception*, *36*(8), 1123–1138. https://doi.org/10.1068/p5659

Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, *50*(24), 2766–2774. https://doi.org/10.1016/j.visres.2010.09.024

Friedman, L., & Komogortsev, O. V. (2025). Fixation drift increases as a function of time-on-task in a brief saccade tracking study. *PLOS ONE*, *20*(6), 1–17. https://doi.org/10.1371/journal.pone.0310619

Fujita, M., Amagai, A., Minakawa, F., & Aoki, M. (2002). Selective and delay adaptation of human saccades. *Cognitive Brain Research*, *13*(1), 41–52.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings. Biological sciences*, *266*(1436), 2435–2440. https://doi.org/10.1098/rspb.1999.0943

Gajewski, D. A., Pearson, A. M., Mack, M. L., Bartlett, F. N., & Henderson, J. M. (2005). Human gaze control in real world search. In L. Paletta, J. K. Tsotsos, E. Rome, & G. Humphreys (Eds.), *Attention and performance in computational vision* (pp. 83–99). Springer Berlin Heidelberg.

Gandrud, J., & Interrante, V. (2016). Predicting destination using head orientation and gaze direction during locomotion in VR. *ACM Symposium on Applied Perception, SAP 2016*, 31–38. https://doi.org/10.1145/2931002.2931010

Gardner, R. A. (1957). Probability-learning with two and three choices. *The American Journal of Psychology*, *70*(2), 174–185. https://doi.org/10.2307/1419319

Gaspar, J. G., Ward, N., Neider, M. B., Crowell, J., Carbonari, R., Kaczmarski, H., Ringer, R. V., Johnson, A. P., Kramer, A. F., & Loschky, L. C. (2016). Measuring the useful field of view during simulated driving with gaze-contingent displays. *Human Factors*, *58*(4), 630–641. https://doi.org/10.1177/0018720816642092

Geisler, W. S., & Perry, J. S. (2002). Real-time simulation of arbitrary visual fields. *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, 83–87. https://doi.org/10.1145/507072.507090

Genç, Ç., Soomro, S., Duyan, Y., Ölçer, S., Balcı, F., Ürey, H., & Özcan, O. (2016). Head mounted projection display & visual attention: Visual attentional processing of head referenced static and dynamic displays while in motion and standing. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1538–1547. https://doi.org/10.1145/2858036.2858449

Georg, K., & Lappe, M. (2007). Spatio-temporal contingency of saccade-induced chronostasis. *Experimental brain research*, *180*(3), 535–539.

Geruschat, D. R., Turano, K. A., & Stahl, J. W. (1998). Traditional measures of mobility performance and retinitis pigmentosa. *Optometry and Vision Science*, *75*(7). https://doi.org/10.1097/00006324-199807000-00022

Ghiani, A., Van Hout, L. R., Driessen, J. G., & Brenner, E. (2023). Where do people look when walking up and down familiar staircases? *Journal of Vision*, *23*(1), 7–7. https://doi.org/10.1167/jov.23.1.7

Gibaldi, A., & Banks, M. S. (2019). Binocular eye movements are adapted to the natural environment. *Journal of Neuroscience*, *39*(15), 2877–2888. https://doi.org/10.1523/JNEUROSCI.2591-18.2018

Gibaldi, A., & Sabatini, S. P. (2021). The saccade main sequence revised: A fast and repeatable tool for oculomotor analysis. *Behavior Research Methods*, *53*(1), 167–187. https://doi.org/10.3758/s13428-020-01388-2

Gibaldi, A., Vanegas, M., Bex, P. J., & Maiello, G. (2017). Evaluation of the Tobii EyeX eye tracking controller and matlab toolkit for research. *Behavior research methods*, *49*(3), 923–946. https://doi.org/10.3758/s13428-016-0762-9

Gilchrist, I. D., Brown, V., & Findlay, J. M. (1997). Saccades without eye movements. *Nature*, *390*(6656), 130–131. https://doi.org/10.1038/36478

Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual Cognition*, *14*(4-8), 704–715. https://doi.org/10.1080/13506280500193719

Goldberg, M. E., & Wurtz, R. H. (1972). Activity of superior colliculus in behaving monkey. ii. effect of attention on neuronal responses. *Journal of Neurophysiology*, *35*(4), 560–574. https://doi.org/10.1152/jn.1972.35.4.560

Grasso, R., Prévost, P., Ivanenko, Y. P., & Berthoz, A. (1998). Eye-head coordination for the steering of locomotion in humans: An anticipatory synergy. *Neuroscience Letters*, *253*(2), 115–118. https://doi.org/10.1016/S0304-3940(98)00625-9

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001

Gremmler, S., & Lappe, M. (2017). Saccadic suppression during voluntary versus reactive saccades. *Journal of Vision*, *17*(8), 8–8. https://doi.org/10.1167/17.8.8

Gresty, M. (1974). Coordination of head and eye movements to fixate continuous and intermittent targets. *Vision Research*, *14*(6), 395–403. https://doi.org/10.1016/0042-6989(74)90238-7

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*, 649–677. https://doi.org/https://doi.org/10.1146/annurev.neuro.27.070203.144220

Groner, R., & Groner, M. T. (1989). Attention and eye movement control: An overview. *European archives of psychiatry and neurological sciences*, *239*(1), 9–16. https://doi.org/10.1007/BF01739737

Gulhan, D., Durant, S., & Zanker, J. M. (2021). Similarity of gaze patterns across physical and virtual versions of an installation artwork. *Scientific Reports*, *11*(1), 18913. https://doi.org/10.1038/s41598-021-91904-x

Haarmeier, T., & Thier, P. (1999). Impaired analysis of moving objects due to deficient smooth pursuit eye movements. *Brain*, *122*(8), 1495–1505. https://doi.org/10.1093/brain/122.8.1495

Haarmeier, T., Thier, P., Repnow, M., & Petersen, D. (1997). False perception of motion in a patient who cannot compensate for eye movements. *Nature*, *389*(6653), 849–852. https://doi.org/10.1038/39872

Hacques, G., Dicks, M., Komar, J., & Seifert, L. (2022). Visual control during climbing: Variability in practice fosters a proactive gaze pattern. *PLOS ONE*, *17*(6), 1–23. https://doi.org/10.1371/journal.pone.0269794

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011). Alpha-oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences*, *108*(48), 19377–19382. https://doi.org/10.1073/pnas.1117190108

Haitao, H., Ruijun, D., Na, H., Jiarong, B., Yulong, W., Ke, L., Chenru, W., Zhanshan, M., Lili, C., & hao, Z. (2022). Research on stray light affecting the imaging of Fresnel lens in virtual reality equipment. *SID Symposium Digest of Technical Papers*, *53*(1), 827–829. https://doi.org/10.1002/sdtp.15620

Hammock, B. (2015). How a film projector works. *Youtube*. https://www.youtube.com/watch?v=En__V0oEJsU&t=250s

Hansen, K. A., Kay, K. N., & Gallant, J. L. (2007). Topographic organization in and near human visual area V4. *Journal of Neuroscience*, *27*(44), 11896–11911. https://doi.org/10.1523/JNEUROSCI.2991-07.2007

Harezlak, K., Kasprowski, P., & Stasch, M. (2014). Towards accurate eye tracker calibration – methods and procedures. *Procedia Computer Science*, *35*, 1073–1081. https://doi.org/10.1016/j.procs.2014.08.194

Harris, L. R. (1994). Visual motion caused by movements of the eye, head and body. *Visual detection of motion*, *1*, 397–435. https://www.yorku.ca/harris/pubs/1994visualmotioncausedbymovementsofeyeheadandbody.pdf

Hashiba, M., Matsuoka, T., Baba, S., & Watanabe, S. (1996). Non-visually induced smooth pursuit eye movements using sinusoidal target motion. *Acta otolaryngologica. Supplementum*, *525*, 158–162. http://europepmc.org/abstract/MED/8908293

Haskins, A., Mentch, J., Botch, T., & Robertson, C. E. (2020). Active vision in immersive, 360° real-world environments. *Scientific Reports*, *10*, 14304. https://doi.org/10.1038/s41598-020-71125-4

Hassan, S. E., Lovie-Kitchin, J. E., Woods, & L., R. (2002). Vision and mobility performance of subjects with age-related macular degeneration. *Optometry and Vision Science*, *79*(11). https://doi.org/10.1097/00006324-200211000-00007

Hayhoe, M. M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*(1-3), 43–64. https://doi.org/10.1080/135062800394676

Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci.*, *9*(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, *38*(1), 125–137. https://doi.org/10.1016/S0042-6989(97)00116-8

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 6–6. https://doi.org/10.1167/3.1.6

Hegdé, J., & Van Essen, D. C. (2005). Role of primate visual area V4 in the processing of 3-d shape characteristics defined by disparity. *Journal of Neurophysiology*, *94*(4), 2856–2866. https://doi.org/10.1152/jn.00802.2004

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In *Eye movements* (pp. 537–562). Elsevier. https://doi.org/10.1016/B978-008044980-7/50027-6

Herwig, A., & Schneider, W. X. (2014). Predicting object features across saccades: Evidence from object recognition and visual search. *Journal of Experimental Psychology: General*, *143*(5), 1903–1922. https://doi.org/10.1037/a0036781

Hessels, R. S., Kemner, C., van den Boomen, C., & Hooge, I. T. (2016). The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior research methods*, *48*(4), 1694–1712. https://doi.org/10.3758/s13428-015-0676-y

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. (2017). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior research methods*, *49*(5), 1802–1823. https://doi.org/10.3758/s13428-016-0822-1

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society open science*, *5*(8), 180502. https://doi.org/10.1098/rsos.180502

Hibbard, P. B., van Dam, L. C., & Scarfe, P. (2020). The implications of interpupillary distance variability for virtual reality. *2020 International Conference on 3D Immersion (IC3D)*, 1–7. https://doi.org/10.1109/IC3D51119.2020.9376369

Higuchi, Y., Inoue, S., Endo, T., & Kumada, T. (2019). Task-irrelevant optic flow guides attention in visual search. *Attention, Perception, & Psychophysics*, *81*(5), 1327–1345. https://doi.org/10.3758/s13414-018-01646-8

