HAND INTERACTION ON THE GO: COMPARING REFERENCE FRAMES IN AUGMENTED REALITY

THESIS

Master Student
Tamás Csizmadia
Aarhus University
au724747@uni.au.dk

Master Student
Mikkel Roos
Aarhus University
201804414@post.au.dk

Main Supervisor
Ken Pfeuffer
Aarhus University
ken@cs.au.dk

Co-Supervisor
Pavel Manakhov
Aarhus University
pmanakhov@cs.au.dk

Aarhus, 14th of June 2024

ABSTRACT

With the increasing accessibility and mobility of head-mounted displays (HMDs), their potential use cases extend beyond stationary applications. This thesis investigates hand interaction in augmented reality (AR) and examines how different reference frames affect input performance while walking. Specifically, we compare three reference frames — *Palm*, *PalmWithoutRotation*, and *Path* — each offering varying degrees of user control. Using a Fitts' law study (ISO 9241-9), we measured the performance of target selection in three movement conditions: standing, linear walking, and circular walking.

Our findings indicate that the choice of reference frame significantly impacts hand input performance during locomotion. The reference frame with the highest level of user control, *Palm*, demonstrated superior performance in terms of speed and throughput while on the go, and the differences get larger as the movement type gets more complex. On the other hand, the results are inconclusive and optimized versions of *Path* could catch up. Based on user preferences, *Palm* still performed better, but we conclude that providing multiple reference frame options as an interface designer might be the optimal route.

Our work contributes to the understanding of AR reference frames and suggests that hand-controlled interfaces might increase target acquisition, but that further research on placing the *Path* reference frame is needed.

Contents

1 Introduction **4**

2 Related work **5**
2.1 Augmented Reality in locomotion . . . 6
2.2 Spatial reference frames . . . 6
2.3 Virtual hand input . . . 6

3 Design of the Reference Frames **7**
3.1 *Palm* . . . 7
3.2 *PalmWORotation* . . . 8
3.3 *Path* . . . 8
3.4 *Chest* – Unused . . . 9

4 Study **10**
4.1 Study Rationale . . . 10
4.2 Evaluation metrics . . . 11
4.2.1 Questionaires . . . 11
4.3 Independent variables . . . 12
4.3.1 Reference frames . . . 12
4.3.2 Movement type . . . 12
4.3.3 Target diameter . . . 13
4.3.4 Balanced Latin Square . . . 13
4.4 Dependent variables . . . 13
4.4.1 Movement time (MT) . . . 13
4.4.2 Success . . . 13
4.4.3 Accuracy (SD_x) . . . 13
4.4.4 Effective amplitude (A_e) . . . 13
4.4.5 Effective width (W_e) . . . 14
4.4.6 Effective Index of Difficulty ID_e . . . 14
4.4.7 Throughput (TP) . . . 14
4.5 System . . . 14
4.5.1 Apparatus . . . 15
4.5.2 Server application . . . 15
4.5.3 Client application . . . 16
4.5.4 Passthrough / VR . . . 16
4.6 Study procedure . . . 17
4.7 Data collection . . . 18
4.7.1 Logging . . . 18
4.8 Participants . . . 18

5 Results **18**
5.1 RQ1: How does movement affect user performance of target selection? . . . 19
5.2 RQ2: How do the reference frames affect user performance? . . . 20
5.3 RQ3: Does having more control over the position of the interface lead to more or less speed and accuracy? . . . 21

5.4 RQ4: Is there a reference frame that is suited more for one movement type and another for a different movement? . . . 21
5.5 RQ5: What is the user experience with the reference frames? . . . 22
5.6 User feedback . . . 23
5.6.1 Physical demand . . . 23
5.6.2 Circle track . . . 24
5.6.3 Path reference frame . . . 25
5.7 Supplementary Metrics . . . 25
5.7.1 Depth . . . 25
5.7.2 Decline . . . 26
5.7.3 Relative target yaw . . . 26

6 Discussion **27**
6.1 Limitations and future work . . . 28

7 Conclusion **29**

8 Acknowledgements **29**

Appendices **33**

A Significance Tables **33**
A.1 Movement time ($Movement : ReferenceFame$) . . . 33
A.2 Success ($Movement : TargetSize$) . . . 34

B Study procedure **35**

C Logging **36**
C.1 Selections logger . . . 37
C.2 High-frequency logger . . . 37

D Questionnaire **39**

E Source Code **39**

1 Introduction

Augmented Reality (AR) and more specifically Head-Mounted Displays (HMDs) have enabled us to interact with digital objects in more natural ways. Augmented Reality combines the real and virtual world in real-time, with interactable 3D elements and UI's. HMDs let people get immersed in this world by strapping onto the user's head. As these virtual objects are more and more life-like it is most intuitive to interact with them using our hands. The use of hand gestures as an input modality in AR has been a topic of interest for researchers and developers alike. Clicking on a button with our fingers is more similar to how we interact with physical objects like a smartphone, or a keyboard. Newer HMDs have hand tracking built-in that is improving with every update. This makes studying how different interfaces affect the target acquisition speed and accuracy much easier. AR hand input has been extensively studied in the context of interaction during sitting or standing [7] and others studied the accuracy of target acquisition with hands during walking [3, 35]. We have less knowledge of AR target acquisition on interfaces in front of a user in locomotion.

Walking during interaction with our everyday devices is a common occurrence. We walk while using our smartphones and smartwatches, and a common vision for the future is that we will walk while using AR glasses. As with the personal computer, the technical advancements of AR HMDs, like the usage of inside-out tracking, imply greater mobility and a trajectory towards being a more common tool in our everyday lives. Additionally, tracking the user's hands has not been feasible while walking, as the algorithm assumed a stationary camera [7], or it required cumbersome equipment, like a glove [5]. With the advent of hand tracking in AR glasses, and more specifically their continuing improvement, we should be able to send emails or book a flight ticket while walking. One big problem remains though, most of the interfaces AR devices use are stationary. This means that the user has to stop walking to interact with the interface. This is not ideal, as it breaks the flow of walking.

