Gaze on the Go: Effect of Spatial Reference Frame on Visual Target Acquisition During Physical Locomotion in Extended Reality

Pavel Manakhov
pmanakhov@cs.au.dk
Aarhus University
Aarhus, Denmark

Ludwig Sidenmark
lsidenmark@dgp.toronto.edu
University of Toronto
Toronto, Ontario, Canada

Ken Pfeuffer
ken@cs.au.dk
Aarhus University
Aarhus, Denmark

Hans Gellersen
h.gellersen@lancaster.ac.uk
Lancaster University
Lancaster, United Kingdom
Aarhus University
Aarhus, Denmark

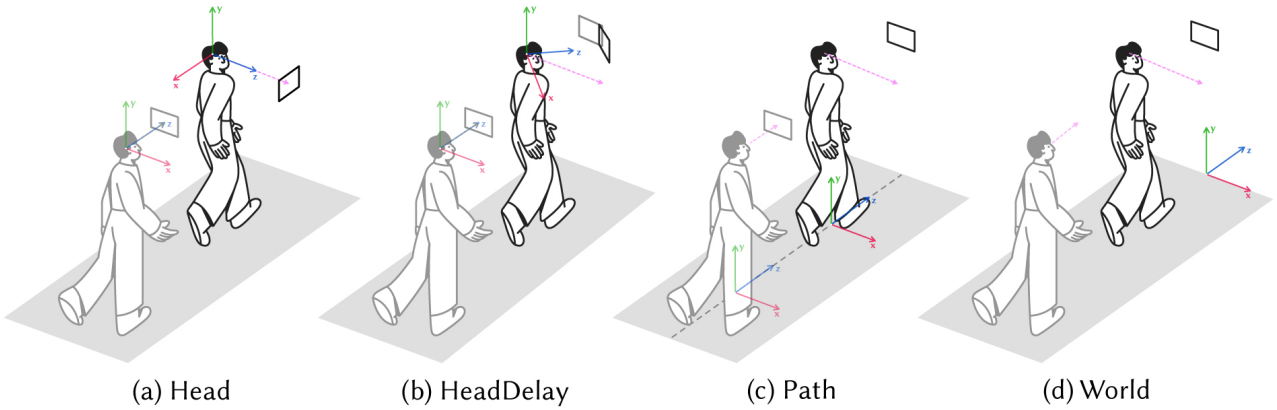


Figure 1: The reference frames we compare in this study: (a) *Head* as the frame of reference where the spatial UI follows head translation and rotation instantaneously. (b) *HeadDelay* is essentially similar to *Head* except it introduces a delay between the head and target's movements to simulate the inertia effect. (c) *Path* is a novel reference frame where the target floats in front of a moving user at a fixed distance and height completely unaffected by head rotation and head translation perpendicular to the direction of locomotion. (d) *World* where the target is placed at a fixed height at the opposite side of the virtual track and is stationary relative to it.

ABSTRACT

Spatial interaction relies on fast and accurate visual acquisition. In this work, we analyse how visual acquisition and tracking of targets presented in a head-mounted display is affected by the user moving linearly at walking and jogging paces. We study four reference frames in which targets can be presented: *Head* and *World* where targets are affixed relative to the head and environment, respectively; *HeadDelay* where targets are presented in the head coordinate system but follow head movement with a delay, and novel *Path* where targets remain at fixed distance in front of the user, in the direction of their movement. Results of our study in virtual reality demonstrate that the more stable the target is relative to the environment, the faster and more precise it can be fixated. The results have practical significance as head-mounted displays enable interaction during mobility, and in particular when eye tracking is considered as input.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Mixed / augmented reality**; **Virtual reality**.

KEYWORDS

spatial UIs, reference frames, UI placement, physical locomotion, extended reality, gaze interaction, eye tracking

ACM Reference Format:

Pavel Manakhov, Ludwig Sidenmark, Ken Pfeuffer, and Hans Gellersen. 2024. Gaze on the Go: Effect of Spatial Reference Frame on Visual Target Acquisition During Physical Locomotion in Extended Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3613904.3642915

1 INTRODUCTION

Head-mounted displays (HMDs) facilitate interaction with extended reality (XR) without users needing to stop in their tracks. This enables many use cases, such as a technician wearing augmented reality (AR) headset and approaching a machine on which a virtual marker indicates a fault, or a player of a location-based virtual reality (VR) shooter game looking at their in-game health on a heads-up display, as well as walking meetings [6], attending notifications [33] and maps [11], interacting with a shopping list in a supermarket [1], and reading [54]. Such interactions rely on gaze, to fixate on targets for a sufficient time to process visual feedback. As spatial input relies on visual guidance, gaze itself can be effective for pointing to assist with target selection [48, 49], and eye-tracking is increasingly integrated as input with HMDs [50, 51, 69]. However, literature suggests that reading and pointing performance on mobile devices decreases when users are in motion [4, 36, 56]. In this work, we study how movement impacts gaze, and analyse the acquisition of targets and the accuracy with which fixations are maintained during linear locomotion. This is fundamental to any visual interaction in HMDs, even when the eyes are not tracked as input.

Objects in spatial user interfaces (UIs) can be presented in different reference frames. A well-known dichotomy is to present objects as fixed in the world versus fixed in the HMD but reference frames can be more nuanced. In this work we consider *World* and *Head*, and additionally *HeadDelay* and *Path* (Figure 1). We use HeadDelay as short for head-referenced target presentation with a delay that simulates inertia. HeadDelay has not been studied in the literature but is recommended in the HMD interface guidelines [30]. We further introduce Path as a novel reference frame where targets are placed at fixed distance and height in front of the user, in the direction of their locomotion, and unaffected by head rotation or lateral translation.

We compared the four reference frames in two related experiments conducted in VR. Participants were asked to track a target located within the headset's field of view (FoV) with their eyes while moving back and forth on an obstacle-free linear track in a virtual environment aligned with its physical counterpart. Our motivation for using VR in the study was to provide the participants with a minimalistic environment that would allow for safe movement but at the same time bring the number of visual features down to a minimum, so not to interfere with the target acquisition task. We studied movement at walking and jogging paces with standing still as a baseline condition. In the first experiment, targets were presented at 1m in front of the users, in the Head, HeadDelay and Path conditions, while in the World condition the targets were places at the opposite side of the 5m track. In the second experiment, targets in these three reference frames were presented at 3.3m distance as the average at which targets were fixated in the World condition.

The study revealed that both target acquisition time and precision of tracking fixation are predicated on the relative movement of the target to the person's head. On the move, acquisition time rises and precision worsens when the target is rigidly affixed to head (Head), or when its relative movement is less predictable (HeadDelay). Conversely, acquisition is faster and precision improves depending on how stable the target is relative to the environment (World), specifically in a plane perpendicular to the direction of movement (Path). Trueness[1] was also affected by the movement pace, while the effect of reference frame on trueness is more nuanced. Distance to the target only had a minute effect on precision. Our results have practical relevance for any visual interaction in HMDs and especially when gaze is used as input modality.

This work contributes to understanding of gaze performance during locomotion and informs interaction design for spatial applications on the go:

- Movement pace has a detrimental effect on both speed with which a person can acquire an object and stability of tracking fixation.
- Choice of reference frame makes a difference. Relative stability of targets in the environment aids acquisition and tracking, while fixation in the HMD, even with inertia, worsens performance.
- The proposed Path reference frame may provide an alternative that combines the mobility of Head reference frames with the stability of World.

2 RELATED WORK

According to the recent studies [44, 64], 26% of people use their mobile devices on the go for eyes-busy mobile interaction, such as social networking, watching videos, and texting. Studies with mobile touchscreen devices have shown that reading and input performance is compromised by locomotion [2, 56]. Walking interaction with HMDs has been investigated with VR and AR studies on the treadmill [34, 36]. Li et al. found that users performed worse in raycasting while walking compared to standing, but relatively better with world-fixed than head-fixed targets, suggesting the difference may be due to proprioception. Closest to our work, Borg et al. studied reading in head versus world reference frames during linear locomotion, finding performance lower for head [4]. Their work is insightful as it explains the effect of stabilisation mechanisms in human vision that compensate for head movement but become undermined when the target the person is fixating on moves with their head.

2.1 Gaze During Locomotion

Gaze has been long considered as modality for input and interaction [23] but only recently started to be considered in contexts where the user is not stationary. Kapp et al. studied the accuracy of HoloLens 2's eye tracker in the context of tracking head-affixed targets located at distances of 0.5-4m from the user during linear

[1]In literature also referred "accuracy". Here, we follow terminology of accuracy as aggregate quality measure composed of "trueness" and "precision".

locomotion. They found that the bigger the distance, the better the tracking fixation trueness, while there was no significant effect on precision [24]. The study also demonstrated that both trueness and precision is, predictably, higher while resting than while walking. With the introduction of headsets such as HoloLens 2, Meta Quest Pro, and Apple Vision Pro, the XR industry has started adopting eye gaze-based interaction techniques proposed by HCI researchers [53], such as Gaze&Dwell [23, 59] and Gaze+Pinch [50]. Gaze also has utility as an implicit input, for example to ensure that visual augmentations do not obstruct the view of the moving person [45, 66]. Even if gaze is not used in this capacity, efficient perception of information placed within different frames of reference is a crucial part of interaction with HMDs on the go.

In order to reason about effect of spatial reference frames, it is important to understand how our visual system functions on the go. Our vision is foveal, and we move our eyes to align objects with the small retinal region on which photoreceptors are most densely clustered. The alignment needs to be maintained for a sufficient time to gain information, which is achieved through oculomotor fixation during which the eyes perform only small fixational movements. However, during locomotion fixations are supported by other types of compensatory mechanisms. When a human walks with a moderate to fast speed (1.4–1.8 m/s), their head translates up and down with an average frequency of 2 Hz, left and right with a frequency of 1 Hz [22, 41]. The head also rotates along with the torso [21]. These head movements need to be counteracted by stabilising eye movements to remain "on target" during a fixation, based primarily on the vestibulo-ocular reflex (VOR). As there are also other mechanisms involved such as smooth pursuit eye movement, we adopt the notion of *enhanced VOR (eVOR)* from Han et al. [20]. Fixations are further overlaid by vergence eye movement, when the distance to the target changes, characterised by movement of the eyes in opposite directions to maintain binocular vision. During locomotion, a fixation is thus supported by a set of eye movements, no longer making it a single oculomotor event. Because of that, in this work, we adopt the notion of *tracking fixation* [28].