Hinkle, D. A., & Connor, C. E. (2001). Disparity tuning in macaque area V4. *NeuroReport*, *12*(2). https://doi.org/10.1097/00001756-200102120-00036

Hinkle, D. A., & Connor, C. E. (2002). Three-dimensional orientation tuning in macaque area V4. *Nature Neuroscience*, *5*(7), 665–670. https://doi.org/10.1038/nn875

Hirasaki, E., Moore, S. T., Raphan, T., & Cohen, B. (1999). Effects of walking velocity on vertical head and body movements during locomotion. *Experimental Brain Research*, *127*(2), 117–130. https://doi.org/10.1007/s002210050781

Hirsch, J., & Curcio, C. A. (1989). The spatial resolution capacity of human foveal retina. *Vision Research*, *29*(9), 1095–1101. https://doi.org/10.1016/0042-6989(89)90058-8

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hollands, M. A., & Marple-Horvat, D. E. (1996). Visually guided stepping under conditions of step cycle-related denial of visual information. *Experimental brain research*, *109*(2), 343–356. https://doi.org/10.1007/BF00231792

Hollands, M. A., Marple-Horvat, D. E., Henkes, S., & Rowan, A. K. (1995). Human eye movements during visually guided stepping. *Journal of motor behavior*, *27*(2), 155–163. https://doi.org/10.1080/00222895.1995.9941707

Hollands, M. A., Patla, A. E., & Vickers, J. N. (2002). "Look where you're going!": Gaze behaviour associated with maintaining and changing the direction of locomotion. *Experimental brain research*, *143*(2), 221–230. https://doi.org/10.1007/s00221-001-0983-7

Hollands, M. A., Sorensen, K., & Patla, A. (2001). Effects of head immobilization on the coordination and control of head and body reorientation and translation during steering. *Experimental Brain Research*, *140*(2), 223–233. https://doi.org/10.1007/s002210100811

Hollands, M. A., Ziavra, N. V., & Bronstein, A. M. (2004). A new paradigm to investigate the roles of head and eye movements in the coordination of whole-body movements. *Experimental Brain Research*, *154*(2), 261–266. https://doi.org/10.1007/s00221-003-1718-8

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 45–52. https://doi.org/10.1145/2168556.2168563

Holt, E. B. (1903). Eye movement and central anaesthesia. the problem of anaesthesia during eye-movement. *Psychol. Monographs*, *4*, 3–46.

Holzwarth, V., Gisler, J., Hirt, C., & Kunz, A. (2021). Comparing the accuracy and precision of SteamVR tracking 2.0 and Oculus Quest 2 in a room scale setup. *Proceedings of the 2021 5th International Conference on Virtual and Augmented Reality Simulations*, 42–46. https://doi.org/10.1145/3463914.3463921

Hooge, I. T. C., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2018). Is human classification by experienced untrained observers a gold standard in fixation detection? *Behavior Research Methods*, *50*(5), 1864–1881. https://doi.org/10.3758/s13428-017-0955-x

Hou, B. J., Abramyan, L., Gurumurthy, P., Adams, H., Tosic Rodgers, I., Gonzalez, E. J., Patel, K., Colaço, A., Pfeuffer, K., Gellersen, H., Ahuja, K., & Gonzalez-Franco, M. (2025). Online-EYE: Multimodal implicit eye tracking calibration for XR. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3706598.3713461

Hu, Z., Bulling, A., Li, S., & Wang, G. (2021). Fixationnet: Forecasting eye fixations in task-oriented virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, *27*(5), 2681–2690. https://doi.org/10.1109/TVCG.2021.3067779

Hu, Z., Li, S., Zhang, C., Yi, K., Wang, G., & Manocha, D. (2020). DGaze: CNN-based gaze prediction in dynamic scenes. *IEEE Transactions on Visualization and Computer Graphics*, *26*(5), 1902–1911. https://doi.org/10.1109/TVCG.2020.2973473

Huey, E. B. (1898). Preliminary experiments in the physiology and psychology of reading. *The American Journal of Psychology*, *9*(4), 575–586. Retrieved September 19, 2024, from http://www.jstor.org/stable/1412192

Hulleman, J., & Olivers, C. N. L. (2017). The impending demise of the item in visual search. *Behavioral and Brain Sciences*, *40*, e132. https://doi.org/10.1017/S0140525X15002794

Hutton, C., & Suma, E. (2016). A realistic walking model for enhancing redirection in virtual reality. *2016 IEEE Virtual Reality (VR)*, 183–184. https://doi.org/10.1109/VR.2016.7504714

Hutton, S. B. (2019). Eye tracking methodology. In C. Klein & U. Ettinger (Eds.), *Eye movement research: An introduction to its scientific foundations and applications* (pp. 277–308). Springer International Publishing. https://doi.org/10.1007/978-3-030-20085-5_8

Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, *51*(10), 1192–1205. https://doi.org/10.1016/j.visres.2011.03.010

Ilg, U. J., & Hoffmann, K. P. (1993). Motion perception during saccades. *Vision research*, *33*(2), 211–220. https://doi.org/10.1016/0042-6989(93)90159-T

Imai, T., Moore, S. T., Raphan, T., & Cohen, B. (2001). Interaction of the body, head, and eyes during walking and turning. *Experimental Brain Research*, *136*(1), 1–18. https://doi.org/10.1007/s002210000533

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*(3), 420–456. https://doi.org/10.1016/0010-0285(91)90015-G

Irwin, D. E., Colcombe, A. M., Kramer, A. F., & Hahn, S. (2000). Attentional and oculomotor capture by onset, luminance and color singletons. *Vision Research*, *40*(10), 1443–1458. https://doi.org/10.1016/S0042-6989(00)00030-4

James, G., & Koehler, D. J. (2011). Banking on a bad bet: Probability matching in risky choice is linked to expectation generation. *Psychological Science*, *22*(6), 707–711. https://doi.org/10.1177/0956797611407933

Jeon, S.-B., Jung, J., Park, J., & Lee, I.-K. (2025). F-rdw: Redirected walking with forecasting future position. *IEEE Transactions on Visualization and Computer Graphics*, *31*(4), 1970–1984. https://doi.org/10.1109/TVCG.2024.3376080

Jia, J., Chan, T. T., Lian, T., & Rio, K. W. (2023). Local pupil swim in virtual- and augmented-reality: Root cause and perception model. *Journal of the Society for Information Display*, *31*(5), 230–240. https://doi.org/10.1002/jsid.1210

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye–hand coordination in object manipulation. *Journal of Neuroscience*, *21*(17), 6917–6932. https://doi.org/10.1523/JNEUROSCI.21-17-06917.2001

Jolly, P., Li, Y. H., Cox, M. A., Clark, A. M., Yang, B., Lin, R., Zhao, Z., & Rucci, M. (2023). Microsaccades in head-free high-acuity tasks. *Journal of Vision*, *23*(9), 5817–5817. https://doi.org/10.1167/jov.23.9.5817

Joukal, M. (2017). Anatomy of the human visual pathway. In K. Skorkovská (Ed.), *Homonymous visual field defects* (pp. 1–16). Springer International Publishing. https://doi.org/10.1007/978-3-319-52284-5_1

Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *Journal of Neuroscience*, *29*(19), 6234–6238. https://doi.org/10.1523/JNEUROSCI.5570-08.2009

Judd, C. H., McAllister, C. N., & Steele, W. (1905). General introduction to a series of studies of eye movements by means of kinetoscopic photographs. *Psychological Review Monographs*, *7*(1), 1–16.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*(6-7), 979–1003. https://doi.org/10.1080/13506280902771138

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, *107*(25), 11163–11170. https://doi.org/10.1073/pnas.1005062107

Karson, C. N., Berman, K. F., Donnelly, E. F., Mendelson, W. B., Kleinman, J. E., & Wyatt, R. J. (1981). Speaking, thinking, and blinking. *Psychiatry Research*, *5*(3), 243–246. https://doi.org/10.1016/0165-1781(81)90070-6

Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology*, *86*(3), 1398–1411. https://doi.org/10.1152/jn.2001.86.3.1398

Katrychuk, D., Griffith, H. K., & Komogortsev, O. V. (2019). Power-efficient and shift-robust eye-tracking sensor for portable VR headsets. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. https://doi.org/10.1145/3314111.3319821

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, *3*(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*(500), 1590–1598. https://doi.org/10.1080/01621459.2012.737745

Kim, S., Nussbaum, M. A., & Ulman, S. (2018). Impacts of using a head-worn display on gait performance during level walking and obstacle crossing. *Journal of Electromyography and Kinesiology*, *39*, 142–148. https://doi.org/10.1016/j.jelekin.2018.02.007

Kim, Y., Hwang, S., Oh, J., & Kim, S. (2024). Gaitway: Gait data-based VR locomotion prediction system robust to visual distraction. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613905.3651073

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. https://hdl.handle.net/11245/1.505367

Kirchner, J., Watson, T., Busch, N. A., & Lappe, M. (2022). Timing and kinematics of horizontal within-blink saccades measured by EOG. *Journal of Neurophysiology*, *127*(6), 1655–1668. https://doi.org/10.1152/jn.00076.2022

Kirchner, J., Watson, T., & Lappe, M. (2022). Real-time MRI reveals unique insight into the full kinematics of eye movements. *eNeuro*, *9*(1). https://doi.org/10.1523/ENEURO.0357-21.2021

Kit, D., Katz, L., Sullivan, B., Snyder, K., Ballard, D., & Hayhoe, M. M. (2014). Eye movements, visual search and scene memory, in an immersive virtual environment. *PLOS ONE*, *9*(4), 1–11. https://doi.org/10.1371/journal.pone.0094362

Koerfer, K., Watson, T., & Lappe, M. (2024). Inability to pursue nonrigid motion produces instability of spatial perception. *Science Advances*, *10*(45), eadp6204. https://doi.org/10.1126/sciadv.adp6204

Komogortsev, O. V., Ryu, Y. S., & Koh, D. H. (2009). Quick models for saccade amplitude prediction. *Journal of Eye Movement Research*, *3*(1), 1–13. https://doi.org/10.16910/jemr.3.1.1

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, *11*(2), 224–231. https://doi.org/10.1038/nn2036