When the user is walking or running instead of sitting or standing, the reference frames we can choose from are limited. One way to provide a stable reference frame is to attach the interface to the user's body. This, though, is not clear-cut. The body has many moving parts, and it is not clear which part of the body is the best reference frame for interaction. Consequently, and despite the current state and direction of AR in society, a knowledge gap remains in understanding the impact of reference frames on hand input during locomotion. This affects both AR researchers and industry professionals as it is not clear what to expect when studying or developing on-the-go applications, that rely on hand interaction. This motivates our study, where we investigate the effect of different reference frames on target selection during walking in AR. The reference frames we compare are *Palm*, *PalmWORotation* and *Path*. These represent a continuum of degrees of control the user has over the interface. We chose these reference frames as two polar opposites and one in-between.

Palm is the most direct way to control the interface. It appears a few centimetres to the side of the palm of the non-dominant hand. The user can move the interface by moving their hand, the interface will remain in the same position relative to the palm. This is as stable as a mobile phone in our hands. However, we do not know, whether providing less control, but more stability would increase or decrease performance. The adaptation of *Path* [25] we implemented moves with the head in the direction of the path, and keeps its rotation locked to be perpendicular to the linear track. We adapted it to work in circular tracks, where its rotation is turned more towards the user than if it was perpendicular to the path. The user's control here is virtually nonexistent, except when the user is moving forward, it also moves. On the flip side, this reference frame is the most stable. *PalmWORotation* is similar to *Palm*, but the interface does not rotate with the hand, it is rotated exactly like *Path*. This restricts the user, trading control for stability. It can shed light on whether positional control or rotational control affects performance more.

Previous research concluded that walking and turning had a negative effect on the user's performance in a target selection task [35]. To this end, we conducted our experiment with 3 path types: *Standing*, *Linear* and *Circular*. The *Standing* path is a control condition where the user is standing still. The *Linear* path is a straight line where the user walks from one end to the other. The *Circular* path is a circular track where the user walks around a circle.

The main research question of this study is: *How do different reference frames affect target selection during walking in AR?* To answer this question, we conduct a study where participants walk on a predefined path and select targets with their dominant hand. We measure the time it takes to acquire the target and the accuracy of the selection. To measure target acquisition speed and accuracy, we use a 2D Fitts' law task [11, 30].

The key findings from our study are, that the choice of reference frame does make a significant difference in hand input performance while walking. *Palm* had significantly better performance in speed and throughput. We can also see that the difference between *Palm* and *Path* gets larger as the movement type gets harder. This suggests *Palm* is better on the go. We also found that *PalmWORotation* is the worst out of the three as it is outperformed by *Palm*, but it requires as much physical effort. Based on these results, we can say that control does influence target acquisition, but the results are inconclusive as better versions of *Path* could close the performance gap. Even with the current reference frames, the user experience is mixed and providing multiple reference frames might be the best way to interact in the future.

Generally, this work deepens the understanding of AR reference frames and provides fundamental knowledge that can serve as the basis for UI design decisions for AR applications on the go.

The specific contributions of this thesis are:

- Empirical data on target acquisition while on the go, testing three reference frames that observe different levels of user control.
- User feedback on three reference frames that try to enable user interaction on the go.
- Adaptation of the *Path* reference frame [25] with a novel method of placing it on a circular track.

2 Related work

Walking is one of the most fundamental activities of humankind and has been the subject of many studies [20, 18]. Lucero et al. [22] experimented with displaying notifications on a simple UI in interactive glasses while walking in public. Since the UI used is purely display-fixed, it falls outside the commonly accepted definition of an augmented reality system, where objects should be registered in 3D [2]. Relevant to this work is their finding that receiving and performing simple interactions with virtual notifications did not distract the users from the simultaneous task of walking. Regarding more complex interactions, research on mobile phone use shows that walking and turning have a negative effect on the user's performance in a target selection task while using a mobile phone [29, 3]. This is because the user has to divide their attention between the physical world and the virtual world. The user has to pay attention to the path they are walking on and the target they are selecting. This can lead to slower and less accurate target selection. This is known as the dual-task paradigm [28], and is useful even in Augmented Reality, where the phenomenon seems to still hold. Kane et al. [12] explored hand interactions on mobile device UI's, that dynamically increased the button sizes when walking to reduce accuracy loss. Research like this might also be relevant in HMD user interfaces but requires more fundamental knowledge of how these UIs perform with hand interaction.

2.1 Augmented Reality in locomotion

Walking has been the focus of many recent Augmented Reality papers, suggesting a growing trend towards walking applications. Müller et al. investigated walking as an input modality by letting the user select different options depending on surface-drawn virtual tracks that the user has to walk on [27]. Lee et al. focused on navigating in a virtual environment with little space available, keeping the movement of the body similar [17]. Chan et al. explored gait analysis helping rehabilitation training [6]. Klose et al. [14] focused on reading performance while walking and the placement of text. Khamis et al. [13] investigated eye pursuit interaction. The aforementioned works share a common focus of studying different aspects of AR interaction when walking, but none of them addresses the relation between hand-based target acquisition and interface stability. Relevant to our work is that some of these studies found target acquisition while walking to negatively affect accuracy, which generalizes the results of the earlier mobile phone research. They also shed light on the problem of interface placement. [35] focused on target acquisition during walking in the peripersonal space. They found that chest-referenced targets showed better performance than head-referenced. Since they only focused on an eyes-free target acquisition task, the question remains if the same results can be seen in within-FOV target selection tasks.

2.2 Spatial reference frames

Reference frames are the different ways the UI is shown to the user and how it translates with the user or objects. The most commonly accepted set of reference frames are *head*, *world*, *body* and *object* [15]. As technology improves we are able to anchor to more complex objects and surfaces without cumbersome hardware. The first Head Mounted Display that could track virtual objects positioned in the *world* was made by Sutherland [31]. Since then we have been able to track *objects* as anchors [26] and even *body* parts [7]. The *head* reference frame is when the interface does not move within the display of the HMD, it moves with the head [9]. It has been shown that this type of reference frame is eye-straining [4]. Reference frames are mobile if they can be used during locomotion. The *egocentric* reference frames [8] such as the *body* and *head* generally fit this criteria. The *device* reference frame [15] is also mobile, where the tracked object is a handheld device. Examples of papers focusing on reference frames during movement include Zhou et al. comparing *head* and *torso* in the peripersonal space [35], Lu et al. [19] compared the same two in information access tasks, Li et al. [18] focused on comparing frames with walking on a treadmill, also comparing ray-casting with virtual hands. While *head* and *torso* reference frames have been compared time and time again, *hand* has not. Finally, one study found that handheld devices (*hand*) were easier to read than HMDs (*head*) [32].