2.2 Spatial Reference Frames in Extended Reality

The current commonly-accepted classification of UI placements distinguishes among world, object, head, body, and device reference frames [29, section 9.5.2]. The very first positionally-tracked AR HMD built by Ivan Sutherland and Robert Sproull in 1968 was capable of displaying virtual objects in the *world* frame of reference which would appear as floating in the mid-air [63]. As opposed to that, Steve Mann's VideoOrbits system could identify physical objects such as billboards and place world-anchored information over them [39], the concept extended to semantic alignment in the more recent research [10]. The movable real-world *objects* can also be used as reference frames. The first mention of this idea in the literature belongs to Steve Feiner et al., who implemented the X11 window system for see-through HMDs capable of attaching 2D windows to positionally tracked objects [13], which the authors referred to as 'world-fixed windows' (p. 148). The similar concept was adopted to provide passive tactile feedback while interacting with virtual UI controls anchored to a physical tablet [57]. Modern AR HMDs such as Magic Leap One and HoloLens 2, while technically capable of tracking objects and images, place spatial apps within World by default and allow their users to customize their placement manually. *World* and *object* are also referred as *exocentric* reference frames highlighting their world-centric nature [12].

Over the years, researchers have introduced multitude of reference frames that we call mobile, i.e. UI placements more suitable for interaction on the go. The simplest — *head* reference frame that is sometimes called a heads-up display (HUD) due to its similarity to this type of display [58, p. 89] — is known from 1993 [13]. However, modern XR design guidelines recommend against using head-locked content citing discomfort caused by eye strain [30, 40]. Instead, Magic Leap's guidelines recommend using soft lock simulating momentum. We adopt the technique for our comparison but describe it as "Head with a delay" to capture how it differs from Head reference. Other mobile reference frames — *body*-anchored content according to LaViola Jr et al. [29] — include positionally-tracked torso [35], hand[2] [51], and forearm [18]. It is possible to simulate body reference frames if the HMD is positionally tracked. Billinghurst et al. used this approach to implement their AR conferencing system [3]. Recent work placed content on a fixed height relative to the ground and allowing it to rotate horizontally following head yaw with a delay [26], or simply when an angular threshold is reached [33]. The latter behaviour is implemented as a ready-available component of Microsoft's Mixed Reality Toolkit (MRTK)[3]. *Head* and *body* are also referred as *egocentric* reference frames [12]. Lastly, the *device* reference frame, as defined by LaViola Jr et al. [29], is relevant to handheld devices but not to HMDs, therefore, is not considered here.

Studies comparing spatial reference frames span various contexts from linear locomotion studied with [4, 16] and without a treadmill [17, 24], to curved paths with [26] and without obstacles [35, 67], to uneven surfaces [25]. Lu et al. compared the influence of the head and torso reference frames on the performance of two information access tasks [35]. The study demonstrated that participants perceive the head-referenced widgets as more usable but somewhat cluttered. Lee & Woo got similar results, where notifications presented within the head frame of reference were found to be more noticeable [33]. Another study found text reading harder on a HMD (i.e. head-referenced) than on a handheld device [67]. Recent work in VR on a treadmill found that text readability is higher with text in world reference than in head reference [16]. Interestingly, other work comparing simulated body and head reference frames did not find a significant difference in readability in walking [26]. Other studies comparing reference frames during locomotion were focused on visual search [17], gait performance [25], and situational awareness [47]. We found only a single study that explains differences between head and world reference frames in terms of compensatory eye movements [4], which we noted above as insightful for our purposes.

In advance over prior work, we focus fundamentally on tracking fixations, as this underpin pointing, reading or any visual or visually guided task. While there has been ample comparison of Head versus

[2] https://learn.microsoft.com/en-us/windows/mixed-reality/design/hand-menu
[3] https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/features/ux-building-blocks/solvers/solver?view=mrtkunity-2022-05#radialview

World reference for interaction in HMDs, we include HeadDelay which has not previously been studied, and Path as a novel reference frame for interaction on the go.

3 SPATIAL REFERENCE FRAME DESIGN

In this study we compared four spatial reference frames: head, head with a delay, world, and novel path. The choice of the spatial reference frames was informed by the gaze stabilization mechanisms that become active while tracking the target with one's eyes on the go. In order to better understand the involvement of those mechanisms, we introduce the frames of reference along with their target movement relative to the head. The description is based on a left-handed coordinate system (as used in, e.g., the Unity game engine[4]). The coordinate system is oriented to the right with a positive X coordinate, up with a positive Y, and forward with a positive Z. We discuss the four reference frames in the order that makes the description of corresponding relative target movements the most succinct.

Path. This reference frame denotes a novel approach where the frame of reference translates according to the user's walking path by taking historical information into account. The Z and X axes of this reference frame lay within the ground plane and the Y axis is oriented upwards (Figure 1c). The origin moves along a predicted path and always stays in the middle of it effectively ignoring lateral oscillation of the person's head. The translation along the path is synchronised with the head position one-to-one. Within this reference frame the target is affixed at a constant distance from the user (the Z coordinate) and height above the ground (the Y coordinate). Since the origin translates along the track with the user's head, the target always stays at the same distance from them.

In case of a motion along a linear track, the predicted path coincides with the track and the Z axis of the reference frame is oriented towards the locomotion direction. In case of free-form locomotion, the direction of the person's movement can be identified by extrapolating the headsets' trajectory. This way, whenever the person turns, it would seem to them that the UI movement anticipates the user's path. When not moving, the UI would not rotate in response to the person turning around. Only after a certain velocity threshold has been reached would the frame of reference orient itself according to the predicted direction of movement.

From the perspective of a moving user, the target exhibits a complex repetitive movement comprised of sinusoidal horizontal and vertical oscillations caused by head translation (Figure 2a) — the type of movements the eVOR is effective at compensating. The amplitude of these oscillations is affected by two factors — the movement pace and the distance to the target — both factors that are covered in the current study.

World. The origin of this reference frame stays still relative to the surrounding environment and, similarly to *Path*, is located on the ground plane (Figure 1d). The target within this coordinate system is located at a fixed height. In our study, the targets were placed at the opposite side of the virtual track. From a moving observer's perspective, the target placed to the side of the track, i.e. not directly in front of the observer, exhibits the same movement as in *Path* plus an additional component originating from the constantly decreasing distance between the person's head and the target. This component from is perceived as if the target is trying to leave the person's FoV (Figure 2b). Thus, in addition to eVOR, in this reference frame, the person's eyes converge the closer the person is to the target.

[4]https://docs.unity3d.com/Manual/QuaternionAndEulerRotationsInUnity.html

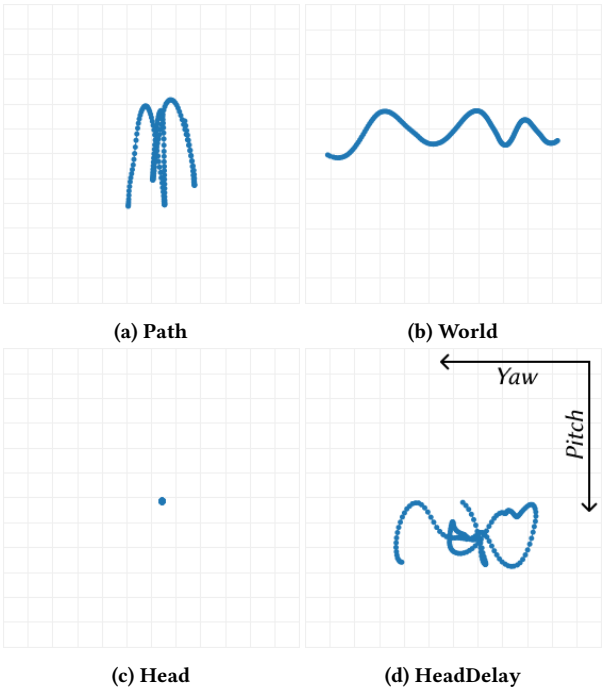


Figure 2: First-person view of the target movement relative to the person's head for the four reference frames.

Head. The head frame of reference's origin coincides with the head position while its orientation follows the head rotation (Figure 1a). In this reference frame the targets are located in front of the user at a specified distance (the Z coordinate). From the user's standpoint, the target does not move at all (Figure 2c). However, as previous studies have suggested [4], even then the eVOR is still active.

HeadDelay. This reference frame is similar to *Head* except it introduces a delay between head and the coordinate system's translation and rotation (Figure 1b). The delay is implemented as recommended by [30] with one exception: the translational delay along the Z axis was suppressed to avoid the targets getting closer or further from the user during the changes in walking speed which effectively let us keep the targets at a specified distance similar to *Head* (see Appendix A for more detail). From the perspective of a moving observer, the target exhibits the same movement as in *Path* plus an additional component caused by the delay itself (Figure 2d). The former implies that the eVOR will still help with stabilizing gaze on the target in this frame of reference, while the latter, since the target exhibits less predictable self-motion, that the contribution of the smooth pursuit eye movement should increase.

Lastly, the head gaze of a walking person usually fixates around one point in front of the person [41]. The two head reference frames differ from *Path* and *World* in one crucial regard: the latter two

allow turning the head towards the target if the user so desires. This is especially useful if the target is located further from the head fixation position because in this case people tend to involve the head and even the torso while acquiring such targets [60].

4 STUDY

In the current study, we examine the gaze behaviour of healthy individuals who were tasked to track virtual targets with their eyes during physical movement along a linear obstacle-free track while wearing a VR headset. The study is designed to address the following questions:

1. How does movement pace affect target acquisition time and tracking fixation accuracy?
2. How do the spatial reference frames compare among each other?
3. Does the possibility of voluntary head movement in the Path and World reference frames affect target acquisition in a different way?
4. Do target acquisition time and tracking fixation accuracy depend on the distance to the target?
5. How easy do the participants find the task of fixating the targets on the go?

We conducted two experiments that had an identical within-subject 3×4×4×2 factorial design with the factors being:

- Movement: 0 steps/min (Standing), 90 steps/min (Walking), 130 steps/min (Jogging)
- Reference Frame: Head, HeadDelay, Path, World
- Gaze Direction: North (N), South (S), East (E), West (W)
- Gaze Angle: 10° (Inner), 20° (Outer)

We varied the distance to the targets between the two experiments. We conducted the first experiment by placing the targets within the Head, HeadDelay, and Path reference frames closer to the participant, i.e. at a distance of 1m, where the effect of eVOR is more pronounced. Considering the nature of the World reference frame, however, within it we kept the 1m distance only for the Standing condition. For Walking and Jogging, the targets were always placed at the opposite end of the 5m track so that the participant would not reach them while moving toward them. After the first experiment was finished, we calculated the average distance from the participant to the target for the World reference frame. This calculation gave us 3m for Jogging and 3.6m for Walking and allowed us to inform the design of the second experiment. In the second experiment, to approximate the distance among all four reference frames during physical movement, we placed the targets at the distance of 3.3m — the average between the two numbers above.

4.1 HMD-administered Task

With this study, our goal was to simulate the acquisition of spatial UI controls, such as buttons, checkboxes, sliders, etc., with one's eyes. For that, we devised a task that is loosely based on the procedure that is used to measure the spatial accuracy of eye trackers (see[5] for an example). We intentionally diverged from the continuous tracking of a series of targets to a shorter task that can be repeated every time the participant is going from one end of the 5m track to another. Our HMD-administered task starts with the participant looking at the circular target located in the centre of the headset's FoV (Figure 3a), which we call the resting point. After looking at it for 0.7 s, the target disappears from the centre and instantaneously reappears at the periphery (Figure 3b). This signifies the beginning of a trial. During the Walking and Jogging conditions, a random seed that spans the range of two steps is added to the time to balance target acquisition in different phases of walking. The participant is tasked with following the target with their eyes and, after acquiring the target, keep fixating on it until they hear a confirmation sound which is played 1.2 s after the acquisition. The sound signifies the end of the current trial. At that moment, the target reappears in the centre, the participant acquires it, and the procedure restarts. For non-stationary conditions that means that the participant acquires one target at the periphery per one go, and their task is to move back and forth along the track until they are done with all targets.