Konrad, R., Angelopoulos, A., & Wetzstein, G. (2020). Gaze-contingent ocular parallax rendering for virtual reality. *ACM Trans. Graph.*, *39*(2). https://doi.org/10.1145/3361330

Koppula, H. S., & Saxena, A. (2015). Anticipating human activities using object affordances for reactive robotic response. *IEEE transactions on pattern analysis and machine intelligence*, *38*(1), 14–29. https://doi.org/10.1109/TPAMI.2015.2430335

Kowacs, P., Piovesan, E., Werneck, L., Fameli, H., & da Silva, H. P. (2004). Headache related to a specific screen flickering frequency band. *Cephalalgia*, *24*(5), 408–410. https://doi.org/10.1111/j.1468-2982.2004.00686.x

Kowacs, P., Piovesan, E., Werneck, L., Fameli, H., Zani, A., & da Silva, H. (2005). Critical flicker frequency in migraine. A controlled study in patients without prophylactic therapy. *Cephalalgia*, *25*(5), 339–343. https://doi.org/10.1111/j.1468-2982.2004.00861.x

Kowalsky, D. B., Rebula, J. R., Ojeda, L. V., Adamczyk, P. G., & Kuo, A. D. (2021). Human walking in the real world: Interactions between terrain type, gait parameters, and energy expenditure. *PLOS ONE*, *16*(1), 1–14. https://doi.org/10.1371/journal.pone.0228682

Krauzlis, R. J., Lovejoy, L. P., & Zénon, A. (2013). Superior colliculus and visual spatial attention. *Annu Rev Neurosci*, *36*, 165–182. https://doi.org/10.1146/annurev-neuro-062012-170249

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*(6483), 742–744. https://doi.org/10.1038/369742a0

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature neuroscience*, *3*(12), 1340–1345. https://doi.org/10.1038/81887

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual neuroscience*, *26*(1), 51–62. https://doi.org/10.1017/s0952523808080899

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision research*, *41*(25-26), 3559–3565. https://doi.org/10.1016/S0042-6989(01)00102-X

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328. https://doi.org/10.1068/p2935

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology*, *11*(15), 1215–1220. https://doi.org/10.1016/S0960-9822(01)00351-7

Land, M. F., & Tatler, B. W. (2009a). *Looking and acting: Vision and eye movements in natural behaviour.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198570943.001.0001

Land, M. F., & Tatler, B. W. (2009b, July). Locomotion on foot. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198570943.003.0006

Langbehn, E., Steinicke, F., Lappe, M., Welch, G. F., & Bruder, G. (2018). In the blink of an eye: leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Trans. Graph.*, *37*(4), 66:1–66:11. https://doi.org/10.1145/3197517.3201335

Langbehn, E., Lubos, P., & Steinicke, F. (2018). Evaluation of locomotion techniques for room-scale VR: Joystick, teleportation, and redirected walking. *Proceedings of the Virtual Reality International Conference-Laval Virtual*, 1–9. https://doi.org/10.1145/3234253.3234291

Lappe, M., Awater, H., & Krekelberg, B. (2000). Postsaccadic visual references generate presaccadic compression of space. *Nature*, *403*(6772), 892–895. https://doi.org/10.1038/35002588

Lappi, O. (2015). Eye tracking in the wild: The good, the bad and the ugly. *Journal of Eye Movement Research*, *8*(5). https://doi.org/10.16910/jemr.8.5.1

Latour, P. (1962). Visual threshold during eye movements. *Vision Research*, *2*(7), 261–262. https://doi.org/10.1016/0042-6989(62)90031-7

Li, C. L., Aivar, M. P., Kit, D. M., Tong, M. H., & Hayhoe, M. M. (2016). Memory and visual search in naturalistic 2d and 3d environments. *Journal of vision*, *16*, 9. https://doi.org/10.1167/16.8.9

Li, C. L., Aivar, M. P., Tong, M. H., & Hayhoe, M. M. (2018). Memory shapes visual search strategies in large-scale environments. *Scientific Reports*, *8*, 4324. https://doi.org/10.1038/s41598-018-22731-w

Lin, J. Y., Franconeri, S., & Enns, J. T. (2008). Objects on a collision path with the observer demand attention. *Psychological Science*, *19*(7), 686–692. https://doi.org/10.1111/j.1467-9280.2008.02143.x

Lisberger, S. G. (2015). Visual guidance of smooth pursuit eye movements. *Annual Review of Vision Science*, *1*(Volume 1, 2015), 447–468. https://doi.org/https://doi.org/10.1146/annurev-vision-082114-035349

Liu, Z., Lin, C.-H., Hyun, B.-R., Sher, C.-W., Lv, Z., Luo, B., Jiang, F., Wu, T., Ho, C.-H., Kuo, H.-C., & He, J.-H. (2020). Micro-light-emitting diodes with quantum dots in display technology. *Light: Science & Applications*, *9*(1), 83. https://doi.org/10.1038/s41377-020-0268-1

Lo, A. W., Marlowe, K. P., & Zhang, R. (2021). To maximize or randomize? an experimental study of probability matching in financial decision making. *PLOS ONE*, *16*(8), 1–20. https://doi.org/10.1371/journal.pone.0252540

Lohr, D. J., Friedman, L., & Komogortsev, O. V. (2019). Evaluating the data quality of eye tracking signals from a virtual reality system: Case study using smi's eye-tracking htc vive. *arXiv preprint arXiv:1912.02083*. https://doi.org/10.48550/arXiv.1912.02083

Longstaffe, K. A., Hood, B. M., & Gilchrist, I. D. (2014). The influence of cognitive load on spatial search performance. *Attention, Perception, & Psychophysics*, *76*(1), 49–63. https://doi.org/10.3758/s13414-013-0575-1

Loschky, L. C., & McConkie, G. W. (2000). User performance with gaze contingent multiresolutional displays. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 97–103. https://doi.org/10.1145/355017.355032

Loschky, L. C., & Wolverton, G. S. (2007). How late can you update gaze-contingent multiresolutional displays without detection? *ACM Transactions on Multimedia Computing, Communications, and Applications*, *3*(4). https://doi.org/10.1145/1314303.1314310

Lukashova-Sanz, O., Agarwala, R., & Wahl, S. (2022). Context matters during pick-and-place in VR: Impact on search and transport phases. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.881269

Lukashova-Sanz, O., & Wahl, S. (2021). Saliency-aware subtle augmentation improves human visual search performance in VR. *Brain Sciences*, *11*(3). https://doi.org/10.3390/brainsci11030283

Luo, Z., Ding, Y., Yang, Q., & Wu, S.-T. (2024). Ghost image analysis for pancake virtual reality systems. *Opt. Express*, *32*(10), 17211–17219. https://doi.org/10.1364/OE.523196

Macele, P., & Mueggenburg, J. (2024). Playing with the eyes. a media history of eye tracking. In M. Spöhrer & B. Ochsner (Eds.), *Disability and video games: Practices of en-/disabling modes of digital gaming* (pp. 117–143). Springer International Publishing. https://doi.org/10.1007/978-3-031-34374-2_5

Mack, D. J., Belfanti, S., & Schwarz, U. (2017). The effect of sampling rate and lowpass filters on saccades–a modeling approach. *Behavior Research Methods*, *49*(6), 2146–2162. https://doi.org/10.3758/s13428-016-0848-4

Magnusson, M., Pyykkö, I., & and, B. N. (1986). The relationship of optokinetic nystagmus to pursuit eye movements, vestibular nystagmus and to saccades in humans: A clinical study. *Acta Oto-Laryngologica*, *101*(5-6), 361–370. https://doi.org/10.3109/00016488609108620

Maiello, G., Harrison, W. J., & Bex, P. J. (2016). Monocular and binocular contributions to oculomotor plasticity. *Scientific reports*, *6*, 31861. https://doi.org/10.1038/srep31861

Manakhov, P., Sidenmark, L., Pfeuffer, K., & Gellersen, H. (2024). Gaze on the go: Effect of spatial reference frame on visual target acquisition during physical locomotion in extended reality. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3613904.3642915

Mankowska, N. D., Marcinkowska, A. B., Waskow, M., Sharma, R. I., Kot, J., & Winklewski, P. J. (2021). Critical flicker fusion frequency: A narrative review. *Medicina*, *57*(10). https://doi.org/10.3390/medicina57101096

Marigold, D. S., & Patla, A. E. (2008). Visual information from the lower visual field is important for walking across multi-surface terrain. *Experimental Brain Research*, *188*(1), 23–31. https://doi.org/10.1007/s00221-008-1335-7

Marigold, D., & Patla, A. (2007). Gaze fixation patterns for negotiating complex ground terrain. *Neuroscience*, *144*(1), 302–313. https://doi.org/10.1016/j.neuroscience.2006.09.006

Marsh, D. (2017). Temporal rate conversion. Retrieved February 21, 2025, from https://learn.microsoft.com/en-us/previous-versions/windows/hardware/design/dn642112(v=vs.85)?redirectedfrom=MSDN

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(3), 229–240. https://doi.org/10.1038/nrn1348

Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: Towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, *14*(2), 83–96. https://doi.org/10.1038/nrn3405

Matin, L., & Pearce, D. G. (1965). Visual perception of direction for stimuli flashed during voluntary saccadic eye movements. *Science*, *148*(3676), 1485–1488. https://doi.org/10.1126/science.148.3676.1485

Matsumiya, K., & Uchikawa, K. (2003). The role of presaccadic compression of visual space in spatial remapping across saccadic eye movements. *Vision Research*, *43*(18), 1969–1981. https://doi.org/10.1016/S0042-6989(03)00301-8

Matthis, J. S., & Fajen, B. R. (2014). Visual control of foot placement when walking over complex terrain. *Journal of experimental psychology. Human perception and performance*, *40*(1), 106–115. https://doi.org/10.1037/a0033101

Matthis, J. S., Barton, S. L., & Fajen, B. R. (2017). The critical phase for visual control of human walking over complex terrain. *Proceedings of the National Academy of Sciences*, *114*(32), E6720–E6729. https://doi.org/10.1073/pnas.1611699114

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, *28*(8), 1224–1233.e5. https://doi.org/10.1016/j.cub.2018.03.008

Mauderer, M., Flatla, D. R., & Nacenta, M. A. (2016). Gaze-contingent manipulation of color perception. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5191–5202. https://doi.org/10.1145/2858036.2858320