2.3 Virtual hand input

We have seen many studies on hand input, underscoring its importance in the field of AR. There are two classes of techniques that enable a user to interact in the 3D space, using virtual hands and ray-casting. [1] In this study, we are using virtual hands as it has been shown to be more intuitive [18]. Providing a way to measure input performance is crucial to comparing study conditions. The most common method is using Fitts' law, the standard ISO 9241-9 [23, 30]. Many studies in Augmented Reality used this metric to assess user performance [33, 18]. Weiss et al. [34] used this metric to directly compare the performance difference between touchscreen (*hand* reference) and AR (*world* reference) in a simple target acquisition task. They found the throughput of a touchscreen to outperform AR in all their measures. This motivates our research to investigate whether positioning an interface in AR, using the *hand* has reference frame, shows similar results.

3 Design of the Reference Frames

In this study, we are using a combination of *body* and *world* reference frames. The following reference frames cover the space between world-controlled and user-controlled UI placement. Our hands are suited to stabilize objects during locomotion. We are used to reading from screens while walking. Multiple studies found that using mobile phones to read or interact resulted in better performance compared to world-referenced AR interfaces [32, 34]. We inform our reference frame choice by this phenomenon, the stabilizing effect of hands which could improve performance over actually stable-relative-to-the-world reference frames.

We considered expanding on this by exploring a design space defined by two axes: One measuring the extent to which a reference frame was user-referenced versus world-referenced, and another assessing the degree of user control. This would also enable us to compare the hand stabilization effect to other body-referenced anchors. We ended up not using it because the design space would be half-baked without more reference frames and excluding it enabled us to focus on movement types.

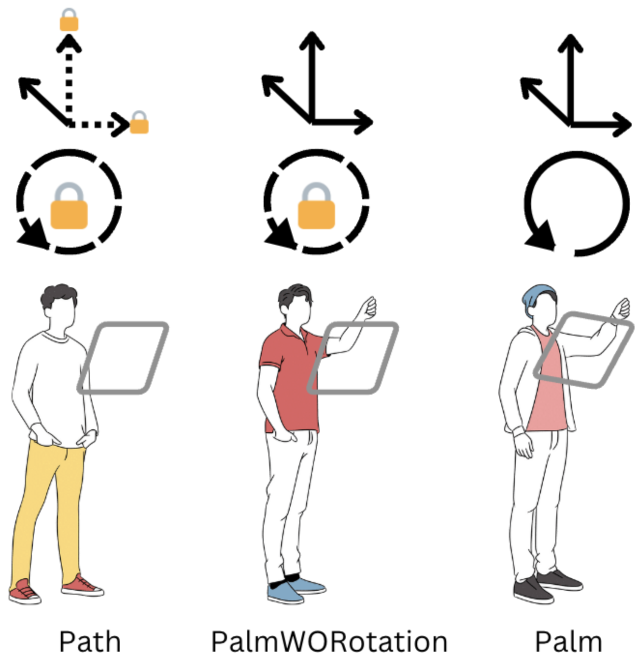


Figure 1: The three reference frames we compare, ranging from only controlling one axis of movement to rotation and position being fully controlled by the user's palm.

3.1 *Palm*

This reference frame is anchored to the centre of the user's palm on their non-dominant hand. It is offset to the side by a few centimetres. When the user's palm faces them, the interface is aligned to face the user. When the user rotates their palm, the interface moves with it. I.e. as seen in Figure 1, no axes are world referenced. In general, the reference frame is positioned within the palm's own geometry. The palm anchor was chosen to offer maximal control and exploit the biomechanical hand stabilization advantages. This is most comparable with using smartphones with two hands, hence participants might feel most used to this reference frame.

The hand tracking of the newest HMDs is improving, and it enables us to do more and more complex interaction types in AR. However, tracking while walking still leaves much to be desired. If we were to track both hands optically from the headset, the tracking errors would be more prevalent than with *Path*, as the interface here is positioned relative to the non-dominant hand. To counter this, we developed a custom hand mount (Figure 6), in which the controller fits snugly. The controllers we use, the Quest Pro controllers, have their own tracking, and they track well even when out of sight. We use this solution so that the participant can still interact with the UI with an open palm facing the user, as that is the standard way to interact.

During development, we discussed whether it would make more sense to put the interface on top of the palm. It would provide haptic feedback, that in theory could make interaction faster and more accurate. We decided otherwise for two reasons, one is that the other two reference frames do not have haptic feedback and it would be harder to compare how much the haptic feedback and the reference frame itself contributed. Secondly, the dominant-hand tracking gets worse if it has objects under it.

3.2 *PalmWORotation*

This reference frame is still anchored to the centre of the user's palm, but now it is always rotated perpendicularly to the track. If the user moves their hand, the interface still moves, but it does not rotate. This provides some world-referenced stability to the user, so when they are walking, small unavoidable hand jitter does not greatly affect the positioning of the interface. Hand stability is less of a factor here, so the comparison between this and *Palm* might illuminate the difference between hand stabilisation and the lack of it.

The implementation of this is very similar to the *Palm* reference frame. The only difference is that it also uses the track's position to calculate its rotation (like *Path*).

3.3 *Path*

Our *Path* [25] uses *world* position except on the axis the participant is moving on. It is anchored to the headset with an offset parallel to the track. So if the participant swerves off the track, the interface stays on top of the track. *Path* mimics the reference frame used by Lu et al. because, while it was fully *world* referenced, the participants were walking on a treadmill [21]. This reference frame is the most world-stable, it does not shake with the user's hands, it does not swerve off, and it is reliably in the expected spot where the user feels it's most comfortable without the instability of walking. On the other hand, the user has no control over its positioning during use. In circular tracks, the reference frame positions itself in the middle of the circle track 30 cm in front of the user. It is rotated so that it looks at the middle of the track under the user, hence stepping sideways does not affect it, but stepping forward moves it along the track.