Each target is displayed as a white circle with a black dot in the centre with the diameter of 1/6 of the outer circle. For Head, HeadDelay, and Path, each target keeps a fixed angular size of 2° regardless of the distance. However, for World the target maintains the same physical size which is equivalent to the angular size of 2° at the moment when the target jumps from the centre to the periphery. We took this approach to imitate the spatial UI of a fixed size that the person approaches.

4.2 Independent Variables

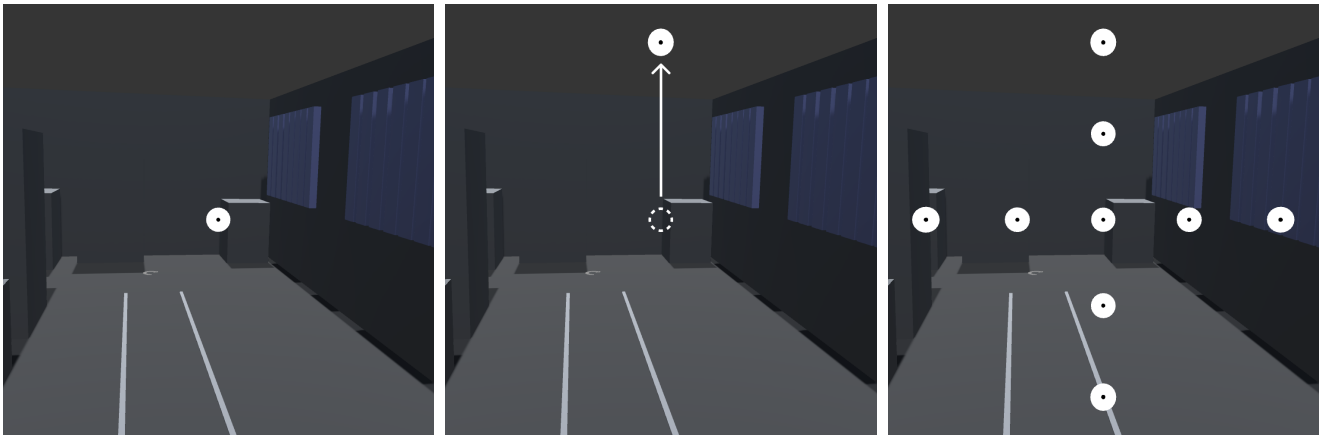
Movement. We used a metronome sound to synchronise the step frequency among the participants, similarly to other studies that involve uninstrumented locomotion [42]. The upper limit for jogging is 140 steps per minute (the pace above this number is considered running) [72], while brisk walking pace is considered to be 100 steps per minute [52]. We decided on values 10 steps per minute lower than these limits. Therefore, the levels of Movement are as follows: standing (0 beats per minute or no metronome sound), walking (90 beats per minute), and jogging (130 beats per minute). Post-hoc analysis demonstrated low standard deviations for both walking and jogging — 93.17 steps/min (SD=5.07) and 134.26 steps/min (SD=4.89), respectively — that indicates that the participants closely followed the set pace.

Reference Frame. The targets were located within three mobile (Head, HeadDelay, Path), and one stationary reference frame (World) which we consider a baseline. All the reference frames are described in detail in section 3.

Gaze Direction. To cover different directions in which gaze can move in a UI, we studied acquisition of targets located in four cardinal directions relative to the resting point: north, south, east, or west (Figure 3c).

Gaze Angle. We varied the visual angle between the resting point and the final target position. We were interested in comparing the reference frames that allow voluntary head rotation (Path and World) with those that do not (Head and HeadDelay) in cases where the target is below and above the angular threshold comfortably reachable by the eyes only (15° according to [60]). Thus, we chose 10° and 20° for Inner and Outer targets, respectively (Figure 3c). The participants were allowed to decide for themselves whether to

[5] https://www.tobii.com/resource-center/reports-and-papers/eye-tracking-performance-assessment



(a) (b) (c)

Figure 3: Representation of the target tracking task: the target moves instantaneously from the centre of the participant's FoV (a) to the periphery (b), in this case to the North. (c) illustrates all target positions.

rotate their head towards the target or not because we were motivated to observe natural target acquisition behaviour. All targets were placed at the same distance from the observer (Figure 4).

12 conditions (4 Reference Frames × 3 levels of Movement) were counterbalanced among participants using the balanced Latin Square[6]. The order of Gaze Angle and Gaze Direction was randomised. However, to avoid jumping from one gaze angle to another, we first randomised the order of Gaze Angles and then, within each Gaze Angle, the order of Gaze Directions.

4.3 Study Setup

4.3.1 Equipment & Study Environment. We used an HTC Vive Pro Eye VR headset with a built-in Tobii eye tracker connected to a tower PC with the AMD Ryzen 5 5600X CPU, Gigabyte GeForce RTX 3070Ti GPU, and 32GB of RAM which is well above the recommended PC requirements[7] for the headset. We used a tethered setup to firstly, avoid introducing an additional delay in data transfer between the headset and the PC, and secondly, to minimise slippage by using the cable as a counterweight for the front-heavy headset.

The study was conducted in a quiet, well-lit room, part of which was equipped to accommodate an 8m long and 3.2m wide obstacle-free walking area. To ensure seamless positional tracking across the area, four Vive Lighthouse base stations were mounted to a ramp at the height of 2.2-2.4m. At least two of the four base stations maintained a direct view of the headset at all times during physical movement.

4.3.2 Prototype. A prototype was created using Unity 2021.3.8f LTS. Within it, we created a virtual environment that represents a simplified replica of the physical study room with 3D geometric primitives (Figure 5). The environment was aligned with the physical space one-to-one. The 0.7m wide and 5m long virtual track represented by two lines on the floor was located symmetrically relative to a real-world position of the PC so that participants could freely reach both ends of the track without overextending the VR headset's cable. The track had 1m wide empty safe space on the sides and 1.5m wide space at both ends.

During each trial we continuously gathered data about the position and rotation of the participant's eyes and head. We turned off the built-in filtering mechanism of the eye tracker to gather raw data. The eye tracking data was sampled at 120Hz via SRanipal API[8] and fused with the head pose we got from the Lighthouse positional tracking system directly via the OpenVR plugin (i.e. by calling openvr_api.dll from C# bypassing Unity XR API).

4.4 Participants

We recruited 36 able-bodied participants (11 women), 24 for the first experiment and 12 for the second, aged 22-37 (M = 28; SD = 3.8). Participation in the study was voluntary and did not imply any monetary reward.

4.5 Procedure

Every study session began with a participant reading and signing an informed consent form and a GDPR-related form informing them about processing of their personal data, followed by filling in a demographics questionnaire. Then the participant would be given a brief summary of the study, including initial guidance on the task, and instructed on how to ensure the headset's fit and visual clarity. The participants were also instructed to put cable aside with one hand before starting to walk or jog and to follow a visual cue located at both ends of the track which suggested the appropriate direction of turning (the arrows on the floor, see Figure 5). To make participants feel more safe in the virtual environment, after putting on and fitting the headset, every participant was given a chance to freely walk around the room and touch virtual walls to reassure them that they coincide with physical ones.

[6]https://cs.uwaterloo.ca/ dmasson/tools/latin_square/

[7]https://www.vive.com/eu/support/vive-pro-eye/category_howto/what-are-the-system-requirements.html

[8]https://developer.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/overview/

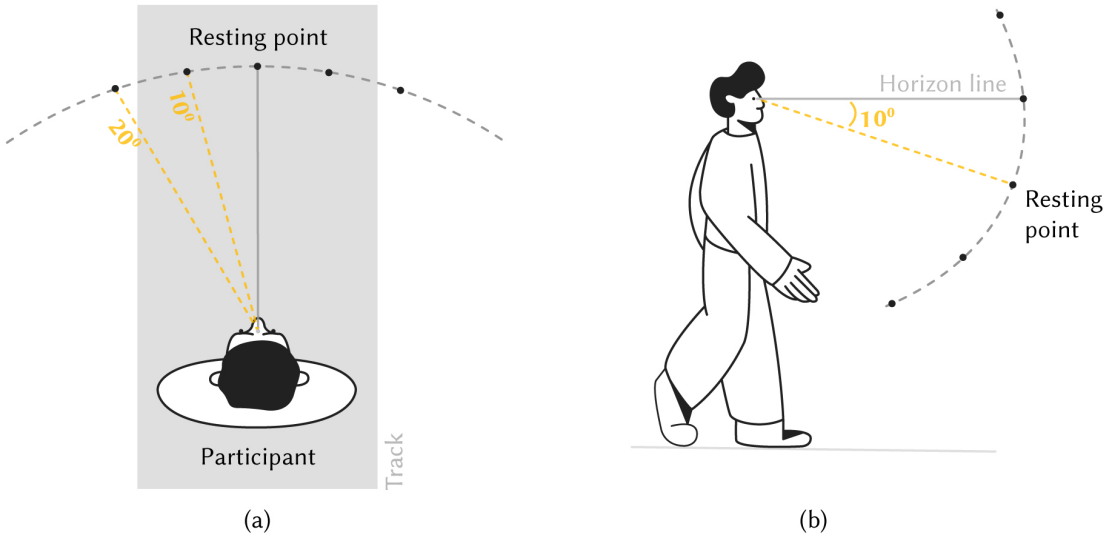


Figure 4: Top view (a) and a view from the side (b) of a participant walking along the track with the targets in the path reference frame. All targets are on to illustrate their relative position. In a real experiment the participant would see only one target at a time.

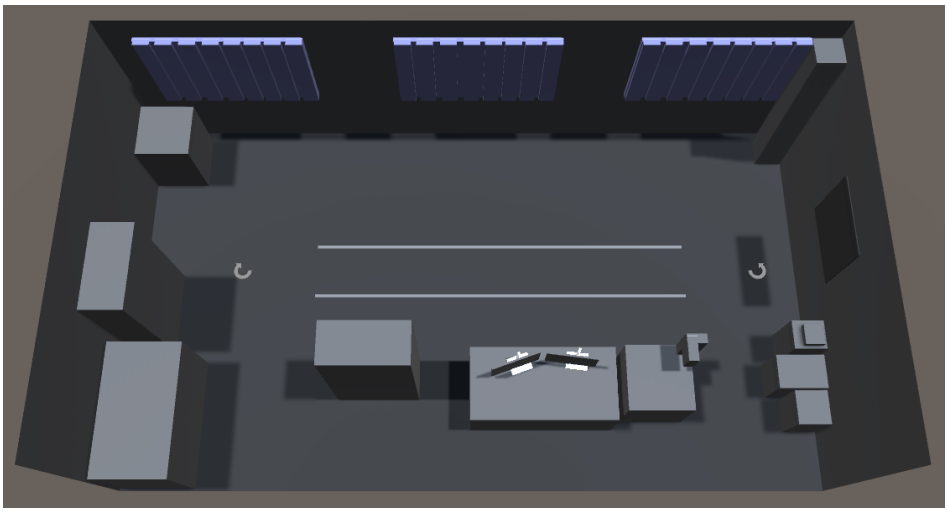


Figure 5: A virtual replica of the room where the study was conducted that represents a configuration of the physical space, as virtual and physical objects were aligned one-to-one.