Mayor, J., Calleja, P., & Fuentes-Hurtado, F. (2024). Long short-term memory prediction of user's locomotion in virtual reality. *Virtual Reality*, *28*(1), 65. https://doi.org/10.1007/s10055-024-00962-9

McAnally, K., Grove, P., & Wallis, G. (2024). Vergence eye movements in virtual reality. *Displays*, *83*, 102683. https://doi.org/10.1016/j.displa.2024.102683

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586. https://doi.org/10.3758/BF03203972

McConkie, G. W., Wolverton, G. S., & Zola, D. (1984). Instrumentation considerations in research involving eye-movement contingent stimulus control. *Center for the Study of Reading Technical Report; no. 305*. https://doi.org/10.1016/S0166-4115(08)61816-6

McLaughlin, S. C. (1967). Parametric adjustment in saccadic eye movements. *Perception & Psychophysics*, *2*(8), 359–362. https://doi.org/10.3758/BF03210071

Melnyk, K., Friedman, L., Katrychuk, D., & Komogortsev, O. (2025). Gaze prediction as a function of eye movement type and individual differences. *Proceedings of the 2025 Symposium on Eye Tracking Research and Applications*. https://doi.org/10.1145/3715669.3723116

Mennie, N., Hayhoe, M., & Sullivan, B. (2007). Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research*, *179*(3), 427–442. https://doi.org/10.1007/s00221-006-0804-0

Merchant, J., Morrissette, R., & Porterfield, J. L. (1974). Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE Transactions on Biomedical Engineering*, *BME-21*(4), 309–317. https://doi.org/10.1109/TBME.1974.324318

Miladinović, A., Quaia, C., Kresevic, S., Ajčević, M., Diplotti, L., Michieletto, P., Accardo, A., & Pensiero, S. (2024). High-resolution eye-tracking system for accurate measurement of short-latency ocular following responses: Development and observational study. *JMIR Pediatr Parent*, *7*, e64353. https://doi.org/10.2196/64353

Miller, D. T., Williams, D. R., Morris, G. M., & Liang, J. (1996). Images of cone photoreceptors in the living human eye. *Vision Research*, *36*(8), 1067–1079. https://doi.org/10.1016/0042-6989(95)00225-1

Montag, J. L. (2021). Limited evidence for probability matching as a strategy in probability learning tasks. In K. D. Federmeier (Ed.), *The psychology of learning and motivation* (pp. 233–273, Vol. 74). Academic Press. https://doi.org/10.1016/bs.plm.2021.02.005

Moore, S. T., Hirasaki, E., Cohen, B., & Raphan, T. (1999). Effect of viewing distance on the generation of vertical eye movements during locomotion. *Experimental Brain Research*, *129*(3), 347–361. https://doi.org/10.1007/s002210050903

Moore, S. T., Hirasaki, E., Raphan, T., & Cohen, B. (2001). The human vestibulo-ocular reflex during linear locomotion. *Annals of the New York Academy of Sciences*, *942*(1), 139–147. https://doi.org/10.1111/j.1749-6632.2001.tb03741.x

Morasso, P., Bizzi, E., & Dichgans, J. (1973). Adjustment of saccade characteristics during head movements. *Experimental Brain Research*, *16*, 492–500. https://doi.org/10.1007/BF00234475

Mordi, J. A., & Ciuffreda, K. J. (1998). Static aspects of accommodation: Age and presbyopia. *Vision Research*, *38*(11), 1643–1653. https://doi.org/10.1016/S0042-6989(97)00336-2

Moreno-Arjonilla, J., López-Ruiz, A., Jiménez-Pérez, J. R., Callejas-Aguilera, J. E., & Jurado, J. M. (2024). Eye-tracking on virtual reality: A survey. *Virtual Reality*, *28*(1), 38. https://doi.org/10.1007/s10055-023-00903-y

Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nature Neuroscience*, *8*(7), 950–954. https://doi.org/10.1038/nn1488

Mulavara, A. P., & Bloomberg, J. J. (2003). Identifying head-trunk and lower limb contributions to gaze stabilization during locomotion. *Journal of Vestibular Research*, *12*(5-6), 255–269. https://doi.org/10.3233/VES-2003-125-606

Muller, K. S., Bonnen, K., Shields, S. M., Panfili, D. P., Matthis, J., & Hayhoe, M. M. (2024). Analysis of foothold selection during locomotion using terrain reconstruction (M. Spering & T. Moore, Eds.). *eLife*, *12*, RP91243. https://doi.org/10.7554/eLife.91243

Murphy, H., & Duchowski, A. T. (2002). Perceptual gaze extent & level of detail in VR: Looking outside the box. *ACM SIGGRAPH 2002 Conference Abstracts and Applications*, 228. https://doi.org/10.1145/1242073.1242241

Nadeau, C., & Bengio, Y. (2003). Inference for the generalization error. *Machine learning*, *52*(3), 239–281. https://doi.org/10.1023/A:1024068626366

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, *8*(3), 4–4. https://doi.org/10.1167/8.3.4

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621. https://doi.org/10.1016/j.visres.2005.08.025

Nescher, T., Huang, Y.-Y., & Kunz, A. (2014). Planning redirection techniques for optimal free walking experience using model predictive control. *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, 111–118. https://doi.org/10.1109/3DUI.2014.6798851

Nguyen, A., Wüest, P., & Kunz, A. (2020). Human following behavior in virtual reality. *26th ACM Symposium on Virtual Reality Software and Technology*, 1–3. https://doi.org/10.1145/3385956.3422099

Niehorster, D. C., Hessels, R. S., & Benjamins, J. S. (2020). GlassesViewer: Open-source software for viewing and analyzing data from the Tobii Pro Glasses 2 eye tracker. *Behavior Research Methods*, *52*(3), 1244–1253. https://doi.org/10.3758/s13428-019-01314-1

Niehorster, D. C., Li, L., & Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception*, *8*(3), 2041669517708205. https://doi.org/10.1177/2041669517708205

Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T., Kasneci, E., & Nyström, M. (2020). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods*, *52*(3), 1140–1160. https://doi.org/10.3758/s13428-019-01307-0

Niehorster, D. C., Zemblys, R., Beelders, T., & Holmqvist, K. (2020). Characterizing gaze position signals and synthesizing noise during fixations in eye-tracking data. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01400-9

Nikonorov, A., Skidanov, R., Fursov, V., Petrov, M., Bibikov, S., & Yuzifovich, Y. (2015). Fresnel lens imaging with post-capture image processing. *Proceedings* *of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. https://doi.org/10.1109/CVPRW.2015.7301373

Nilsson, N. C., Peck, T., Bruder, G., Hodgson, E., Serafin, S., Whitton, M., Steinicke, F., & Rosenberg, E. S. (2018). 15 years of research on redirected walking in immersive virtual environments. *IEEE computer graphics and applications*, *38*(2), 44–56. https://doi.org/10.1109/MCG.2018.111125628

Noton, D., & Stark, L. (1971a). Scanpaths in eye movements during pattern perception. *Science*, *171*(3968), 308–311. https://doi.org/10.1126/science.171.3968.308

Noton, D., & Stark, L. (1971b). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*(9), 929–IN8. https://doi.org/10.1016/0042-6989(71)90213-6

Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image viewing–both initially and overall. *Journal of Eye Movement Research*, *2*(2), 1–11. https://doi.org/10.16910/jemr.2.2.2

Nyström, M., Hooge, I. T. C., Hessels, R. S., Andersson, R., Hansen, D. W., Johansson, R., & Niehorster, D. C. (2025). The fundamentals of eye tracking part 3: How to choose an eye tracker. *Behavior Research Methods*, *57*(2), 67. https://doi.org/10.3758/s13428-024-02587-x

Olk, B., Dinu, A., Zielinski, D. J., & Kopper, R. (2018). Measuring visual search and distraction in immersive virtual reality. *Royal Society Open Science*, *5*(5), 172331. https://doi.org/10.1098/rsos.172331

O'Sullivan, C., Dingliana, J., & Howlett, S. (2003). Eye-movements and interactive graphics (J. Hyönä, R. Radach, & H. Deubel, Eds.). https://doi.org/10.1016/B978-044451020-4/50030-X

Otto, A. R., Taylor, E. G., & Markman, A. B. (2011). There are at least two kinds of probability matching: Evidence from a secondary task. *Cognition*, *118*(2), 274–279. https://doi.org/10.1016/j.cognition.2010.11.009

Paeye, C., Collins, T., Cavanagh, P., & Herwig, A. (2018). Calibration of peripheral perception of shape with and without saccadic eye movements. *Attention, perception & psychophysics*, *80*(3), 723–737. https://doi.org/10.3758/s13414-017-1478-3

Palacios-Ibáñez, A., Marín-Morales, J., Contero, M., & Alcañiz, M. (2023). Predicting decision-making in virtual environments: An eye movement analysis with household products. *Applied Sciences*, *13*(12). https://doi.org/10.3390/app13127124

Palmer, E. M., Van Wert, M. J., Horowitz, T. S., & Wolfe, J. M. (2019). Measuring the time course of selection during visual search. *Attention, Perception, & Psychophysics*, *81*(1), 47–60. https://doi.org/10.3758/s13414-018-1596-6

Palmero, C., Komogortsev, O. V., Escalera, S., & Talathi, S. S. (2023). Multi-rate sensor fusion for unconstrained near-eye gaze estimation. *Proceedings of the 2023*

Symposium on Eye Tracking Research and Applications. https://doi.org/10.1145/3588015.3588407

Patla, A. E. (1998). How is human gait controlled by vision. *Ecological Psychology*, *10*(3-4), 287–302. https://doi.org/10.1080/10407413.1998.9652686

Patla, A. E., & Vickers, J. N. (1997). Where and when do we look as we approach and step over an obstacle in the travel path? *Neuroreport*, *8*(17), 3661–3665. https://doi.org/10.1097/00001756-199712010-00002

Patla, A. E., & Vickers, J. N. (2003). How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Experimental Brain Research*, *148*(1), 133–138. https://doi.org/10.1007/s00221-002-1246-y

Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., & Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)*, *35*(6), 179. https://doi.org/10.1145/2980179.2980246