Implementing this frame was the simplest as it needs the head position for only one axis and derives the other axes and rotation from the track. However, introducing the circular path greatly increased the complexity. We came up with 3 different design options (See Figure 2), all starting from the adjusted position of the user, which is the closest point in the middle of the track:

- One where we calculate an offset forward on a tangent drawn from under the user. The position of the interface is on the closest point from this to the middle track. The interface is rotated perpendicularly to the track. This one turned out to rotate away from the participant too much. and it also rotates in *PalmWORotation* when moving the palm back and forth, even if the user is not moving.

- One where we draw a tangent under the user and the reference frame is located on top of that tangent, perpendicular to the tangent. We discarded this because humans tend to turn into their path. This means that the head always looks slightly into the circle rather than strictly forward. Users could have perceived it as the interface uncomfortably outside of their turning circle.
- We ended up with the current one because it is a good way of balancing the above two. We offset the user position 30 cm forward on a tangent drawn from the user, then we find its projection on the track. The position is then calculated by drawing a line between the adjusted user position and the offset position and adding another offset. This way in *PalmWORotation*, the reference frame does not rotate when standing and moving the palm back and forth. This might be explained better in Figure 2.

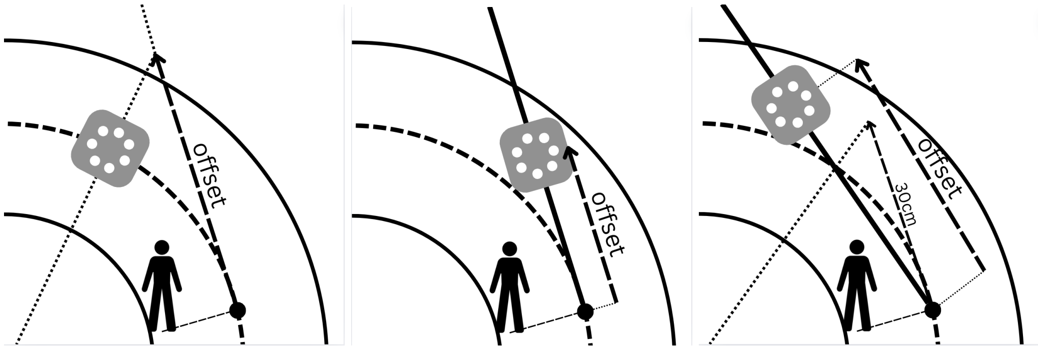


Figure 2: Path positioning considered: The leftmost uses the desired position of the reference frame and just puts it on the closest point on the track, perpendicularly to it. The middle uses a tangent drawn on the track from the user. The rightmost uses a more balanced combination of the two that tries to be right in front of the user. We used the left one in the study.

This and *PalmWORotation* both are rotated 40 degrees up so that they face the user even when positioned at a lower point. To avoid problems with user height and arm length differences, the participant has an extra step before using this reference frame, where they have to set a comfortable height and distance from themselves. The position set here will offset the default position with the frame's height and distance from the user, but it will remain in the same vertical plane.

We did consider keeping the reference frame non-customisable, either with a fixed offset, or scaled to the height of the participant, under the assumption that this would increase the internal validity of the study. However, since the desired position of the targets is very subjective, as verified by our results, the interval validity is arguably higher when adding customizability.

3.4 *Chest* – Unused

The *Chest* reference frame was considered, since it has been explored before by [36] on circular tracks, where it showed promising results. It was motivated by an alternative design space where *Palm* and *Chest* are fully referenced to the user, whereas *PalmWORotation* and *Path* restrict some degrees of freedom to be world referenced. The biomechanical mobility of the hand and palm significantly improves user control of the UI, placing it opposite to the more restrictive Chest and Palm reference frames on the second axis of this particular design space. As described in the beginning of this section, we decided to exclude it from our study. The reasons for excluding the Chest reference frame include:

- A simplified design space, now focusing on the Palm, PalmWORotation, and Path frames, effectively spans a continuum from fully user-referenced (Palm) to nearly fully world-referenced (Path). This reduction in independent variables allows for a clearer focus on the essential properties that influence hand interaction.

- The introduction of a curved track was prioritized, as it more accurately simulates real-world locomotion in AR environments, thereby enhancing the study's practical application to everyday AR use.

4 Study

In this study, we examine the target acquisition performance of individuals who had time to train with the specified tasks, comparing the reference frames. We address the following questions:

- How does movement affect user performance of target selection?
- How do the reference frames affect user performance?
- Does having more control over the position of the interface lead to more or less speed and accuracy?
- Is there a reference frame that is more suited for one movement type and another for a different movement?
- What is the user experience with the reference frames?

4.1 Study Rationale

Basing the answer to the above questions on empirical research requires rigorous data-gathering methods to ensure the validity of the results. We are looking to compare three reference frames and gain insight into how the hand input performance compares. The most externally valid way of answering the questions would be using AR glasses – as we expect AR glasses are going to be the outcome of AR research – and using real-world applications to compare our reference frames. AR glasses have not reached the level of maturity to be able to run more complex applications, so we had to use contemporary HMDs. The best hand tracking available to us was the Quest 3, even though it required using virtual reality with video passthrough instead of see-through AR. Additionally, using complex applications to compare reference frames would have been less quantitative, as time measurements in studies like that are less reliable as they include thinking time and other factors. To measure significant differences between the reference frames we would have had to study a greater amount of participants. For this reason, using Fitts' law to measure user performance is an easy way to get quantitative data for comparison. It is the standard way to evaluate input devices, hence there is a great amount of research behind it and it provides a few robust numbers that make comparison easy. Regarding user experience, we are interested in two things: How difficult the users found the task (perceived workload), and subjective preference. NASA-TLX and questionnaires are standard methods for answering this and will be covered in more detail below.

We employed a factorial within-subject experiment design with 3x3x4 conditions, where the independent variables with their corresponding ranges are:

- *Movement type*:{*Standing*, *Linear*, *Circle*}
- *Reference frame*:{*Palm*, *PalmWORotation*, *Path*}
- *Target Diameter*:{2cm, 3cm, 4cm, 5cm}

4.2 Evaluation metrics

Following in the footsteps of other research [25, 21], we use Fitts' law and the ISO 9241-9 standard [30] to measure user performance between conditions. Fitts' law is a well-known method for quantifying performance, measured in a unit called throughput. This single measure of performance allows for easier comparisons between different reference frames. It depends on another measure called the Index of Difficulty, which quantifies the difficulty of a given selection task. We calculate the Index of Difficulty as $log_2(A/W+1)$ [30, 24] where D is the distance between buttons and W is the button width. We choose conditions so that the Index of Difficulty covers a wide range of values by varying the target width, as explained in Sub Section 4.3.3. Throughput also requires us to record movement time data and the actual position of the pointing device (finger).