Regardless of the order of the experimental conditions, the study would begin with a training of the target acquisition task in the stationary condition when a participant would be instructed to look at the targets as they would normally do, and minimize blinking while fixating the target after it reappears at the periphery. The training of walking and jogging to the sound of a metronome without the targets present would commence right before the first walking and jogging conditions, respectively. Shorter retraining with no targets was conducted every time when the condition would change from walking to jogging or the other way around, because during pilots we noticed that some participants are having troubles switching between the rhythms. Every one out of 12 conditions would start with a training. The training was shorter towards the end of the study and could consist of a single run (i.e., one target) with the specified reference frame and the rhythm. Eye tracking calibration was ran before every condition. For each condition, we maintained a queue of target locations. Whenever a trial failed, it would be sent to the end of the queue. To ensure that the task is completed accurately, the conditions of failure include the gaze dropping off

the target more than a set threshold for 3 frames in a row, the participant stopping in the middle of a trial, or walking outside of the track width. Furthermore, a moderator could mark a trial as failed if, for example, they noticed that the participant did not maintain the set pace.

Participants were informed that they could ask to take a break at any time they desired. However, if they did not, a single planned break was conducted halfway through the 12 conditions. After finishing all 12, the participant was invited to fill in the final questionnaire asking them to compare reference frames among themselves. After that the debriefing would commence and the participant was given their copies of the documents they signed at the beginning. The experimental session took 1.5 hours on average.

4.6 Dependent Variables

In this study, we use *target acquisition time* to characterize acquisition speed, and *trueness* and *precision* to characterize tracking fixation accuracy, as measures widely used in literature [14, 24]. Before defining them, it is important to note that we consider tracking fixation precision a measure of how stable gaze fixates on the object of interest, similar to the retinal slip used elsewhere [4]. This means that the results described below pertaining to precision affect such activities of an HMD user as reading text or looking for a UI control within a spatial interface, even if it is not controlled by gaze, studying information presented graphically, etc. At the same time, the combination of tracking fixation trueness and precision, and target acquisition time can help characterize the performance of gaze when used as input.

4.6.1 Target Acquisition Time. We define target acquisition time as the time required for gaze to settle on the target. A typical trial consists of three phases (Figure 6a): (1) The participant keeps looking at the resting point "by inertia". Note that 0ms in every trial represents the moment when the target moves from the center of the FoV to the periphery. The length of this period is mainly determined by individual reaction time. (2) A series of gaze shifts when the participant acquires the target with their eyes which can be as short as the time needed for a single saccade. (3) Tracking fixation which on the go consists of pursuit and vestibulo-ocular eye movements combined to various degrees depending on the reference frame. In the stationary condition, this period consists mostly of oculomotor fixation. We define borders among these three regions as the beginning of the first and the end of the last gaze shift. Thus, *the target acquisition time is the time passed from the moment the target moved to the periphery until the end of the last saccade.*

To identify saccades during locomotion we used the algorithm from [62] which is based on simultaneously applying the thresholds for angular gaze velocity (> 240 deg/s) and acceleration (> 3000 deg/s^2). However, before employing this algorithm we filtered data using a 9-point wide median filter (75 ms delay at 120Hz) to remove smaller jumps in the gaze data as recommended by [46]. The beginnings and ends of saccades were identified as the frames before and after the angular acceleration local maxima and minima respectively, as described in [7, p. 12]. The last thing we did to calculate the target acquisition time is we capped the period within which we looked for saccade endings by 700ms which contains 91.3% of all identified gaze shifts as another outlier removal step.

4.6.2 Tracking Fixation Precision. We define precision as the closeness of measured values to each other according to [15] and use *the standard deviation of angular differences between the mean gaze direction and each data point starting from the end of the last gaze shift until the end of a trial* as a measure of tracking fixation precision:

$$precision = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(g_i \angle \overline{g})^2} \tag{1}$$

where $g_i = (x_i, y_i, z_i)$ is the i-th gaze direction, $\overline{g}$ is the mean gaze direction, i.e. $\overline{g} = \frac{1}{N}\sum_{i=1}^{N} g_i$, while $\angle$ denotes an angle between two vectors. We performed the calculation of precision within a constructed coordinate system known as the target forward coordinate system [27]. The coordinate system's origin coincides with the gaze origin, and is always oriented towards the current target, allowing us to account for the target's movement relative to the head.

4.6.3 Tracking Fixation Trueness. We define trueness (a.k.a. spatial accuracy) according to [15] as the closeness of the average of measured values to the reference value, which in our case is the target position, and calculate it as *an angle between the vector from the gaze origin to the target and the mean gaze direction computed on a set of data points starting from the end of the last gaze shift until the end of a trial*, as follows:

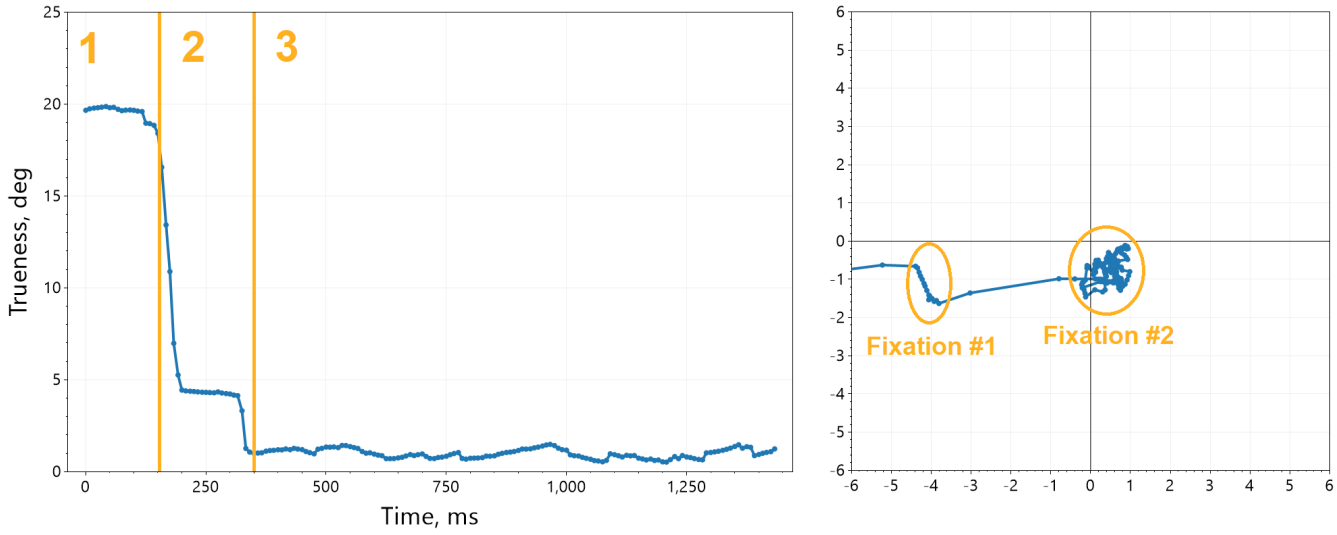
$$trueness = \overline{g} \angle t' \tag{2}$$

where $\overline{g}$ is the mean gaze direction, t' is the vector from the gaze origin to the target, $\angle$ denotes an angle between two vectors. We calculated trueness within the same constructed coordinate system as we do the precision, therefore, t' is always equal to $(0, 0, 1)$. It is important to note that we calculated both trueness and precision on raw gaze data with outliers removed as described in the next section.

4.6.4 Subjective Assessment. We asked the participants to rate reference frames on a 7-point Likert scale against the following criteria: Ease of gaze pointing; Ease of movement; Single Ease Question (SEQ) [55].

4.7 Results

All data was pre-processed before analysis. Only successful trials were used, i.e. 3,448 trials in total (3 movement paces × 4 reference frames × 4 gaze directions × 2 gaze angles × 36 participants). Within each trial, data points marked by the eye tracker as 'Invalid' at least for one eye along with 5 data points before and after flags were removed. We did not interpolate the resulting gaps due to occasional long gaps of missing data. Local (eyes-in-head) combined gaze direction and origin were calculated from the gaze directions and origins of the left and right eyes. The data in the local coordinate system was converted to a global one (eyes-in-world) using the known head pose. One of the 432 total conditions was missing, i.e. 8 trials. We used winsorization to fill in the values, i.e. the maximum or minimum values over all participants for each dependent variable.



(a) (b)

Figure 6: The plot to the left (a) represents how far the gaze is from the target depending on time. The vertical lines are the beginning of the first and the end of the second and last gaze shift. The plot to the right (b) represents a scanpath for the same trial. The first tracking fixation occurs during the second period while the second fixation during the third as it is shown on the plot (a).

Outliers were removed using the threshold detection algorithm described in [62, Section 3E]. The only difference from this algorithm was that we would merge outliers if the gap between the two consecutive ones would be less than or equal to 3 frames (25ms at 120Hz). Angular velocity and acceleration were calculated using the eyes-in-world data as recommended by [7, Eq. 13]. The threshold value of 750 deg/sec for the highest possible velocity of a gaze shift was taken from [8], while the value of 75 000 deg/sec^2 for acceleration was calculated from the velocity threshold similar to [62]. The resulting gaps would be linearly interpolated.

The analysis for the main experiment was performed employing a 4-way repeated measures ANOVA ($\alpha = .05$) using Movement, Reference Frame, Gaze Angle, Gaze Direction as within-subject factors. We used the aligned rank transform [9, 70] on all continuous variables that according to the Shapiro-Wilk test were not normally distributed within most of the conditions. Whenever the assumption of sphericity was violated, we applied the Greenhouse-Geisser correction. Bonferroni-corrected pairwise comparisons were used when applicable.

After the second experiment, we combined the data from all 36 participants to investigate the effect of distance between the two studies. Therefore, we treated Distance as an additional independent variable and analyzed the data using a 5-way mixed ANOVA using Movement, Reference Frame, Gaze Angle, Gaze Direction as within-subject factors and Distance as a between-subject factor. Regarding the effects of the first within-subject factors, the results reiterated the results of the first study (Table 1). Friedman tests were used to analyze subjective factors within Reference Frames. Conover's post-hoc tests were employed when applicable. A Kruskal-Wallis test was used to analyze Distance as a between-subject factor. Below we report on the effects that are relevant to the study's research questions (Table 1). The results for other effects can be found in the supplementary material.