Pélisson, D., Alahyane, N., Panouillères, M., & Tilikete, C. (2010). Sensorimotor adaptation of saccadic eye movements. *Neuroscience and biobehavioral reviews*, *34*(8), 1103–1120. https://doi.org/10.1016/j.neubiorev.2009.12.010

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*(25), 3587–3596. https://doi.org/10.1016/S0042-6989(01)00245-0

Pelz, J. B., & Rothkopf, C. (2007). Oculomotor behavior in natural and man-made environments. In R. P. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements* (pp. 661–676). Elsevier. https://doi.org/10.1016/B978-008044980-7/50033-1

Piumsomboon, T., Lee, G., Lindeman, R. W., & Billinghurst, M. (2017). Exploring natural eye-gaze-based interaction for immersive virtual reality. *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, 36–39. https://doi.org/10.1109/3DUI.2017.7893315

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*(6741), 233–238. https://doi.org/10.1038/22268

Poletti, M., Listorti, C., & Rucci, M. (2013). Microscopic eye movements compensate for nonhomogeneous vision within the fovea. *Current Biology*, *23*(17), 1691–1695. https://doi.org/10.1016/j.cub.2013.07.007

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Prasad, S., & Galetta, S. L. (2011). Chapter 1 - anatomy and physiology of the afferent visual system. In C. Kennard & R. J. Leigh (Eds.), *Neuro-ophthalmology* (pp. 3–19, Vol. 102). Elsevier. https://doi.org/10.1016/B978-0-444-52903-9.00007-8

Press, W. A., Brewer, A. A., Dougherty, R. F., Wade, A. R., & Wandell, B. A. (2001). Visual areas and spatial summation in human visual cortex. *Vision Research*, *41*(10), 1321–1332. https://doi.org/10.1016/S0042-6989(01)00074-8

Ranti, C., Jones, W., Klin, A., & Shultz, S. (2020). Blink rate patterns provide a reliable measure of individual engagement with scene content. *Scientific Reports*, *10*(1), 8267. https://doi.org/10.1038/s41598-020-64999-x

Rashbass, C. (1961). The relationship between saccadic and smooth tracking eye movements. *The Journal of Physiology*, *159*(2), 326–338. https://doi.org/10.1113/jphysiol.1961.sp006811

Rayner, K. (2014). The gaze-contingent moving window in reading: Development and review. *Visual Cognition*, *22*(3-4), 242–258. https://doi.org/10.1080/13506285.2013.879084

Razzaque, S., Kohn, Z., & Whitton, M. C. (2001). Redirected Walking. In *Eurographics 2001 - short presentations*. Eurographics Association. https://doi.org/10.2312/egs.20011036

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological science*, *8*(5), 368–373. https://doi.org/10.1111/j.1467-9280.1997.tb00427

Reynolds, J. H., & Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron*, *37*(5), 853–863. https://doi.org/10.1016/S0896-6273(03)00097-7

Richardson, D., & Spivey, M. (2008, May). Eye-tracking: Characteristics and methods. In *Encyclopedia of biomaterials and biomedical engineering*. https://doi.org/10.1201/b18990-101

Rigas, I., Raffle, H., & Komogortsev, O. V. (2017). Hybrid PS-V technique: A novel sensor fusion approach for fast mobile eye-tracking with sensor-shift aware correction. https://arxiv.org/abs/1707.05411

Robinson, D. A. (1963). A method of measuring eye movemnent using a scieral search coil in a magnetic field. *IEEE Transactions on Bio-medical Electronics*, *10*(4), 137–145. https://doi.org/10.1109/TBMEL.1963.4322822

Roe, A. W., Chelazzi, L., Connor, C. E., Conway, B. R., Fujita, I., Gallant, J. L., Lu, H., & Vanduffel, W. (2012). Toward a unified theory of visual area V4. *Neuron*, *74*(1), 12–29. https://doi.org/10.1016/j.neuron.2012.03.011

Rolff, T., Schmidt, S., Steinicke, F., & Frintrop, S. (2023). A deep learning architecture for egocentric time-to-saccade prediction using weibull mixture-models and historic priors. *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*. https://doi.org/10.1145/3588015.3588408

Rolff, T., Stein, N., Lappe, M., Steinicke, F., & Frintrop, S. (2022). Metrics for time-to-event prediction of gaze events. *NeuRIPS 2022 Workshop on Gaze Meets ML*. https://openreview.net/forum?id=snClL0drE-A

Rolff, T., Steinicke, F., & Frintrop, S. (2022). When do saccades begin? prediction of saccades as a time-to-event problem. *2022 Symposium on Eye Tracking Research and Applications*. https://doi.org/10.1145/3517031.3529627

Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Research*, *49*(20), 2415–2441. https://doi.org/10.1016/j.visres.2009.08.010

Ross, J., Morrone, M. C., & Burr, D. C. (1997). Compression of visual space before saccades. *Nature*, *386*(6625), 598–601. https://doi.org/10.1038/386598a0

Ross, J., Morrone, M., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, *24*(2), 113–121. https://doi.org/10.1016/S0166-2236(00)01685-4

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2016). Task and context determine where you look. *Journal of Vision*, *7*(14), 16–16. https://doi.org/10.1167/7.14.16

Rottach, K. G., Das, V. E., Wohlgemuth, W., Zivotofsky, A. Z., & Leigh, R. J. (1998). Properties of horizontal saccades accompanied by blinks. *Journal of Neurophysiology*, *79*(6), 2895–2902. https://doi.org/10.1152/jn.1998.79.6.2895

Rucci, M., & Poletti, M. (2015). Control and functions of fixational eye movements. *Annual Review of Vision Science*, *1*(Volume 1, 2015), 499–518. https://doi.org/https://doi.org/10.1146/annurev-vision-082114-035742

Rui, C., Zhao, L., Hou, T., Liang, J., & Wang, C. (2023). Pupil swim measurement and analysis of near-eye displays. *SID Symposium Digest of Technical Papers*, *54*(S1), 371–374. https://doi.org/10.1002/sdtp.16306

Sakai, H. (2023). Perception of brightness when the eyes are closed. *Color Research & Application*, *48*(1), 63–68. https://doi.org/10.1002/col.22832

Saldana, C., Claidière, N., Fagot, J., & Smith, K. (2022). Probability matching is not the default decision making strategy in human and non-human primates. *Scientific Reports*, *12*(1), 13092. https://doi.org/10.1038/s41598-022-16983-w

Scarfe, P., & Glennerster, A. (2015). Using high-fidelity virtual reality to study perception in freely moving observers. *Journal of Vision*, *15*(9), 3. https://doi.org/10.1167/15.9.3

Scarfe, P., & Glennerster, A. (2019). The science behind virtual reality displays. *Annual Review of Vision Science*, *5*(Volume 5, 2019), 529–547. https://doi.org/https://doi.org/10.1146/annurev-vision-091718-014942

Schuetz, I., & Fiehler, K. (2022). Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability. *Journal of Eye Movement Research*, *15*(3). https://doi.org/10.16910/jemr.15.3.3

Schulze, C., James, G., Koehler, D. J., & Newell, B. R. (2019). Probability matching does not decrease under cognitive load: A preregistered failure to replicate. *Memory & Cognition*, *47*(3), 511–518. https://doi.org/10.3758/s13421-018-0888-3

Schulze, C., & Newell, B. R. (2016). Taking the easy way out? increasing implementation effort reduces probability maximizing under cognitive load. *Memory & Cognition*, *44*(5), 806–818. https://doi.org/10.3758/s13421-016-0595-x

Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Engbert, R., & Wichmann, F. A. (2019). Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, *19*(3), 1–1. https://doi.org/10.1167/19.3.1

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, *11*(5), 9–9. https://doi.org/10.1167/11.5.9

Schütz, A. C., Trommershäuser, J., & Gegenfurtner, K. R. (2012). Dynamic integration of information about salience and value for saccadic eye movements. *Proceedings of the National Academy of Sciences*, *109*(19), 7547–7552. https://doi.org/10.1073/pnas.1115638109

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in cognitive sciences*, *10*(1), 38–45. https://doi.org/10.1016/j.tics.2005.11.008

Sharma, M., Martínez Martínez, C. A., Wirth, B. E., Krüger, A., & Müller, P. (2024). Distinguishing target and non-target fixations with eeg and eye tracking in realistic visual scenes. *Proceedings of the 26th International Conference on Multimodal Interaction*, 459–468. https://doi.org/10.1145/3678957.3685728

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *The Quarterly Journal of Experimental Psychology Section A*, *38*(3), 475–491. https://doi.org/10.1080/14640748608401609

Shioiri, S., Kobayashi, M., Matsumiya, K., & Kuriki, I. (2018). Spatial representations of the viewer's surroundings. *Scientific reports*, *8*(1), 7171. https://doi.org/10.1038/s41598-018-25433-5

Siegel, S., & Goldstein, D. A. (1959). Decision-making behavior in a two-choice uncertain outcome situation. *Journal of Experimental Psychology*, *57*(1), 37. https://doi.org/10.1037/h0045959

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in cognitive sciences*, *1*(7), 261–267. https://doi.org/10.1016/S1364-6613(97)01080-2

Sipatchin, A., Wahl, S., & Rifai, K. (2021). Eye-tracking for clinical ophthalmology with virtual reality (vr): A case study of the htc vive pro eye's usability. *Healthcare*, *9*(2). https://doi.org/10.3390/healthcare9020180

Skavenski, A., Hansen, R., Steinman, R., & Winterson, B. (1979). Quality of retinal image stabilization during small natural and artificial body rotations in man. *Vision Research*, *19*(6), 675–683. https://doi.org/10.1016/0042-6989(79)90243-8

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *3*(2), 130–144. https://doi.org/10.1162/pres.1994.3.2.130

Sloot, L., van der Krogt, M., & Harlaar, J. (2014). Effects of adding a virtual reality environment to different modes of treadmill walking. *Gait & Posture*, *39*(3), 939–945. https://doi.org/10.1016/j.gaitpost.2013.12.005

Smit, A., Van Gisbergen, J., & Cools, A. (1987). A parametric analysis of human saccades in different experimental paradigms. *Vision Research*, *27*(10), 1745–1762. https://doi.org/10.1016/0042-6989(87)90104-0

Smith, A., Singh, K., Williams, A., & Greenlee, M. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*, *11*(12), 1182–1190. https://doi.org/10.1093/cercor/11.12.1182