The 2D Fitts Law task from the ISO 9241-9 standard uses a circular arrangement of targets as shown in Figure 3. The starting position is fixed and the selection order of the targets is always the same. The targets are positioned at the circumference of a circle with a diameter of 15cm. While this circle diameter stayed fixed throughout the study, the target diameter fell in the range of $\{2cm, 3cm, 4cm, 5cm, \}$.

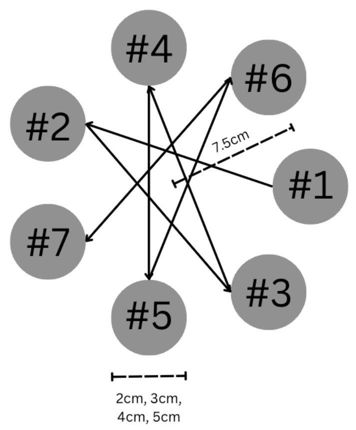


Figure 3: 2D Fitts law targets

4.2.1 Questionaires

We administered two questionnaires during the study, first the demographic questionnaire, where we just collected basic information about the participants. For the other questionnaire we used the standard NASA-TLX [10] form to measure the perceived workload of the participants, for each reference frame. The questionnaire aims to assess this workload by measuring the subjective perception of the task on 6 different scales: Mental demand, physical demand, temporal demand, performance, effort, and frustration. Each subscale is ranked from 1 to 7, resulting in ordinal data. See the appendix for the full questionnaires.

User feedback was also measured in the final questionnaire and included a subjective ranking of the three reference frames, as well as an overall impressions field, to gather general user feedback and comments to contextualize the chosen reference frame ranking.

4.3 Independent variables

In this section, we will list the independent variables of the study, i.e. the factors we systematically manipulate to observe their effect on the dependent variables.

4.3.1 Reference frames

We use *Palm*, *PalmWORotation* and *Path* reference frames in this study. The full descriptions of these are found in Section 3.

4.3.2 Movement type

As mentioned, our study attempts to assess how spatial reference frames affect hand input performance during locomotion. An ideal method for investigating this in a realistic setting would be to let study participants walk freely as pedestrians on authentic routes and measure the results. Because this would be a great source of extraneous variables, as well as no suitable AR HMD currently recommends outside use near roads or streets, we simulated this in a controlled environment.

We used linear and circular virtual tracks to simulate walking straight and turning while walking. So we let the independent variable *Movement* range over 3 values: *Standing*, *Linear*, *Circular*.

The linear track is visualized by two track borders (5cm wide, 1cm tall), which are positioned parallel with a 65cm gap between them (Figure 4). The track length is 7 meters and was chosen as we concluded that we did not need more to be able to finish one trial in time. Figure 8 shows what it looks like to the participant. The circular track is visualized by the same borders, arranged as concentric circles, such that the centre of the path (between the two borders) forms a circle with a radius of 1.33 meters (See Figure 4). A starting position was created for the participant by removing border segments defined by a 70 cm circle chord on the inferred circle that is the centre of the path. This circular track length was chosen to be of similar path length as the linear track, but slightly longer to account for the potential shorter actual path participants might take when walking closer to the inner border. We acknowledge that the radius of the circle is more important in this measurement, but we concluded that basing the radius on the length is also valid, and the turning seemed great enough for larger differences between it and the linear path.

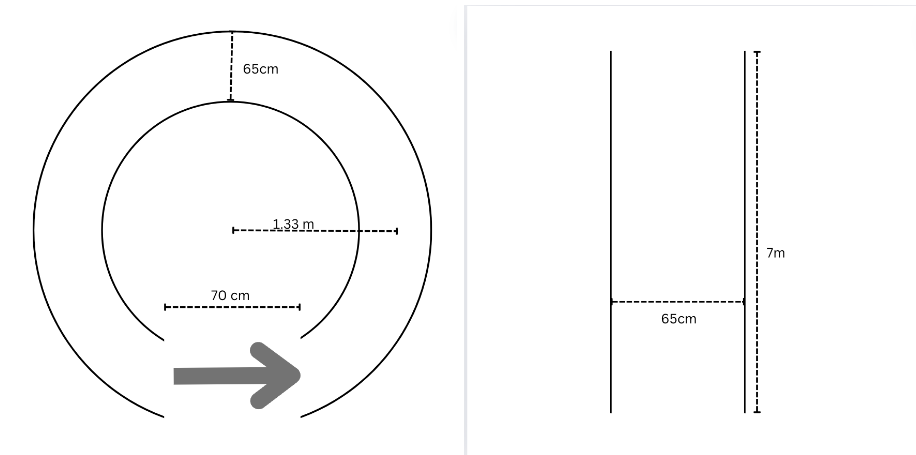


Figure 4: Circle track and straight track

In *Linear* and *Circular*, during training and measurement we play a metronome sound to help participants step at a constant $100bpm$. This is to help the internal validity of the study. While we cannot restrict walking speed, as participants have differing natural step lengths, we have to make sure they do not slow down at a harder task or speed up at an easier one.

We considered linear track *Jogging* instead of *Circular* in the beginning, but after realising that the headset would be too heavy to jog with and that it would move up and down even after a tighter fit, we decided to replace it.

4.3.3 Target diameter

In accordance with Fitts' law and the ISO 9241-9 standard [30], we use 4 different target sizes to make sure the Index of Difficulty covers a wide range of values. We chose 2cm ($ID = 3.06$), 3cm ($ID = 2.55$), 4cm ($ID = 2.22$), 5cm ($ID = 1.97$). Note, that the ID is still pretty low for even the highest ID, that is because we found that the task is inherently very difficult. Making it harder would have resulted in no accuracy.