RQ1. How does movement pace affect target acquisition time and tracking fixation accuracy? Our results showed that Movement has a detrimental effect on all dependent variables in various degrees. While acquisition time is affected by Movement, we did not find a significant difference between Walking and Jogging (p=.538). It took 243ms (SD=73) to stabilize gaze on the target while standing, and 275s (SD=97.4) on average while moving (p<.001). Trueness deteriorated at a higher pace, starting at 1.01±0.76° for Standing, which corresponds to the value provided by the manufacturer (between 0.5 and 1.1°[9]), followed by 1.13±0.86° for Walking, and 1.31±0.91° for Jogging. Precision decreased substantially when the pace increased — Standing (0.6±0.45°), Walking (0.9±0.42°), and Jogging (1.41±0.69°), all being significantly different (p<.001).

RQ2. How do the spatial reference frames compare among each other? Reference Frame was significant for both target acquisition time and tracking fixation accuracy. For the former, HeadDelay, Path, and World showed no significant differences (257.92±84.37ms on average), while Head took longer at 283.34±107.16ms (p<.001). Figure 7 showcases the different behaviors: 1) the time distribution for Head has a thicker tail, 2) the later peak for Head indicates a higher probability of at least one follow-up gaze shift (see Figure 6a). Trueness was least affected by Reference Frame with Head (1.23±0.88°) and World (1.06±0.83°) being the only significantly different pair (p=.004), while the rest occupied the space between them. For precision, no significant difference was revealed for Path and World (p=.093). Their precision of 0.83±0.54° (combined) was followed

[9]https://developer.vive.com/eu/support/sdk/category_howto/trackable-field-of-view.html

Table 1: Statistical analysis of the effects that are relevant to the study's research questions.

24 Participants

Variable	ANOVA			
	Effect	F-value	p	η_p^2
Acquisition Time	M	F(2, 46)=27.811	<.001	.547
	RF	F(3, 69)=20.804	<.001	.475
	GA	F(1, 23)=145.151	<.001	.863
	M×RF	F(6, 138)=10.298	<.001	.309
	RF×GA	F(3, 69)=3.294	.026	.125
Trueness	M	F(2, 46)=16.526	<.001	.418
	RF	F(3, 69)=4.339	.007	.159
	GA	F(1, 23)=46.677	<.001	.67
	M×RF	F(6, 138)=2.256	.042	.089
	RF×GA	F(3, 69)=0.848	.472	.036
Precision	M	F(1.61, 37.022)=295.34	<.001	.928
	RF	F(3, 69)=89.294	<.001	.795
	GA	F(1, 23)=106.353	<.001	.822
	M×RF	F(6, 138)=41.351	<.001	.643
	RF×GA	F(3, 69)=1.286	.286	.053

36 participants, Subjective Assessment

Variable	Non-parametric Tests			
	Effect	χ^2	p	
Ease of Gaze Pointing	RF	$\chi^2(3)$=0.106	.01	-
	D	$\chi^2(1)$=1.197	.274	-
Ease of Movement	RF	$\chi^2(3)$=0.105	.01	-
	D	$\chi^2(1)$=9.906	.002	-
SEQ	RF	$\chi^2(3)$=0.124	.004	-
	D	$\chi^2(1)$=0.014	.905	-

36 Participants

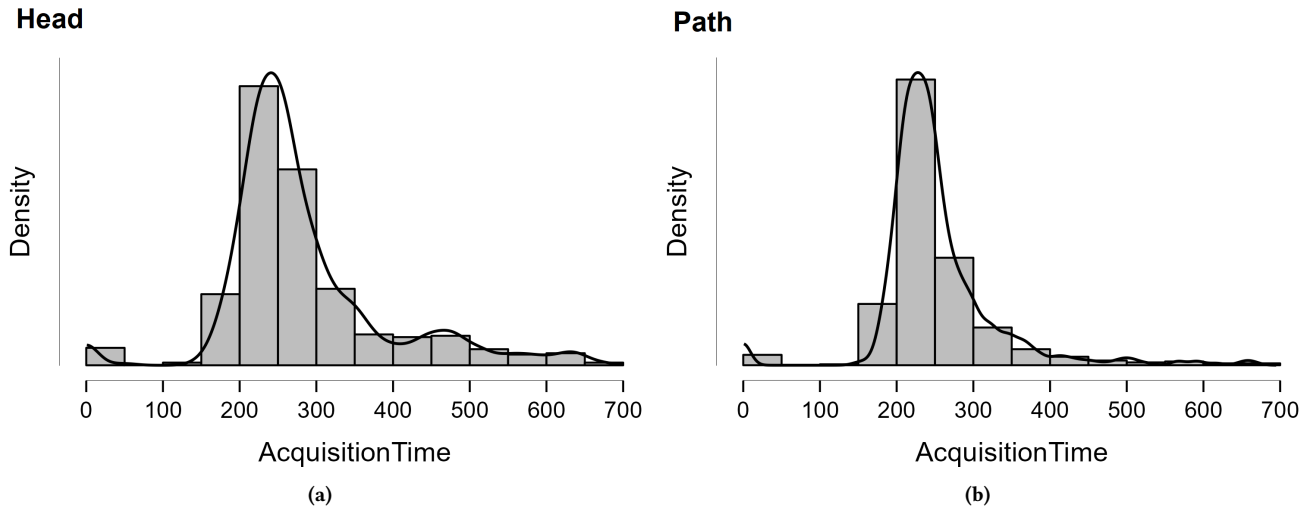
Variable	ANOVA			
	Effect	F-value	p	η_p^2
Acquisition Time	M	F(2, 68)=48.988	<.001	.59
	RF	F(3, 102)=25.114	<.001	.425
	GA	F(1, 34)=165.155	<.001	.829
	D	F(1, 34)=1.584	.217	.045
	M×RF	F(6, 204)=12.724	<.001	.272
	RF×GA	F(2.352, 79.975)=6.063	.002	.151
	D×RF	F(3, 102)=1.035	.38	.03
Trueness	M	F(2, 68)=26.699	<.001	.44
	RF	F(3, 102)=5.505	.002	.139
	GA	F(1, 34)=56.295	<.001	.623
	D	F(1, 34)=0.379	.542	.011
	M×RF	F(6, 204)=2.191	.045	.061
	RF×GA	F(3, 102)=1.37	.256	.039
	D×RF	F(3, 102)=0.142	.934	.004
Precision	M	F(2, 68)=432.318	<.001	.927
	RF	F(2.204, 74.923)=171.478	<.001	.835
	GA	F(1, 34)=177.727	<.001	.839
	D	F(1, 34)=1.209	.279	.034
	M×RF	F(4.677, 159.016)=78.77	<.001	.699
	RF×GA	F(3, 102)=.934	.427	.027
	D×RF	F(2.13, 72.422)=8.407	<.001	.198

by HeadDelay (1.01±0.59°), and finally, Head (1.21±0.75°), all being significantly different among each other (p<.001). Notably, neither in terms of trueness, nor precision, Path did not differ from World. This can indicate that vergence does not affect the tracking fixation accuracy as was previously thought [24].

Diving deeper into the differences, we found the Movement×Reference Frame interaction to be significant for acquisition time. As shown in Figure 8a, the difference mainly stems from gaze taking longer to settle on the targets within Head on the move than within the other reference frames. Notably, Path is the only frame of reference that does not differ across any of movement levels. Similarly, we found a significant Movement×Reference Frame interaction for trueness (Figure 8b). The interaction effect mainly originates from tracking fixation being less true during Jogging than during Standing for Head (p=.001) and Path (p=.003). We also found a significant Movement×Reference Frame interaction for precision (Figure 8c). Post-hoc tests revealed no significant differences among the four reference frames in the Standing condition (0.6±0.45° on average). For Walking, the reference frames with the highest precision — Path and World — did not differ from each other (p=1.0). For Jogging, only Head and HeadDelay that showed the lowest performance were not significantly different (p=.495).

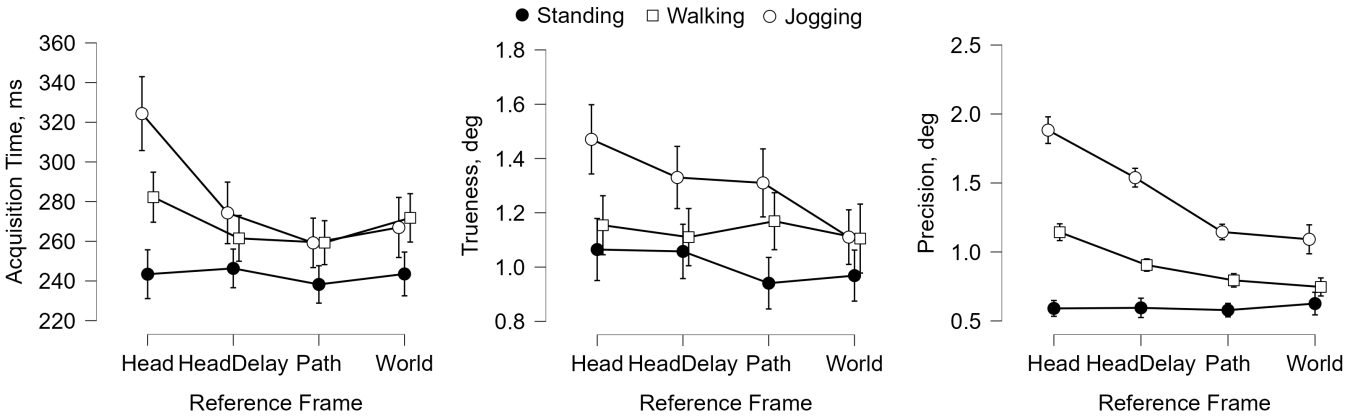
Precision deteriorating faster with higher pace for both head reference frames than for Path and World can be explained by the effect of eVOR. To test this conjecture, we calculated the correlation between eye-in-head and head-in-world rotations during tracking fixation in horizontal and vertical planes separately. To avoid the effects of voluntary head rotation common to Path and World, when analyzing the correlation in horizontal plane (yaw) we excluded the targets E and W. Respectively, we excluded N and S, when conducting calculations for the vertical plane (pitch). As can be seen in Table 2, the existence of negative correlation during both walking and jogging indicates that eVOR is present regardless of the reference frame. The vertical and lateral head oscillations caused by natural locomotion are compensated by it within Path and World, while within Head, where the lowest precision can be reached by minimizing eye movement, the same eVOR 'swings' gaze around the target negatively affects tracking fixation precision. Within HeadDelay, where the target exhibits both the repetitive movement predicated on the head oscillation and the movement caused by the delay itself (Figure 2d), the effect of eVOR is diminished, however, as opposed to Head, it still works towards stabilizing gaze on the target. Thus, the HeadDelay results for precision are in between Head and Path/World. Trueness, as our results indicate, is less affected by eVOR because gaze oscillations affect the averaged gaze less (Equation 2).