Solomon, S. G. (2021). Chapter 3 - retinal ganglion cells and the magnocellular, parvocellular, and koniocellular subcortical visual pathways from the eye to the brain. In J. J. Barton & A. Leff (Eds.), *Neurology of vision and visual disorders* (pp. 31–50, Vol. 178). Elsevier. https://doi.org/10.1016/B978-0-12-821377-3.00018-0

Spauschus, A., Marsden, J., Halliday, D. M., Rosenberg, J. R., & Brown, P. (1999). The origin of ocular microtremor in man. *Experimental Brain Research*, *126*(4), 556–562. https://doi.org/10.1007/s002210050764

Sprague, W. W., Cooper, E. A., Tošić, I., & Banks, M. S. (2015). Stereopsis is adaptive for the natural environment. *Science advances*, *1*(4), e1400254. https://doi.org/10.1126/sciadv.1400254

SR Research. (2020). Eyelink 1000 plus technical specifications. https://www.sr-research.com/wp-content/uploads/2017/11/eyelink-1000-plus-specifications.pdf

SR Research Ltd. (2009). Eyelink 1000 User Manual 1.5.0. http://sr-research.jp/support/EyeLink%201000%20User%20Manual%201.5.0.pdf

Stein, N., Niehorster, D. C., Watson, T., Steinicke, F., Rifai, K., Wahl, S., & Lappe, M. (2021). A comparison of eye tracking latencies among several commercial head-mounted displays. *i-Perception*, *12*(1). https://doi.org/10.1177/204166952098333

Steinbach, M. J. (1976). Pursuing the perceptual rather than the retinal stimulus. *Vision Research*, *16*(12), 1371–1376. https://doi.org/10.1016/0042-6989(76)90154-1

Steinicke, F., Bruder, G., Jerald, J., Frenz, H., & Lappe, M. (2009). Estimation of detection thresholds for redirected walking techniques. *IEEE transactions on visualization and computer graphics*, *16*(1), 17–27. https://doi.org/10.1109/TVCG.2009.62

Steinicke, F., Visell, Y., Campos, J., & Lécuyer, A. (2013). *Human walking in virtual environments* (Vol. 2). Springer. https://doi.org/10.1007/978-1-4419-8432-6

Steinman, R. M., & Collewijn, H. (1980). Binocular retinal image motion during active head rotation. *Vision Research*, *20*(5), 415–429. https://doi.org/10.1016/0042-6989(80)90032-2

Stevenson, S., Volkmann, F., Kelly, J., & Riggs, L. (1986). Dependence of visual suppression on the amplitudes of saccades and blinks. *Vision Research*, *26*(11), 1815–1824. https://doi.org/10.1016/0042-6989(86)90133-1

Stuphorn, V., Taylor, T. L., & Schall, J. D. (2000). Performance monitoring by the supplementary eye field. *Nature*, *408*(6814), 857–860. https://doi.org/10.1038/35048576

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*(5678), 1782–1787. https://doi.org/10.1126/science.1094765

Suma, E. A., Lipps, Z., Finkelstein, S., Krum, D. M., & Bolas, M. (2012). Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture. *IEEE Transactions on Visualization and Computer Graphics*, *18*(4), 555–564. https://doi.org/10.1109/TVCG.2012.47

Sun, Q., Patney, A., Wei, L.-Y., Shapira, O., Lu, J., Asente, P., Zhu, S., Mcguire, M., Luebke, D., & Kaufman, A. (2018). Towards virtual reality infinite walking: Dynamic saccadic redirection. *ACM Trans. Graph.*, *37*(4). https://doi.org/10.1145/3197517.3201294

Sutherland, I. E. (1965). The ultimate display. *Proceedings of the IFIP Congress*, *2*, 506–508.

Sutherland, I. E. (1968). A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, 757–764. https://doi.org/10.1145/1476589.1476686

’t Hart, B. M., & Einhäuser, W. (2012). Mind the step: Complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Experimental Brain Research*, *223*(2), 233–249. https://doi.org/10.1007/s00221-012-3254-x

Tatler, B. W., Baddeley, R. J., & Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, *46*(12), 1857–1862. https://doi.org/10.1016/j.visres.2005.12.005

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5), 5–5. https://doi.org/10.1167/11.5.5

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *366*(1564), 596–610. https://doi.org/10.1098/rstb.2010.0188

Tatler, B. W., & Tatler, S. L. (2013). The influence of instructions on object memory in a real-world setting. *Journal of Vision*, *13*(2), 5–5. https://doi.org/10.1167/13.2.5

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception*, *1*(1), 7–27. https://doi.org/10.1068/i0382

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica*, *135*(2), 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385. https://doi.org/10.1111/1467-9280.00071

Theeuwes, J., Kramer, A. F., Hahn, S., Irwin, D. E., & Zelinsky, G. J. (1999). Influence of attentional capture on oculomotor control. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1595–1608. https://doi.org/10.1037/0096-1523.25.6.1595

Theeuwes, J., Van der Burg, E., & Belopolsky, A. (2008). Detecting the presence of a singleton involves focal attention. *Psychonomic Bulletin & Review*, *15*(3), 555–560. https://doi.org/10.3758/PBR.15.3.555

Thier, P., & Ilg, U. J. (2005). The neural basis of smooth-pursuit eye movements. *Current Opinion in Neurobiology*, *15*(6), 645–652. https://doi.org/10.1016/j.conb.2005.10.013

Thomas, N. D., Gardiner, J. D., Crompton, R. H., & Lawson, R. (2020a). Physical and perceptual measures of walking surface complexity strongly predict gait and gaze behaviour. *Human Movement Science*, *71*, 102615. https://doi.org/10.1016/j.humov.2020.102615

Thomas, N. D., Gardiner, J. D., Crompton, R. H., & Lawson, R. (2020b). Look out: An exploratory study assessing how gaze (eye angle and head angle) and gait speed are influenced by surface complexity. *PeerJ*, *8*, e8838. https://doi.org/10.7717/peerj.8838

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. https://doi.org/10.1038/381520a0

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of experimental psychology: human perception and performance*, *16*(3), 459. https://doi.org/10.1037/0096-1523.16.3.459

Tripathi, S., & Guenter, B. (2017). A statistical approach to continuous self-calibrating eye gaze tracking for head-mounted virtual reality systems. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 862–870. https://doi.org/10.1109/WACV.2017.101

Tuhkanen, S., Pekkanen, J., Rinkkala, P., Mole, C., Wilkie, R. M., & Lappi, O. (2019). Humans use predictive gaze strategies to target waypoints for steering. *Scientific reports*, *9*(1), 1–18. https://doi.org/10.1038/s41598-019-44723-0

Tuten, W. S., & Harmening, W. M. (2021). Foveal vision. *Current Biology*, *31*(11), R701–R703. https://doi.org/10.1016/j.cub.2021.03.097

Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., & Brooks Jr, F. P. (1999). Walking > Walking-in-place > Flying, in virtual environments. *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 359–364. https://doi.org/doi.org/10.1145/311535.311589

Valsecchi, M., & Gegenfurtner, K. R. (2016). Dynamic re-calibration of perceived size in fovea and periphery through predictable size changes. *Current Biology*, *26*(1), 59–63. https://doi.org/10.1016/j.cub.2015.10.067

van Ede, F., de Lange, F. P., & Maris, E. (2012). Attentional cues affect accuracy and reaction time via different cognitive and neural processes. *Journal of Neuroscience*, *32*(30), 10408–10412. https://doi.org/10.1523/JNEUROSCI.1337-12.2012

van Heusden, E., Donk, M., & Olivers, C. N. L. (2021). The dynamics of saliency-driven and goal-driven visual selection as a function of eccentricity. *Journal of Vision*, *21*(3), 2–2. https://doi.org/10.1167/jov.21.3.2

van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 746–759. https://doi.org/10.1037/0096-1523.30.4.749

Vinnikov, M., Allison, R. S., & Swierad, D. (2008). Real-time simulation of visual defects with gaze-contingent display. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 127–130. https://doi.org/10.1145/1344471.1344504

Virk, R. (2025). Chapter 4 - intelligent virtual characters, AGI, and the future of metaverse. In C. S. Nam, D. Song, & H. Jeong (Eds.), *Human-centered metaverse* (pp. 51–73). Morgan Kaufmann. https://doi.org/10.1016/B978-0-443-21996-2.00008-5

Vogel, J., & de Freitas, N. (2008). Target-directed attention: Sequential decision-making for gaze planning. *2008 IEEE International Conference on Robotics and Automation*, 2372–2379. https://doi.org/10.1109/ROBOT.2008.4543568

Volkmann, F. C. (1986). Human visual suppression. *Vision Research*, *26*(9), 1401–1416. https://doi.org/10.1016/0042-6989(86)90164-1

Volkmann, F. C., Riggs, L. A., Ellicott, A. G., & Moore, R. K. (1982). Measurements of visual suppression during opening, closing and blinking of the eyes. *Vision Research*, *22*(8), 991–996. https://doi.org/10.1016/0042-6989(82)90035-9

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*(1), 101–118. https://doi.org/10.1111/1467-6419.00106

Wade, N. J. (2010). Pioneers of eye movement research. *i-Perception*, *1*(2), 33–68. https://doi.org/10.1068/i0389

Wallach, H., & Lewis, C. (1966). The effect of abnormal displacement of the retinal image during eye movements. *Perception & Psychophysics*, *1*(1), 25–29. https://doi.org/10.3758/BF03207816

Wandell, B. A., & Winawer, J. (2011). Imaging retinotopic maps in the human brain. *Vision Research*, *51*(7), 718–737. https://doi.org/10.1016/j.visres.2010.08.004

Warburton, M., Mon-Williams, M., Mushtaq, F., & Morehead, J. R. (2023). Measuring motion-to-photon latency for sensorimotor experiments with virtual reality systems. *Behavior Research Methods*, *55*(7), 3658–3678. https://doi.org/10.3758/s13428-022-01983-5

Warren, W. H., Kay, B. A., Zosh, W. D., Duchon, A. P., & Sahuc, S. (2001). Optic flow is used to control human walking. *Nature neuroscience*, *4*(2), 213–216. https://doi.org/10.1038/84054