4.3.4 Balanced Latin Square

For user studies in general, it is well known that the order of conditions, if neglected, might introduce undesirable confounding effects in the measurements. The learning effect is one such extraneous variable, which describes the increase in user performance when repeating the same task. This effect is more frequently seen in within-subject studies [16], like ours. To avoid this, we applied a balanced Latin square design to counterbalance the order in which each participant encountered the 3 reference frames. The carry-over effect, which the balanced Latin square aims to avoid, is still present in a 3x3 Latin square when the number of variables is odd. We therefore needed a balanced square of size 6 in our case, which also meant that the number of participants in our study had to be divisible by 6.

4.4 Dependent variables

We will now define the dependent variables we are interested in measuring. We use throughput as a dependent variable to measure input performance. Computing throughput requires multiple measurements, most of which we decided to keep as dependent variables.

4.4.1 Movement time (MT)

This is a measure of the time it takes from pressing one button to the next. We exclude this measurement for the first button press in each trial since it does not have a previous button to measure from.

4.4.2 Success

The fraction of successful button presses. If the user's index finger crosses the 2D plane of the button, a press is logged. It is successful if the position of the index finger during crossing is within the target's boundaries and unsuccessful if outside.

4.4.3 Accuracy (SD_x)

The standard deviation of dx values for each trial. The dx is the projection of the previous button centre to the current hit vector on the center-to-center vector minus the center-to-center vector. [24] This is clearer in Figure 5. It measures how close the hit was to the centre of the button on one axis.

4.4.4 Effective amplitude (A_e)

Effective amplitude covers the distance from the previous target centre to the current target selection point projected on the line between the two target centres. In Figure 5 it is $a + dx$.

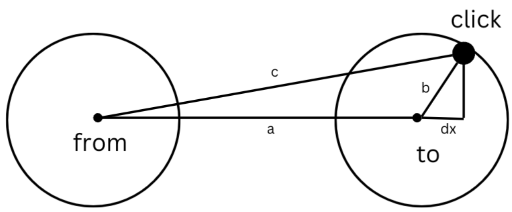


Figure 5: We calculate these variables for every button press. The two circles are the targets, the left one is the previous press, and the right one is the current. a is the distance between the two targets, b is the distance of the press from the centre of the target, more detail in Section 4.4.3 Graphic inspired by [24].

Using just the amplitude, from target centre to target centre, will not account for the scenarios where a participant might hit a button on the closest edge, which will effectively reduce the actual difficulty of the performed selection. This effect will of course be more prominent on larger target diameters, and thus using amplitude could result in an artificially high ID and thus TP [24].

4.4.5 Effective width (W_e)

Effective width indicates the width of the actual selection points in one condition. As mentioned in [24], using this instead of the target width will usually reduce the variation in the resulting TP. For example, hitting all selections in a condition close to the centre of the targets effectively increases the task difficulty, and using effective width will account for this, resulting in a higher and more appropriate ID. We calculate it with $4.133 \cdot SD_x$ [24].

4.4.6 Effective Index of Difficulty ID_e

Effective Index of Difficulty is simply ID but computed using the effective values W_e and A_e described above. We use the so-called Shannon formulation [24].

$$ID_e = \log_2 \left(\frac{A_e}{W_e} + 1 \right)$$

ID_e (and ID) is therefore a way to quantify the difficulty of a selection task in the unit of bits.

4.4.7 Throughput (TP)

Throughput according to the Fitts' law standard [30, 24] is:

$$\log_2 \left(\frac{A_e}{4.133 \cdot SD_x} \right) / MT$$

This can be interpreted as bits-per-second. It combines the movement time and accuracy so this should reflect overall performance the most.

4.5 System

Our system is based on Artem Rozhenkov's Unity prototype for a similar study. Our codebase is a fork of his GitHub repository. We changed a lot of the code to be able to update Unity and the Meta Quest SDK, but the underlying architecture remained the same.

4.5.1 Apparatus

For the study, we used a Meta Quest 3 HDM on build version v65, equipped with a Meta Quest 3 Elite strap. Regarding this choice, this was the setup that provided us the best hand-tracking accuracy, while also providing the possibility to conduct an AR study. It was considered to conduct the study in a fully occluded environment (See Section 4.5.4), but the main goal was always an externally valid AR experiment. High-quality passthrough and hand tracking were crucial, but resource availability also played a role and the Quest 3s provided us with the best balance. The controllers we used were the Meta Quest Touch Pro controllers, selected specifically for their self-tracking capabilities. This feature was essential because the mounting point on the back of the hand, as described below, would obstruct the line of sight required by the standard Quest 3 controllers. Initial hardware testing indicated that these controllers had significantly lower latency compared to the Quest 3 hand tracking. While we did not find official latency figures or research that compared this, our observations of this were consistent. This motivated the need for the hand controller mount that would simulate lower hand tracking latency, while not having to hold the controller.

Figure 6: Custom-made plastic controller mount. Shown on both left and right hand

The left controller was secured in place using a custom-fabricated moulded plastic mount, which positioned the controller on the back side of the hand. The controller mount was constructed to be usable by both left and right-handed participants. Additionally, the controller was positioned such that the self-tracking cameras on the controller were unobstructed, and pointing away from the user in most positions.

We did consider conducting a more complete preliminary HMD comparison study between the Meta Quest 2 and Meta Quest 3, with the specific purpose of identifying the one with the best hand tracking. However, we found through initial experimentation that there were no perceivable differences, so we decided to go with the Quest 3 for the improved passthrough capabilities.

4.5.2 Server application

We administered the study from a server that would connect to the client side on Quest 3, which showed the conditions the participant had to go through. This remote access allowed us to control the study without having to remove the headset, which could influence the performance of the participant. The responsibility of the server was purely to administer the conditions, reset the track positions, and set initial info like participant ID and left- and right-handedness. All other computations would be done locally in the client application running on the Quest 3.

Both the server and client applications were implemented in Unity, version: 2022.3.19f1. The communication between the server and client applications was purely message-based and used the Unity $NetworkDriver$ API from the $Unity.Networking.Transport$ package.