RQ3. Does the possibility of voluntary head movement in the Path and World reference frames affect target acquisition in a different way? The Reference Frame×Gaze Angle interaction characterizes the effect of voluntary head movement. We found that the interaction for target acquisition time was significant (Figure 9a). It is mostly based on the fact that the further the target from the resting point, the closer HeadDelay is getting to Head: Δ28ms for Inner (p < 0.001) vs. Δ17ms for the Outer targets (p = 0.14). For Path and World, the



(a) (b)

Figure 7: Distribution of the target acquisition time for the Head (a) and Path (b) reference frames.



(a) (b) (c)

Figure 8: Mean acquisition time (a), trueness (b), and precision (c) for four Reference Frames depending on Movement. Error bars indicate 95% confidence intervals.

Table 2: Correlation between head and eye movements in horizontal (yaw) and vertical (pitch) planes. The first number represents mean Pearson's correlation coefficient for each condition, the second — standard deviation.

	Horizontal				Vertical			
	Head	*HeadDelay*	*Path*	*World*	*Head*	*HeadDelay*	*Path*	*World*
Walking	-0.399±0.233	-0.433±0.313	-0.441±0.301	-0.775±0.182	-0.578±0.16	-0.515±0.218	-0.615±0.2	-0.762±0.127
Jogging	-0.518±0.177	-0.75±0.199	-0.735±0.181	-0.856±0.081	-0.425±0.172	-0.376±0.25	-0.526±0.269	-0.644±0.302

acquisition time increases similarly to Head, which indicates that head movement supporting a saccade in these reference frames did not have an effect on acquisition speed. Moreover, we did not find Reference Frame×Gaze Angle interaction to be significant for both trueness and precision, which is puzzling. One would assume that the deterioration of tracking fixation accuracy for the Outer targets would be less pronounced because the participant was able to move the target closer to the FoV center by turning their head where both trueness and precision are higher [61], while for both head reference frames the difference between the Inner and Outer targets should be significant. One possible explanation for this is that not all participants used this possibility, which according to our data is true for Standing but not for on the go conditions. Thus, we conducted an additional analysis having the data for the stationary condition excluded for all reference frames. Even then, the ART RM-ANOVA did not reveal a significant Reference Frame×Gaze Angle

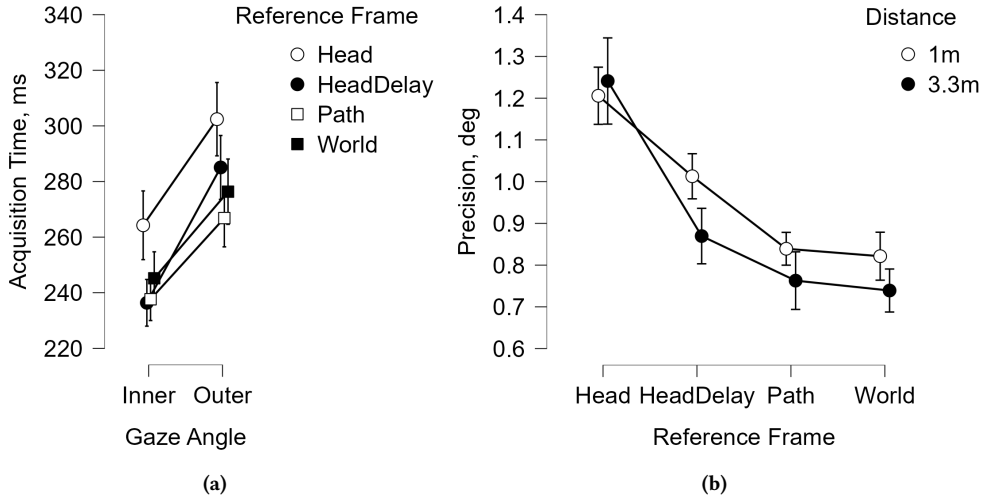


Figure 9: Mean acquisition time for the four Reference Frames depending on Gaze Angle (a) and mean precision for the four Reference Frames depending on Distance (b). Error bars indicate 95% confidence intervals.

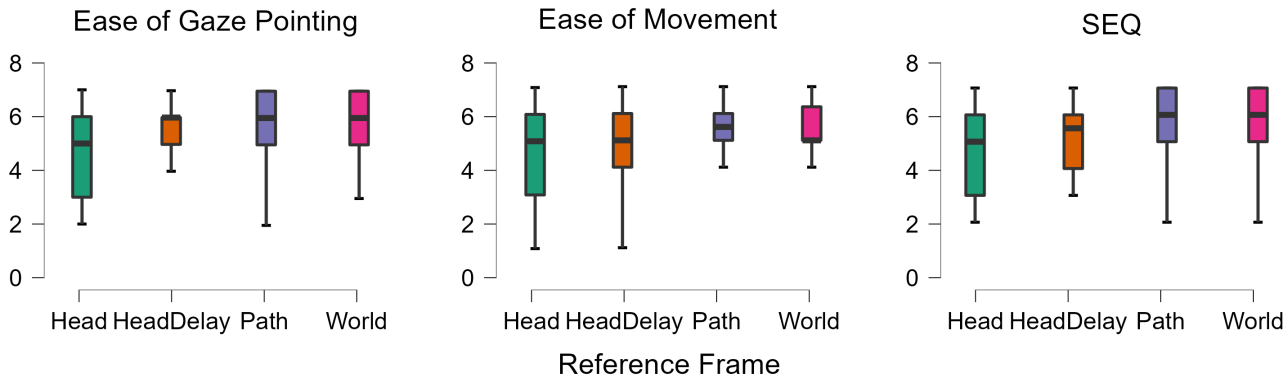


Figure 10: Boxplots depicting median values, interquartile ranges (the 25th and 75th percentiles), and minima/maxima of the responses on the Single Ease Question and questions about ease of eye gaze pointing and physical movement.

interaction for both the trueness ($F_{3,99}$=.29, p=.835) and precision ($F_{3,99}$=.72, p=.542).

The collected data indicates that acquiring the Inner targets takes 245.88±80.43ms, as opposed to the Outer targets that take 282.67±97.54ms to acquire (p<.001). Fixating the Inner targets was also more accurate, i.e. 0.32° more true for Inner (0.99±0.64°) than for Outer (1.31±0.99°), and 0.2° more precise for Inner (0.87±0.59°) than for Outer (1.07±0.66°). However, the former says more about the longer saccades taking more time to land, while the latter about the characteristics of the built-in eye tracker in the HTC Vive Pro Eye. Thus, no matter the reference frame, acquisition of the Outer targets takes more time, and the following tracking fixation is less accurate.

RQ4. Do target acquisition time and tracking fixation accuracy depend on the distance to the target? We found a significant interaction effect of Reference Frame×Distance only on precision (Figure 9b). This interaction mainly originates from HeadDelay at 3.3m not being significantly different from Path and World at 1m, while the same HeadDelay at 1m differs from Path and World at 1m (p<.001). We did not find Distance to be significant for any of the dependent variables, indicating that neither acquisition speed, nor accuracy are dependent on the distance to the target.

RQ5. How easy do the participants find the task of fixating the targets on the go? The participants perceived both tracking targets with their eyes and moving physically to be easier with Path and World than with Head, while HeadDelay occupies the middle ground. In particular, *gaze pointing* (Figure 10, left) with Head was perceived as significantly more difficult than HeadDelay (p=.018), Path (p=.032), and World (p=.006). 11 out of 36 participants mentioned that the targets within the Head reference frame felt too jumpy, nine participants said the same about HeadDelay, and only two about Path. Eight participants specified that World felt stable, while seven participants mentioned the same about Path. Interestingly, 13 out of 36 participants mentioned that it is harder to track targets within

Table 3: Time in milliseconds when gaze settles on the target. The numbers below are equal to the 95th percentile of target acquisition times for different conditions.

	Head	*HeadDelay*	*Path*	*World*
Standing	294	277	270	289
Walking	347	309	324	369
Jogging	425	326	337	339

World as they leave the FoV during an approach. Five participants found that rotating the head towards a target made it easier to focus on it. Distance did not influence the ease of pointing.

Regarding *ease of movement* (Figure 10, centre), Head was perceived as harder than Path (p=.045) and World (p=.045). This may be due to dual-task performance: the harder it is to track the targets, the less attention is left to give to the movement. Similar results were found by Mustonen et al. [42] who found that the amount of cognitive workload was responsible for the deterioration of walking performance. We also found that Distance had a significant effect. Participants found it easier to maintain their pace while fixating on targets located at 3.3m (M = 6) than on targets located at 1m (M = 5). A possible explanation could be that focusing on farther targets minimises the parallax effect, thus making it easier to control the gait. Finally, in our *SEQ* results (Figure 10, right), Head was significantly more difficult than Path (p= .008), and World (p=.017).

5 DISCUSSION

Implications for Gaze-based Selection. If gaze is used as an input modality for pointing, one should take into account both the deterioration of tracking fixation accuracy with higher pace and the effect of reference frame on precision on the go. One possible way to counter the former is to control the scale of a spatial UI depending on the current pace. Not only can it potentially make reading of UI text easier [4] but also increase selection performance [5]. Another promising approach is to stop using the line of sight as a direct pointing mechanism altogether and rely on the correlation between eye movement and a UI control trajectory [68]. It is interesting to see how this approach works during locomotion, specifically for HeadDelay where one can assign different parameters of inertia for different objects of interest to affect their self-motion tracked by the smooth pursuit system to a greater extent. Using both approaches, the pointing will depend on the tracking fixation precision. The use of run-time filtering techniques has been shown to successfully increase precision [14]. However, simply transferring the filtering techniques that were developed for use in stationary conditions to use on the go will not work because of increased noise levels caused by locomotion itself. More research is necessary to explore potential run-time filters suitable for use on the move.

The use of dwell time as an activation mechanism should take into account our results for target acquisition time. Considering that gaze stabilizes on the target with different speeds, dwell times shorter than the target acquisition time may lead to the selection coinciding with a gaze shift. Using the 95th percentile (Table 3) as the threshold value for fixed or adaptive dwell time [37], can help minimize such selection error.

Implications for Spatial UI Placement on the Go. If an HMD supports only rotational tracking (e.g., Xreal Air, Rokid Air Pro, etc.), HeadDelay can be used for 3D UI placement instead of the Head reference frame. With it, at the walking pace, 3D UI controls and information can be fixated as quickly as within Path and negligibly less precise. However, for positionally tracked devices, Path provides a better alternative. Within it, the speed and precision with which content can be fixated degrades slower with the higher pace. It also demonstrates superior precision across various distances, which can be important if the spatial content is presented at different depths, be it the placement of 2D widgets [35] or a 3D visualization [71]. We can speculate that with a longer delay for linear locomotion the performance of HeadDelay will get closer to Path because it should decrease the negative influence of the target self-motion on fixation precision. However, selecting the optimal amount of delay is something that should be studied additionally. Our results also indicate that in all these frames of reference, it is advantageous to place a UI closer to the head fixation position, i.e. the point where head gaze stabilizes in front of the person during locomotion.