Wei, S., Bloemers, D., & Rovira, A. (2023). A preliminary study of the eye tracker in the Meta Quest Pro. *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*, 216–221. https://doi.org/10.1145/3573381.3596467

Wexler, M., & van Boxtel, J. J. A. (2005). Depth perception by the active observer. *Trends in Cognitive Sciences*, *9*(9), 431–438. https://doi.org/10.1016/j.tics.2005.06.018

Wheatstone, C. (1838). On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, *128*, 371–394. https://doi.org/10.1098/rstl.1838.0019

Wiener, J., De Condappa, O., & Holscher, C. (2011). Do you have to look where you go? gaze behaviour during spatial decision making. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*. https://escholarship.org/uc/item/9n91h72n

Wierts, R., Janssen, M. J., & Kingma, H. (2008). Measuring saccade peak velocity using a low-frequency sampling rate of 50 hz. *IEEE Transactions on Biomedical Engineering*, *55*(12), 2840–2842. https://doi.org/10.1109/TBME.2008.925290

Winawer, J., & Witthoft, N. (2015). Human V4 and ventral occipital retinotopic maps (2015/08/04). *Visual Neuroscience*, *32*. https://doi.org/10.1017/S0952523815000176

Wolf, C., & Lappe, M. (2020). Top-down control of saccades requires inhibition of suddenly appearing stimuli. *Attention, Perception, & Psychophysics*, *82*(8), 3863–3877. https://doi.org/10.3758/s13414-020-02101-3

Wolf, C., & Lappe, M. (2021). Salient objects dominate the central fixation bias when orienting toward images. *Journal of Vision*, *21*(8), 23–23. https://doi.org/10.1167/jov.21.8.23

Wolfe, J. M. (2020). Visual search: How do we find what we are looking for? *Annual Review of Vision Science*, *6*(1), 539–562. https://doi.org/10.1146/annurev-vision-091718-015048

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 0058. https://doi.org/10.1038/s41562-017-0058

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Wu, R.-J., Clark, A. M., Cox, M. A., Intoy, J., Jolly, P. C., Zhao, Z., & Rucci, M. (2023). High-resolution eye-tracking via digital imaging of Purkinje reflections. *Journal of Vision*, *23*(5), 4–4. https://doi.org/10.1167/jov.23.5.4

Wu, T. C., & Tsotsos, J. K. (2025). Real-world visual search goes beyond eye movements: Active searchers select 3D scene viewpoints too. *PLOS ONE*, *20*(7), 1–30. https://doi.org/10.1371/journal.pone.0319719

Wurtz, R. H., & Goldberg, M. E. (1972). Activity of superior colliculus in behaving monkey. iv. effects of lesions on eye movements. *Journal of Neurophysiology*, *35*(4), 587–596. https://doi.org/10.1152/jn.1972.35.4.587

Wurtz, R. H., & Mohler, C. W. (1976). Organization of monkey superior colliculus: Enhanced visual response of superficial layer cells. *Journal of Neurophysiology*, *39*(4), 745–765. https://doi.org/10.1152/jn.1976.39.4.745

Xu, Y., Dong, Y., Wu, J., Sun, Z., Shi, Z., Yu, J., & Gao, S. (2018). Gaze prediction in dynamic 360 immersive videos. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5333–5342. https://doi.org/10.1109/CVPR.2018.00559

Yarbus, A. L. (1967). Saccadic eye movements. In *Eye movements and vision* (pp. 129–146). Springer US. https://doi.org/10.1007/978-1-4899-5379-7_5

Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature*, *414*, 302–305. https://doi.org/10.1038/35104551

Young, L. R., & Sheena, D. (1975a). Eye-movement measurement techniques. *American Psychologist*, *30*(3), 315–330. https://doi.org/10.1037/0003-066X.30.3.315

Young, L. R., & Sheena, D. (1975b). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, *7*(5), 397–429. https://doi.org/10.3758/BF03201553

Yu, C., Ma, X., Ren, J., Zhao, H., & Yi, S. (2020). Spatio-temporal graph transformer networks for pedestrian trajectory prediction. *European Conference on Computer Vision*, 507–523. https://doi.org/10.1007/978-3-030-58610-2_30

Zank, M., & Kunz, A. (2016a). Eye tracking for locomotion prediction in redirected walking. *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, 49–58. https://doi.org/10.1109/3DUI.2016.7460030

Zank, M., & Kunz, A. (2016b). Where are you going? Using human locomotion models for target estimation. *The Visual Computer*, *32*(10), 1323–1335. https://doi.org/10.1007/s00371-016-1229-9

Zank, M., & Kunz, A. (2017). Optimized graph extraction and locomotion prediction for redirected walking. *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, 120–129. https://doi.org/10.1109/3DUI.2017.7893328

Zeki, S., Watson, J. D., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. (1991). A direct demonstration of functional specialization in human visual cortex. *J Neurosci*, *11*(3), 641–649.

Zietz, D., & Hollands, M. A. (2009). Gaze behavior of young and older adults during stair walking. *Journal of Motor Behavior*, *41*(4), 357–366. https://doi.org/10.3200/JMBR.41.4.357-366

Zimmermann, E., & Lappe, M. (2016). Visual space constructed by saccade motor maps. *Frontiers in human neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00225

Zmuda, M. A., Wonser, J. L., Bachmann, E. R., & Hodgson, E. (2013). Optimizing constrained-environment redirected walking instructions using search techniques. *IEEE transactions on visualization and computer graphics*, *19*(11), 1872–1884. https://doi.org/10.1109/TVCG.2013.88

Zuber, B., & Stark, L. (1966). Saccadic suppression: Elevation of visual threshold associated with saccadic eye movements. *Experimental Neurology*, *16*(1), 65–79. https://doi.org/10.1016/0014-4886(66)90087-2

Zuber, B., Stark, L., & Lorber, M. (1966). Saccadic suppression of the pupillary light reflex. *Experimental Neurology*, *14*(3), 351–370. https://doi.org/10.1016/0014-4886(66)90120-8

Acknowledgements

I dedicate this thesis, first and foremost, to my wife, Vera. Thank you for your unwavering love, your constant words of encouragement, and the infectious joy of life you share with me every day. I am the luckiest person in the world to have you by my side.

My deepest gratitude goes to my supervisor, Markus Lappe. I hold him in the highest regard as a mentor and scientific role model. Since my first tentative steps as a scientist, Markus has patiently and trustingly opened so many doors and illuminated countless paths, thereby immeasurably expanding my scientific world. I particularly value our many insightful discussions, which consistently challenged my thinking and greatly enriched the research presented in this thesis.

I am profoundly grateful to the many other mentors who shaped my professional development through insightful discussions and unwavering support: Steffen Nestler and Mitja Back were the first to place their trust in me, encouraging me to take the leap and simply begin working on scientific projects. Katharina Rifai taught me how to ambitiously carry out a project from start to finish and how to continually develop myself in the process. Siegfried Wahl showed me how to be innovative at the organizational level and how to structure teams to foster the development of new ideas. Diederick Niehorster motivated me to understand even the trickiest topics in exhaustive detail. Gerd Bruder taught me how to maintain motivation and momentum, even when time became scarce in the final stages of the project. Frank Steinicke demonstrated the art of rapid innovation. Tamara Watson gave me significant self-confidence and trust in my own abilities, showing me how scientific skills can be applied across a variety of projects, and how strong collaboration leads to superior results. Szonya Durant taught me the profound importance of a supportive social environment for project progress and motivated me to pay close attention to the smallest details, ultimately helping me complete this thesis. Heiko Wagner provided interesting insights into biomechanics that inspired my thinking about perception during locomotion. Finally, Christian Becker-Carus, who showed me how to remain curious and engaged, even in old age. All of these individuals took me seriously from the very beginning and offered generous, friendly support. They are all fantastic and very different role models who collectively showed me new ways to approach science and made me a better researcher.

My daily work has been immeasurably enriched by the colleagues I have worked alongside in Münster. I want to thank Gianni Bremer, Sabine Tepper, Annegret Meermeier, Phil Wieland, Maren Westendorf, Taravat Anvari, Svenja Gremmler, Malte Scherff, Krischan Koerfer, Frauke Heins, Christian Wolf, Jana Masselink, Anna Hülemeier, Johannes Kirchner, Joscha Stecker, Sigurd Hadamus, Jaclyn Oltmann, Fabian Ptok, and Sergej Buchhorn. Thank you for your collaboration, the many fruitful discussions, and for making my time there so much more enjoyable. Finally, I would like to specifically remember Harry for his friendship. I still miss him every time I walk past his office.

I am deeply and eternally grateful to my family. To my mother, thank you for the unconditional support that accompanied every endeavour I ever pursued. You instilled in me curiosity, creativity, humour, and taught me to treat others with care. To my sisters, Rabea and Flora, thank you for your unconditional sibling love. I couldn't imagine better sisters, and I am immensely proud of the different paths we are taking in our lives. To my father, thank you for teaching me to think outside the box and for demonstrating that most technical problems can be solved with perseverance and pragmatic, trial-and-error problem-solving. I am also eternally grateful to my grandparents, who were always so proud of me, encouraged my ambition, and supported my adventurous spirit. Finally, thank you to the rest of my very large family, who have shown me since birth that friendship and love speak many languages and do not end at national borders. I am also grateful to Vera's family for welcoming me with such open arms.

My sincere gratitude extends also to my friends, who have been invaluable pillars of support. First and foremost, Hendrik, Jann, and Paul, who have been with me for as long as I can remember. Having friends with whom you can share your entire life is one of the greatest gifts in the world, and I would not be the person I am today without you. Thank you to Hannah, Lena, and Luke, who taught me how to navigate university and who have always been there to offer the right advice at the right time. Without them, I certainly would not have had the courage to even begin a PhD. Thank you to Pauline, Johannes, Lydia, Diogo, Laura, Freddy, Judith, Sarah, Andi, and Milena, who made Münster my true home for so many years and who ensure today that I can feel at home in many places at once. The same goes for Caterina, Marinke, Neil, François, Aash, Isabelle, Will and Anne, Jakob, Matthi, Theresa, Sandra, Fabian, Matt, Eleanor, and Marianne, who all welcomed me with open arms when I arrived in cities where I did not know anyone. Having so many loved ones and friends who feel like one big family spread across the globe is truly invaluable.