A sequence diagram can be seen in Figure 7 showing a snippet of the network communication that happens between server and Client during one condition. Notice that the logging only happens

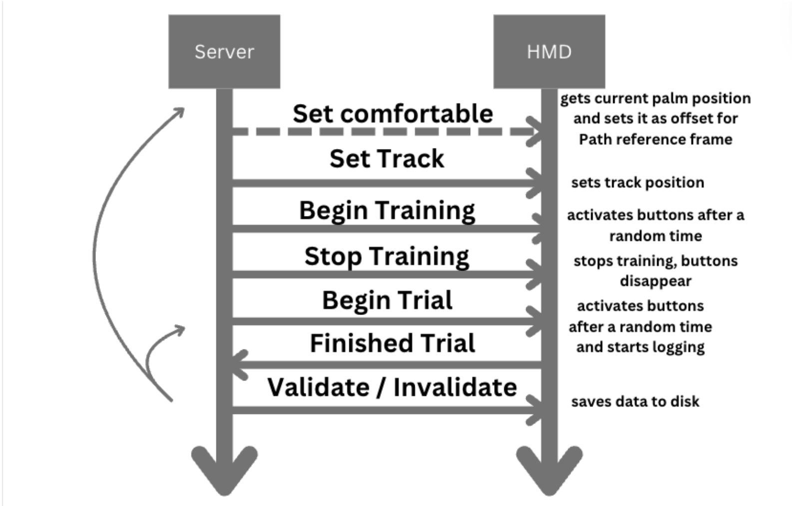


Figure 7: Sequence diagram showing message passing between server and client

during the selection period, and no messages are being explicitly sent during that time. However, if an error occurs during the selection phase, a message is sent to the server, that effectively invalidates the running trial immediately, and a new trial begins automatically (either after a random time period or after leaving and re-entering the track from the beginning). If no error occurred during selection, validation is required from the server, which will write the logs to a file, and start the next condition.

4.5.3 Client application

The client application would render all 3D objects, based on parameters given in the Unity editor, i.e. before building the application. The targets would be rendered on a 3D box, which we chose to keep opaque so that the "passthrough" hands of the participant are not distracting from the virtual hands. If both were visible, users might try to press buttons thinking their finger is closer than actually. We display the user's virtual hands only for the dominant hand, as the other does not have any effect on the interface. but, we do display the controller strapped to the non-dominant hand so that if it tracks incorrectly, participants can immediately notice and notify us.

The application has a text visible at startup so that participants putting on the headset for the first time have a reference to adjust the lens distance. As the server sends requests, the reference frames get activated, buttons get shown and tracks are made visible. On walking tasks, it plays a metronome (set to 100 BPM), so that the participant has a beat to step on, it is made loud enough so that the moderators also hear it.

If the participants were to always see the first button at the same time, they would always press on the same step, potentially making results less valid. For this reason, we wait 0.5 seconds and a random amount of time ranging from 0 to 1.2 seconds (time that equals 2 steps).

The errors that the user makes that can be detected by the HMD are, swerving off track, clicking on the side of the box, and stopping mid-track. If they activate, the screen turns a yellow hue and beeps. If this happens the measurements start again, randomizing the remaining target sizes.

4.5.4 Passthrough / VR

The current technological limitations of the Quest 3 are a relatively low passthrough video quality and framerate. While it just received an update before the study, which improved the quality immensely, it still leaves much to be desired. In order to overcome these concerns and simulate

(a) (b) (c)

Figure 8: Examples of: (a) First-person view of the linear track with *Palm* (b) Linear track with *Path* (c) Circular track with *Path*

a perfect video passthrough, a VR representation of the real world was considered. The built-in Assisted Space Setup feature of the Quest 3 was used to 3D scan the laboratory environment, and a greyscale polygon mesh was added to all scanned surfaces. Although the VR approach would slightly improve the internal validity of the study, as the environment was completely controlled, the perceived detachment from the real world would arguably decrease external validity. For this reason, the video passthrough approach was used in the study.

A better virtual representation of the study environment was considered, which would e.g. include appropriate textures for walls, ceilings and furniture, but time limitations ruled out this approach. The participants could also feel less safe not knowing where exactly a moveable object or a person is.

4.6 Study procedure

The participant was greeted and asked to read and sign the informed consent form and fill out the demographic questionnaire. A quick verbal introduction was given to the study, after which the HMD and palm mount were fitted. A short acclimation walk was done if the participant had less experience with AR. An initial training period was done for the participants to familiarize themselves with hand input and the target selection task. This was always done in the standing context, and using the first reference frame that the participant was supposed to see. The purpose of this training was to reduce the learning effect, by overcoming the initial steep part of the learning curve. The instructions given in this training period and throughout the study were to select the targets with a performance that balances speed and accuracy. The initial training took around 15 minutes and the study would continue when the participants felt confident in the task and had found an appropriate speed/accuracy balance. Before each new condition, a new training period was done, with the same instructions, which lasted until the participant agreed that they had found an appropriate speed for each target size. A mandatory 5-minute break was added around halftime – usually, after the first reference frame as that takes the longest to train – to reduce fatigue. A short walking training was added when the participant first saw each of the two different tracks, to verify that the participant was able to walk to the step of the metronome. This was done without showing the target selection task box. The path customization step was done just before the first target selection task using that reference frame. A full study procedure takes around 1 and a half to two hours, the whole procedure can be found in appendix B. 4 pilot studies were done to identify and prevent any bias in the moderation.

4.7 Data collection

Gathering complete and comprehensive data from any user study is crucial for thorough analysis afterwards. We generally gathered qualitative data in questionnaires before and after the experimental tasks, and quantitative data by logging during the study.

4.7.1 Logging

Naturally, a large amount of quantitative data had to be logged during the study. This was generally divided into two groups: Selection data, and high-speed data. The same logging class was instantiated once for each of these, and logging was only enabled during an actual trial and not during training. Selection data was logged at every button press and included metrics like movement time from the last target (ms), current target and selection point positions (x,y coordinates on the selection task plane), and target selection success (0/1).
Continuous data was also logged at each frame and included metrics like positions of relevant game objects: track, head, reference frame anchors, targets, controller, dominant hand index finger and palm. See the appendix for a full list of data columns. Logged data was kept in memory and only written to a local file on the headset after the trial was validated by the server (see Figure 7).

Current condition information was also added to each measurement in both logs: Participant ID, measurement ID, movement type, circle direction, reference frame, target size, dominant hand, and system time.