Although not inherently mobile, the world reference frame provides 3D content placement that allows for its quick and precise acquisition. It seems to be especially relevant, when the content is semantically coupled to the environment, such as visualizations placed near a whiteboard in an office [10]. One can imagine an application in which spatial UI is distributed among several reference frames, e.g., a navigational app where a recommended path is laid out in World, while a mini-map floats in front of the user slightly lower than the normal line-of-sight during walking and is path-referenced — the former is integrated into the surroundings, the latter is available at a glance and does not occlude the view.

Knowing what we know now, we can speculate about the overall visual performance of some existing reference frames during linear locomotion, specifically, the ones proposed by Klose et al. [26] and in Microsoft MR Design Guidelines [40], and the 'true' body reference frame where torso position is tracked [35, 73]. Klose et al.'s simulated body reference frame lands itself in between HeadDelay and Path because within it, content does not move vertically as in Path but moves horizontally following lateral head oscillation with a delay as in HeadDelay. It is interesting to see how close it can get to Path by varying the amount of delay. Microsoft's solution that uses a threshold value for reference frame horizontal rotation should perform on a par or worse than Klose et al.'s because in it 3D content can be several degrees off the direction of locomotion which, according to our results, has a detrimental effect on overall visual performance. The 'true' body reference frame should demonstrate the lowest performance among these three frames of reference because on the go human's torso oscillates left and right, and up and down, similar to the head, keeping 3D content less stable than Klose et al. and Microsoft's solutions do.

Potential Uses for the Path Reference Frame. In AR context, the use of path-referenced content is advantageous in cases where the user needs to actively switch attention between the virtual content and surrounding environment without experiencing occlusion by the content as in the aforementioned navigational app or, for instance, during walking meetings [6] where participants might need access

to the content relevant to their dialog. In line with the Head-Glance design [35], when information is needed quickly but checked sporadically as, for example, today's schedule, placing it within Path outside of the user's FoV can potentially affords fast access without cluttering the view. When gaze is used as an input modality, path-referenced UIs afford higher pointing accuracy which is especially important when the number of options in a UI increases as, for example, during eye typing [43]. The same advantages are applicable in VR context, given that natural locomotion is used. Moreover, it would be interesting to see whether our findings also stand for locomotion enabled by omni-directional treadmills.

Limitations & Future Work. Several limitations need to be considered. Although mainly motivated by use cases for AR on the move, our study was conducted in VR. The use of VR to simulate AR conditions afforded the control needed to get detailed insight. Other studies have demonstrated the generalizability of results from VR to AR [19, 31, 32, 38] and we would expect that our conclusions generalize accordingly.

We used linear obstacle-free locomotion in our study. While linear movement is common, real-world movement will naturally involve variations in speed and direction. Investigating the effect of spatial reference frames during free-form locomotion along with studies on their obstacle avoidance ability is the logical next step. Real usage may also involve situational influences such as wayfinding and obstacles to negotiate. However, research on mobile device use while walking has shown that people are skilled in adapting their visual search strategies to incorporate device interaction in a safe manner [65].

Our study results were obtained with specific hardware, i.e. HTC Vive Pro's eye tracker designed by Tobii, on which reported absolute values of both tracking fixation trueness and precision are predicated. Our conclusions are drawn from relative values of these measures and thus independent of the device used. The absolute values of the target acquisition time reported should be robust to change of the eye tracker as it was calculated based on analysis of saccadic eye movements.

6 CONCLUSION

In this study, we compared how quickly and accurately users can acquire and track targets affixed in the world-, novel path-, and two head-based reference frames with their eyes while moving along a virtual path with two different paces and while stationary. Our results demonstrate that the more stable the target relative to the environment along the axes perpendicular to the direction of locomotion, the faster and more precise the gaze fixation. On the contrary, tracking fixation trueness depends mostly on the pace of movement and changes based on the reference frame in a more nuanced way. Participants perceived the differences between reference frames in terms of how hard it was to track targets on the move. Their subjective assessment corresponds to the objective results. Our results can be applied in the design of spatial applications used during physical movement.

ACKNOWLEDGMENTS

This work was partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant no. 101021229 GEMINI). The authors would like to express their gratitude to Bekir Tekmen for his help with conducting the study. The first author wants to give special thanks to Stasia Manakhova for her patience and support.

REFERENCES

[1] Rony Abovitz, Brian T. Schowengerdt, and Matthew D. Watson. 2015. Planar waveguide apparatus with diffraction element(s) and system employing same. https://patents.google.com/patent/US20150016777A1/en

[2] Joanna Bergstrom-Lehtovirta, Antti Oulasvirta, and Stephen Brewster. 2011. The Effects of Walking Speed on Target Acquisition on a Touchscreen Interface. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) *(MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/2037373.2037396

[3] M. Billinghurst, J. Bowskill, M. Jessop, and J. Morphett. 1998. A wearable spatial conferencing space. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*. 76–83. https://doi.org/10.1109/ISWC.1998.729532

[4] Olivier Borg, Remy Casanova, and Reinoud J. Bootsma. 2015. Reading from a Head-Fixed Display during Walking: Adverse Effects of Gaze Stabilization Mechanisms. *PLOS ONE* 10, 6 (June 2015), e0129902. https://doi.org/10.1371/journal.pone.0129902

[5] Stephen Brewster. 2002. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous computing* 6 (2002), 188–205.

[6] Ida Damen, Carine Lallemand, Rens Brankaert, Aarnout Brombacher, Pieter van Wesemael, and Steven Vos. 2020. Understanding Walking Meetings: Drivers and Barriers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376141

[7] Gabriel Diaz, Joseph Cooper, Dmitry Kit, and Mary Hayhoe. 2013. Real-time recording and classification of eye movements in an immersive virtual environment. *Journal of Vision* 13, 12 (10 2013), 5–5. https://doi.org/10.1167/13.12.5 arXiv:https://arvojournals.org/arvo/content_public/journal/jov/932808/i1534-7362-13-12-5.pdf

[8] Andrew T. Duchowski. 2003. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, Berlin, Heidelberg.

[9] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 754–768. https://doi.org/10.1145/3472749.3474784

[10] Mats Ole Ellenberg, Marc Satkowski, Weizhou Luo, and Raimund Dachselt. 2023. Spatiality and Semantics - Towards Understanding Content Placement in Mixed Reality. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 254, 8 pages. https://doi.org/10.1145/3544549.3585853

[11] Roar Gauslå Engell. 2018. *Head-mounted Augmented Reality for Outdoor Pedestrian Navigation*. Master's thesis. Technische Universität München.

[12] Barrett Ens, Juan David Hincapié-Ramos, and Pourang Irani. 2014. Ethereal Planes: A Design Framework for 2D Information Space in 3D Mixed Reality Environments. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction* (Honolulu, Hawaii, USA) *(SUI '14)*. Association for Computing Machinery, New York, NY, USA, 2–12. https://doi.org/10.1145/2659766.2659769

[13] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the World: 2D Windows for 3D Augmented Reality. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology* (Atlanta, Georgia, USA) *(UIST '93)*. Association for Computing Machinery, New York, NY, USA, 145–155. https://doi.org/10.1145/168642.168657

[14] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1118–1130. https://doi.org/10.1145/3025453.3025599

[15] Joint Committee for Guides in Metrology. 2021. International Vocabulary of Metrology, Fourth edition – Committee Draft (VIM4 CD).

[16] Shogo Fukushima, Takeo Hamada, and Ari Hautasaari. 2020. Comparing World and Screen Coordinate Systems in Optical See-Through Head-Mounted Displays for Text Readability while Walking. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 649–658. https://doi.org/10.1109/ISMAR50242.2020.00093

[17] Çağlar Genç, Shoaib Soomro, Yalçın Duyan, Selim Ölçer, Fuat Balcı, Hakan Ürey, and Oğuzhan Özcan. 2016. Head Mounted Projection Display & Visual Attention:

Visual Attentional Processing of Head Referenced Static and Dynamic Displays While in Motion and Standing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1538–1547. https://doi.org/10.1145/2858036.2858449

[18] Jens Grubert, Matthias Heinisch, Aaron Quigley, and Dieter Schmalstieg. 2015. MultiFi: Multi Fidelity Interaction with Displays On and Around the Body. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3933–3942. https://doi.org/10.1145/2702123.2702331

[19] Uwe Gruenefeld, Jonas Auda, Florian Mathis, Stefan Schneegass, Mohamed Khamis, Jan Gugenheimer, and Sven Mayer. 2022. VRception: Rapid Prototyping of Cross-Reality Systems in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 611, 15 pages. https://doi.org/10.1145/3491102.3501821

[20] Yanning H. Han, Arun N. Kumar, Millard F. Reschke, Jeffrey T. Somers, Louis F. Dell'Osso, and R. John Leigh. 2005. Vestibular and non-vestibular contributions to eye movements that compensate for head rotations during viewing of near targets. *Experimental Brain Research* 165 (2005), 294–304.

[21] Eishi Hirasaki, Steven T Moore, Theodore Raphan, and Bernard Cohen. 1999. Effects of walking velocity on vertical head and body movements during locomotion. *Experimental Brain Research* 127 (1999), 117–130.

[22] Takao Imai, Steven T Moore, Theodore Raphan, and Bernard Cohen. 2001. Interaction of the body, head, and eyes during walking and turning. *Experimental brain research* 136 (2001), 1–18.

[23] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

[24] Sebastian Kapp, Michael Barz, Sergey Mukhametov, Daniel Sonntag, and Jochen Kuhn. 2021. ARETT: Augmented Reality Eye Tracking Toolkit for Head Mounted Displays. *Sensors* 21, 6 (2021). https://doi.org/10.3390/s21062234

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

[26] Elisa Maria Klose, Nils Adrian Mack, Jens Hegenberg, and Ludger Schmidt. 2019. Text Presentation for Augmented Reality Applications in Dual-Task Situations. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 636–644. https://doi.org/10.1109/VR.2019.8797992

[27] Maurice Lamb, Malin Brundin, Estela Perez Luque, and Erik Billing. 2022. Eye-tracking beyond peripersonal space in virtual reality: validation and best practices. *Frontiers in Virtual Reality* 3 (2022), 864653.

[28] Otto Lappi. 2015. Eye Tracking in the Wild: the Good, the Bad and the Ugly. *Journal of Eye Movement Research* 8, 5 (Oct. 2015). https://doi.org/10.16910/jemr.8.5.1

[29] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional.