Then there are my track and field coaches and mentors: Sina, Dennis, Robert, and Micha. Together, you taught me the discipline required to maintain ambition over long periods and how to organize myself in a goal-oriented manner. You also enabled me to gain significant experience and self-confidence in leading groups. Thank you!

Finally, thank you to my teachers who provided the foundation for this entire journey. Special thanks go to Felix Platzköster, who sparked my interest in mathematics and statistics; Annette Bagusche, who first inspired me with empiricism, theories, and biological principles; Eckard Köhler, who taught me the most important basics of programming and sparked my enthusiasm for open source and finally Stefanie Tiemann, the best primary school teacher I could have wished for.

Appendix

Declaration of Academic Integrity

I hereby confirm that this thesis, entitled, "How We See, Perceive and Act: Analysing Eye Movements in Virtual Reality" is solely my own work and that I have used no sources or aids other than the ones stated. All passages in my thesis for which other sources, including electronic media, have been used, be it direct quotes or content references, have been acknowledged as such and the sources cited. I am aware that plagiarism is considered an act of deception which can result in sanction in accordance with the examination regulations. I confirm that I am aware that my work may be cross-checked with other texts to identify possible similarities and that it may be stored in a database for this purpose. I confirm that I have not submitted the following thesis in part or whole as an examination paper before.

Date Signature

Erklärung des Promovenden / der Promovendin zur Dokumentation von **Open Science-Aktivitäten** und der Erklärung zur **Beschäftigung mit ethischen Aspekten** im Rahmen der Promotion und zum **eigenen Anteil** an den vorgelegten wissenschaftlichen Abhandlungen mit zwei oder mehr Autor(inn)en (kumulative Dissertation)

Promovend/Promovendin: Niklas Hypki

Titel der Dissertation: How We See, Perceive and Act: Analysing Eye Movements in Virtual Reality

1. Dokumentation von **Open Science-Aktivitäten**

Wissenschaftliche Abhandlung 1

	ja	nein	Wenn ja, unter welcher Quelle verfügbar
Präregistrierung	○	●	
Veröffentlichung von Daten	○	●	
Veröffentlichung von Auswertungsskripten	○	●	
Veröffentlichung von Materialien	○	●	
Open Access Publikation	●	○	https://doi.org/10.1177/2041669520983338
Preprint	○	●	

Wissenschaftliche Abhandlung 2

	ja	nein	Wenn ja, unter welcher Quelle verfügbar
Präregistrierung	○	●	
Veröffentlichung von Daten	●	○	https://osf.io/b43uv/
Veröffentlichung von Auswertungsskripten	○	●	
Veröffentlichung von Materialien	○	●	
Open Access Publikation	○	●	
Preprint	●	○	https://nbn-resolving.de/urn:nbn:de:hbz:6-74019498937

Wissenschaftliche Abhandlung 3

	ja	nein	Wenn ja, unter welcher Quelle verfügbar
Präregistrierung	○	●	
Veröffentlichung von Daten	●	○	https://osf.io/ewsj5/
Veröffentlichung von Auswertungsskripten	○	●	
Veröffentlichung von Materialien	○	●	
Open Access Publikation	●	○	https://doi.org/10.1038/s41598-024-59657-5
Preprint	○	●	

2. Erklärung zur **Beschäftigung mit ethischen Aspekten**

Studien-nummer	**Quellenangabe (Manuskript / Kapitel der Promotion):** z.B. Studie 1 in Paper 2, Studie 1 beschrieben in Kapitel 4	**Begutachtung der Studie durch eine Ethikkommission**	
		Ja	**nein**
1	Studie 1 in wissenschaftlicher Abhandlung 1 (Kapitel I)	●	○
2	Studie 2 in wissenschaftlicher Abhandlung 1 (Kapitel I)	●	○
3	Experiment in wissenschaftlicher Abhandlung 2 (Kapitel II)	●	○
4	Experiment 1 in wissenschaftlicher Abhandlung 3 (Kapitel III)	●	○
5	Experiment 2 in wissenschaftlicher Abhandlung 3 (Kapitel III)	●	○

3. Erklärung zum eigenen Anteil an den vorgelegten wissenschaftlichen Abhandlungen mit zwei oder mehr Autor(inn)en

Wissenschaftliche Abhandlung 1

Titel	A Comparison of Eye Tracking Latencies among several commercial Head-mounted Displays
Autor(en)	Niklas Stein, Diederick C. Niehorster, Tamara Watson, Frank Steinicke, Katharina Rifai , Siegfried Wahl, Markus Lappe

Publikationsstatus:	nicht eingereicht	○
	eingereicht	○
	in Begutachtung	○
	in Revision	○
	angenommen	○
	publiziert	●

Journal	i-Perception
Publikationsjahr	2021

Beschreibung des eigenen Anteils, wenn keine Alleinautorenschaft vorliegt:

N.S.
Konzeption der Studie, Datenerhebung, Analyse und Interpretation der Daten, Formulierung & Überarbeitung des Manuskripts

D.N.
Konzeption der Studie, Analyse und Interpretation der Daten, Überarbeitung des Manuskripts

T.W.
Konzeption der Studie, Überarbeitung des Manuskripts

F.S.
Konzeption der Studie, Überarbeitung des Manuskripts

K.R.
Konzeption der Studie, Überarbeitung des Manuskripts

S.W.
Überarbeitung des Manuskripts

M.L.
Konzeption der Studie, Formulierung & Überarbeitung des Manuskripts

Wissenschaftliche Abhandlung 2

Titel	Eye Tracking-based LSTM for Locomotion Prediction in VR
Autor(en)	Niklas Stein & Gianni Bremer, Markus Lappe

Publikationsstatus:	nicht eingereicht	○
	eingereicht	○
	in Begutachtung	○
	in Revision	○
	angenommen	○
	publiziert	●

Journal	2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)
Publikationsjahr	2022

Beschreibung des eigenen Anteils, wenn keine Alleinautorenschaft vorliegt:

N.S.
Konzeption der Studie, Datenerhebung, Analyse und Interpretation der Daten, Formulierung & Überarbeitung des Manuskripts

G.B.
Konzeption der Studie, Datenerhebung, Analyse und Interpretation der Daten, Formulierung & Überarbeitung des Manuskripts

M.L.
Konzeption der Studie, Analyse und Interpretation der Daten, Formulierung & Überarbeitung des Manuskripts

Wissenschaftliche Abhandlung 3

Titel	Eye and head movements in visual search in the extended field of view	
Autor(en)	Niklas Stein, Tamara Watson, Markus Lappe, Maren Westendorf, Szonya Durant	
Publikationsstatus:	nicht eingereicht	○
	eingereicht	○
	in Begutachtung	○
	in Revision	○
	angenommen	○
	publiziert	●
Journal	Scientific Reports	
Publikationsjahr	2024	

Beschreibung des eigenen Anteils, wenn keine Alleinautorenschaft vorliegt:

N.S.
Konzeption der Studie, Analyse und Interpretation der Daten, Formulierung & Überarbeitung des Manuskripts

T.W.
Konzeption der Studie, Formulierung & Überarbeitung des Manuskripts

M.L.
Konzeption der Studie, Formulierung & Überarbeitung des Manuskripts

M.W.
Konzeption der Studie, Datenerhebung, Analyse und Interpretation der Daten

S.D.
Konzeption der Studie, Datenerhebung, Formulierung & Überarbeitung des Manuskripts

Ort, Datum
Lüneburg, 29.11.2025

Unterschrift
Promovend/in N. Hypki

Curriculum Vitae

Niklas Hypki

Personal

Occupation	Research Associate at University of Münster
Nationality	German
Residence	Lüneburg, Germany
Contact	Niklas.Hypki@uni-muenster.de, niklas-hypki.de

Education

2019 – 2026	PhD, University of Münster
2016 – 2019	Master of Science Psychology, University of Münster
2013 – 2016	Bachelor of Science Psychology, University of Münster
2004 – 2013	Allgemeine Hochschulreife, Theodor-Heuss-Gymnasium Waltrop

Work experience

2019 – 2026	Research Associate at University of Münster
2024 – 2025	Research Associate at Royal Holloway University of London
2022 – 2023	Research Associate at Western Sydney University
2017 – 2018	Internship at Zeiss Vision Science Lab in Tübingen

Awards

2021	Best Paper Award at the IEEE International Conference on Artificial Intelligence and Virtual Reality
2020	Early Career Advancement Prize by the Chief Editors and Board of i-Perception

Major publications

Stein, N., Watson, T., Lappe, M. et al. (2024). Eye and head movements in visual search in the extended field of view. Scientific Reports 14, 8907. doi: https://doi.org/10.1038/s41598-024-59657-5

Stein, N., Bremer, G. and Lappe, M. (2022). Eye Tracking-based LSTM for Locomotion Prediction in VR, 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Christchurch, New Zealand, 493-503, doi: https://doi.org/10.1109/VR51125.2022.00069

Stein, N., Niehorster, D. C., Watson, T. et al. (2021). A Comparison of Eye Tracking Latencies Among Several Commercial Head-Mounted Displays. i-Perception, 12(1). doi: https://doi.org/10.1177/2041669520983338

Full publication list: https://scholar.google.com/citations?user=_7KeMeIAAAAJ

Project participation

2023 – 2026	EYERES - Vorhersagen, Hervorrufen und Ausnutzen von Augenbewegungen und Blinzeln für Redirektionssysteme
2023 – 2026	PLACES - PLAsticity of perception in real and virtual spaCES
2022 – 2023	MAIA - Multifunctional, adaptive and interactive AI system for Acting in multiple contexts
2018 – 2023	PLATYPUS - PLAsticiTY of Perceptual space Under Sensorimotor interactions
2018 – 2023	iLUI - interaktive Lokomotionbenutzerschnittstellen zum realen Gehen durch virtuelle Welten - Von der Wahrnehmung zur Anwendung

Teaching

2025 – 2026	Update kognitive Neurowissenschaft, Seminar (M.Sc. Kognitive Neurowissenschaften)
2020 – 2024	Aufmerksamkeit/ Attention, Seminar (B.Sc. Psychology)
2023	Empirisches Praktikum, Seminar (B.Sc. Psychology)