[30] Magic Leap. 2020. Head-Locked Content. Website. Retrieved September 1, 2023 from https://ml1-developer.magicleap.com/en-us/learn/guides/head-locked-content-tutorial-unity

[31] Cha Lee, Scott Bonebrake, Tobias Hollerer, and Doug A. Bowman. 2009. A replication study testing the validity of AR simulation in VR for controlled experiments. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. 203–204. https://doi.org/10.1109/ISMAR.2009.5336464

[32] Cha Lee, Gustavo A. Rincon, Greg Meyer, Tobias Höllerer, and Doug A. Bowman. 2013. The Effects of Visual Realism on Search Tasks in Mixed Reality Simulation. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (2013), 547–556. https://doi.org/10.1109/TVCG.2013.41

[33] Hyunjin Lee and Woontack Woo. 2023. Exploring the Effects of Augmented Reality Notification Type and Placement in AR HMD while Walking. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 519–529. https://doi.org/10.1109/VR55154.2023.00067

[34] Yang Li, Juan Liu, Jin Huang, Yang Zhang, Xiaolan Peng, Yulong Bian, and Feng Tian. 2023. Evaluating the Effects of User Motion and Viewing Mode on Target Selection in Augmented Reality. https://doi.org/10.2139/ssrn.4514609

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

[36] Yujun Lu, BoYu Gao, Huawei Tu, Huiyue Wu, Weiqiang Xin, Hui Cui, Weiqi Luo, and Henry Been-Lirn Duh. 2022. Effects of physical walking on eyes-engaged target selection with ray-casting pointing in virtual reality. *Virtual Reality* (Aug. 2022). https://doi.org/10.1007/s10055-022-00677-9

[37] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast Gaze Typing with an Adjustable Dwell Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 357–360. https://doi.org/10.1145/1518701.1518758

[38] Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt, and Florian Alt. 2020. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376796

[39] Steve Mann and James Fung. 2001. Videoorbits on eye tap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene. In *International Symposium on Mixed Reality, 2001*. 48–55.

[40] Microsoft. 2021. Comfort - Mixed Reality. Website. Retrieved September 4, 2023 from https://web.archive.org/web/20220925015757/https://learn.microsoft.com/en-us/windows/mixed-reality/design/comfort/

[41] Steven T. Moore, Eishi Hirasaki, Theodore Raphan, and Bernard Cohen. 2001. The Human Vestibulo-Ocular Reflex during Linear Locomotion. *Annals of the New York Academy of Sciences* 942, 1 (2001), 139–147. https://doi.org/10.1111/j.1749-6632.2001.tb03741.x arXiv:https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.2001.tb03741.x

[42] Terhi Mustonen, Mikko Berg, Jyrki Kaistinen, Takashi Kawai, and Jukka Häkkinen. 2013. Visual task performance using a monocular see-through head-mounted display (HMD) while walking. *Journal of Experimental Psychology: Applied* 19 (2013), 333–344. https://doi.org/10.1037/a0034635

[43] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 15, 7 pages. https://doi.org/10.1145/3448018.3457998

[44] Judith Mwakalonge, Saidi Siuhi, and Jamario White. 2015. Distracted walking: Examining the extent to pedestrian safety problems. *Journal of Traffic and Transportation Engineering (English Edition)* 2, 5 (2015), 327–337. https://doi.org/10.1016/j.jtte.2015.08.004

[45] Masayuki Nakao, Tsutomu Terada, and Masahiko Tsukamoto. 2014. An Information Presentation Method for Head Mounted Display Considering Surrounding Environments. In *Proceedings of the 5th Augmented Human International Conference* (Kobe, Japan) *(AH '14)*. Association for Computing Machinery, New York, NY, USA, Article 47, 8 pages. https://doi.org/10.1145/2582051.2582098

[46] Anneli Olsen. 2012. The Tobii I-VT fixation filter.

[47] Joris Peereboom, Wilbert Tabone, Dimitra Dodou, and Joost de Winter. 2023. Head-locked, world-locked, or conformal diminished-reality? An examination of different AR solutions for pedestrian safety in occluded scenarios. https://www.researchgate.net/publication/371509780_Head-locked_world-locked_or_conformal_diminished-reality_An_examination_of_different_AR_solutions_for_pedestrian_safety_in_occluded_scenarios Preprint..

[48] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-Touch: Combining Gaze with Multi-Touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

[49] Ken Pfeuffer and Hans Gellersen. 2016. Gaze and Touch Interaction on Tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 301–311. https://doi.org/10.1145/2984511.2984514

[50] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[51] Ken Pfeuffer, Jan Obernolte, Felix Dietz, Ville Mäkelä, Ludwig Sidenmark, Pavel Manakhov, Minna Pakanen, and Florian Alt. 2023. PalmGazer: Unimanual Eye-hand Menus in Augmented Reality. arXiv:2306.12402 [cs.HC]

[52] Julian D. Pillay, Tracy L. Kolbe-Alexander, Karin I. Proper, Willem van Mechelen, and Estelle V. Lambert. 2014. Steps That Count: Physical Activity Recommendations, Brisk Walking, and Steps Per Minute—How Do They Relate? *Journal of Physical Activity and Health* 11, 3 (2014), 502 – 508. https://doi.org/10.1123/jpah.2012-0210

[53] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-Worn Extended Reality. *ACM Comput. Surv.* 55, 3, Article 53 (mar 2022), 39 pages. https://doi.org/10.1145/3491207

[54] Rufat Rzayev, Paweł W. Woźniak, Tilman Dingler, and Niels Henze. 2018. Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*

(Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3173574.3173619

[55] Jeff Sauro and Joseph S. Dumas. 2009. Comparison of Three One-Question, Post-Task Usability Questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1599–1608. https://doi.org/10.1145/1518701.1518946

[56] Bastian Schildbach and Enrico Rukzio. 2010. Investigating Selection and Reading Performance on a Mobile Phone While Walking. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. Association for Computing Machinery, New York, NY, USA, 93–102. https://doi.org/10.1145/1851600.1851619

[57] Dieter Schmalstieg, Anton Fuhrmann, Gerd Hesina, Zsolt Szalavári, L. Miguel Encarnação, Michael Gervautz, and Werner Purgathofer. 2002. The Studierstube Augmented Reality Project. *Presence: Teleoperators and Virtual Environments* 11, 1 (02 2002), 33–54. https://doi.org/10.1162/105474602317343640 arXiv:https://direct.mit.edu/pvar/article-pdf/11/1/33/1623621/105474602317343640.pdf

[58] Dieter Schmalstieg and Tobias Hollerer. 2016. *Augmented Reality: Principles and Practice*. Addison-Wesley Professional.

[59] William E. Schroeder. 1993. Head-mounted computer interface based on eye tracking. In *Visual Communications and Image Processing '93*, Barry G. Haskell and Hsueh-Ming Hang (Eds.), Vol. 2094. International Society for Optics and Photonics, SPIE, 1114 – 1124. https://doi.org/10.1117/12.157867

[60] Ludwig Sidenmark and Hans Gellersen. 2019. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 27, 1, Article 4 (dec 2019), 40 pages. https://doi.org/10.1145/3361218

[61] Alexandra Sipatchin, Siegfried Wahl, and Katharina Rifai. 2020. Accuracy and precision of the HTC VIVE PRO eye tracking in head-restrained and head-free conditions. *Investigative Ophthalmology & Visual Science* 61, 7 (June 2020), 5071.

[62] Samuel Stuart, Brook Galna, Sue Lord, Lynn Rochester, and Alan Godfrey. 2014. Quantifying saccades while walking: Validity of a novel velocity-based algorithm for mobile eye tracking. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 5739–5742. https://doi.org/10.1109/EMBC.2014.6944931

[63] Ivan E. Sutherland. 1968. A Head-Mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I* (San Francisco, California) *(AFIPS '68 (Fall, part I))*. Association for Computing Machinery, New York, NY, USA, 757–764. https://doi.org/10.1145/1476589.1476686

[64] Leah L Thompson, Frederick P Rivara, Rajiv C Ayyagari, and Beth E Ebel. 2013. Impact of social and technological distraction on pedestrian crossing behaviour: an observational study. *Injury Prevention* 19, 4 (2013), 232–237. https://doi.org/10.1136/injuryprev-2012-040601 arXiv:https://injuryprevention.bmj.com/content/19/4/232.full.pdf

[65] Matthew A. Timmis, Herre Bijl, Kieran Turner, Itay Basevitch, Matthew J.D. Taylor, and Kjell N. van Paridon. 2017. The impact of mobile phone use on where we look and how we walk when negotiating floor based obstacles. *PLoS one* 12, 6 (2017), e0179802.

[66] Takumi Toyama, Daniel Sonntag, Jason Orlosky, and Kiyoshi Kiyokawa. 2015. Attention Engagement and Cognitive State Analysis for Augmented Reality Text Display Functions. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, Georgia, USA) *(IUI '15)*. Association for Computing Machinery, New York, NY, USA, 322–332. https://doi.org/10.1145/2678025.2701384

[67] Kristin Vadas, Kenton Michael Lyons, Daniel Ashbrook, Ji Soo Yi, Thad Starner, and Julie A. Jacko. 2006. Reading on the Go: An Evaluation of Three Mobile Display Technologies. http://hdl.handle.net/1853/13112

[68] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Zurich, Switzerland) *(UbiComp '13)*. Association for Computing Machinery, New York, NY, USA, 439–448. https://doi.org/10.1145/2493432.2493477

[69] Uta Wagner, Mathias N. Lystbæk, Pavel Manakhov, Jens Emil Sloth Grønbæk, Ken Pfeuffer, and Hans Gellersen. 2023. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 252, 15 pages. https://doi.org/10.1145/3544548.3581423

[70] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[71] Lijie Yao, Anastasia Bezerianos, Romain Vuillemot, and Petra Isenberg. 2022. Visualization in Motion: A Research Agenda and Two Evaluations. *IEEE Transactions on Visualization and Computer Graphics* 28, 10 (2022), 3546–3562. https://doi.org/10.1109/TVCG.2022.3184993

[72] A Yefremenko, L Shesterova, S Lebediev, I Zakharina, A Apaichev, Y Krajnik, T Samolenko, and S Pyatisotskaya. 2019. Evaluation of characteristics of running with an audio stimulation in prepared students from various sports. *Journal of Physical Education and Sport* 19, 1 (2019), 696–702.

[73] Qiushi Zhou, Difeng Yu, Martin N Reinoso, Joshua Newn, Jorge Goncalves, and Eduardo Velloso. 2020. Eyes-free Target Acquisition During Walking in Immersive Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3423–3433. https://doi.org/10.1109/TVCG.2020.3023570

A IMPLEMENTATION DETAILS OF THE HEADDELAY REFERENCE FRAME

HeadDelay introduces a delay between the head and the targets' movement. While the end position and rotation of the targets were the same as with Head, to smooth their movement, we linearly interpolated their position and spherically interpolated rotation with the constant speed of 5 (the variables *_positionLerpSpeed* and *_rotationLerpSpeed* below; was selected empirically to resemble a moderately quick inertia effect) between consecutive frames:

```
// Simulate inertia by interpolating position & rotation
// from the current state to the desired state
position = Vector3.LerpUnclamped(transform.position,
    position, Time.deltaTime * _positionLerpSpeed);
rotation = Quaternion.SlerpUnclamped(transform.rotation,
   head.rotation, Time.deltaTime * _rotationLerpSpeed);
```

It is important to note that we suppressed the movement along the Z axis, i.e. in depth so that the participant's body acceleration and deceleration during locomotion would not bring the targets closer or further from them, and thus, keeping the targets equidistant.