4.8 Participants

We had 24 participants, out of which 9 were women. 22 participants were recruited by word of mouth. Additionally, 2 form respondents were recruited from the university. 19 participants were students, 11 participants were studying and 3 were working in the Department of Computer Science at Aarhus University. Around half of them had considerable experience in VR or AR (on a scale from 1 (novice) to 6 (expert), mean: 2.92, SD: 1.61) and around half had experience using hands to interact in AR or VR (on a scale from 1 (novice) to 6 (expert), mean: 2.62, SD: 1.64). One participant was left-handed. 7 of them wore glasses and another 7 wore contact lenses. The participants generally use their smartphones while walking regularly (on a scale from 1 (never) to 4 (frequently), mean: 3.08, SD: 0.93). All participants were in their twenties, the mean age being 24.6 years (SD: 1.76). The participants did not receive monetary compensation for taking part in the study.

5 Results

In total, there were $24 * 3 * 3 * 4 = 864$ trials. Because of our invalidation during the experiment, only a few data points were deemed invalid. Out of $864 * 6 = 5184$, we invalidated the clicks that were as a whole farther than reasonable (see Figure 9), a total of 8 clicks. These clicks were probably on the wrong button, we asked participants to report these occurrences, alas some were inevitable. We calculated that the step frequency of the participants on average was 103.82 with a standard deviation of 3.264. This means we managed to keep the participants' pace equal. The speed of the participants was instructed to be natural, on the *Linear* path the mean was $4.0km/h$ (STD: 0.36) and on the *Circular* it was $3.5km/h$ (STD: 0.36). This was probably because the participants instinctively slowed down upon the harder task.

The main quantitative analysis was done with RM-ANOVA using ARTools (R package) in Python. Our within-subject factors were movement type, reference frame and target size. We used Aligned

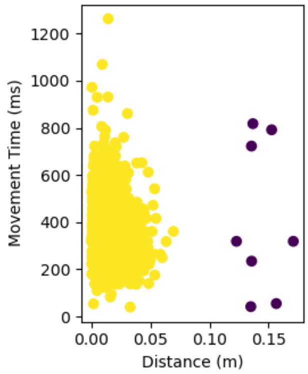


Figure 9: Outliers that we deemed too far, probably misclicks. X axis is the proximity to the centre of the target, and Y is movement time.

Rank Transform (ART) on continuous variables that were not normally distributed according to the Shapiro-Wilk test. We corrected pairwise comparisons with the Bonferroni correction method. We corrected the data with the Greenhouse-Geisser method if the assumption of sphericity was violated. The user questionnaire feedback was analysed with the Friedman statistical method and pairwise differences with the Wilcoxon signed-rank test and Bonferroni corrections. The significance notations mean *: $0.05 \geq p > 0.01$ **: $0.01 \geq p > 0.001$ ***: $0.001 \geq p > 0.0001$ ****: $0.0001 \geq p$

5.1 RQ1: How does movement affect user performance of target selection?

Our results generally showed that the movement type had a significant effect on both Movement time and success rate and therefore also Throughput. When standing, the average time between target selections was $0.321ms$(SD: 0.081), while the *Linear* movement showed a significantly slower mean of $0.340ms$ (SD: 0.089). The *Circular* movement showed an even slower movement time of $0.373ms$ (SD: 0.103).

Movement type also had an adverse effect on success rate, where we measured a mean success rate of 0.769 (SD: 0.243) in standing and a significant decrease in Circular and Linear, which had the values 0.647 (SD: 0.259) and 0.674 (0.254) respectively. Our findings showed a slight decrease in *Circular* success rate compared to *Linear*, but the difference was not significant. Throughput (TP) had statistically significant decreases from Standing $7.489bits/s$ (SD: 2.027) to Linear $6.542bits/s$ (SD: 2.044) and to Circular $5.900bits/s$ (SD: 1.811).

We found interaction between movement type and target size for success rate. Figure 11 shows that in the *Standing* task, button sizes performed similarly well, except the smallest, but as we change movement type, the medium-sized buttons perform worse. This is because, in *Standing*, some participants biologically could not go faster even if their success rate was perfect. While in harder tasks, hitting the buttons accurately meant slowing down. This effect can be summarized as the saturation of accuracy with larger buttons and easier movement tasks.

These results ultimately infer a detrimental effect of movement type on performance, which we can see in the throughput results. It shows a significant difference in performance between all 3 types of movement.

Table 1: Statistical results of our trial measurements, tested with ANOVA

24 participants			
Variable	ANOVA		
	Effect	F-value	p
Movement time	M	65.293	<.001
	RF	29.398	<.001
	T	193.966	<.001
	M:RF	3.265	.011
	M:T	0.800	.569
	RF:T	1.404	.209
Success	M	45.722	<.001
	RF	1.254	.285
	T	249.206	<.001
	M:RF	1.858	.115
	M:T	2.846	.009
	RF:T	1.010	.416
Accuracy (SD_x)	M	17.677	<.001
	RF	1.570	.208
	T	12.834	<.001
	M:RF	0.907	.459
	M:T	2.101	.050
	RF:T	1.232	.287
Effective width (W_e)	M	17.677	<.001
	RF	1.570	.208
	T	12.834	<.001
	M:RF	0.907	.459
	M:T	2.101	.050
	RF:T	1.232	.287
Effective ID (ID_e)	M	18.779	<.001
	RF	1.114	.328
	T	11.700	<.001
	M:RF	0.897	.464
	M:T	1.917	.075
	RF:T	1.193	.307
Throughput (TP)	M	80.620	<.001
	RF	4.806	.008
	T	37.858	<.001
	M:RF	2.174	.070
	M:T	0.759	.602
	RF:T	1.118	.349

5.2 RQ2: How do the reference frames affect user performance?

Looking at movement time, we found that *Palm* had the lowest mean of $0.331ms \pm 0.090ms$, while *PalmWORotation* was slightly slower at $0.342ms \pm 0.093ms$ and the slowest was *Path* at $0.361ms \pm 0.096ms$ As seen in Figure 10, these values differ significantly between *Palm* and *Path* and *PalmWORotation* and *Path*. The choice of reference frame did not show any significant effect on success rate, but this is most probably because this was the most apparent measure to the participant, thus keeping this similar between conditions was one way to "keep the balance between speed and accuracy". As throughput is made up of movement time as well as other measurements, it is not unexpected that *Palm* performs better here too at $6.870bits/s$ (SD: 2.141), higher than *PalmWORotation* at $6.624bits/s$ (SD: 2.076) and *Path* at $6.437bits/s$ (SD: 1.966).

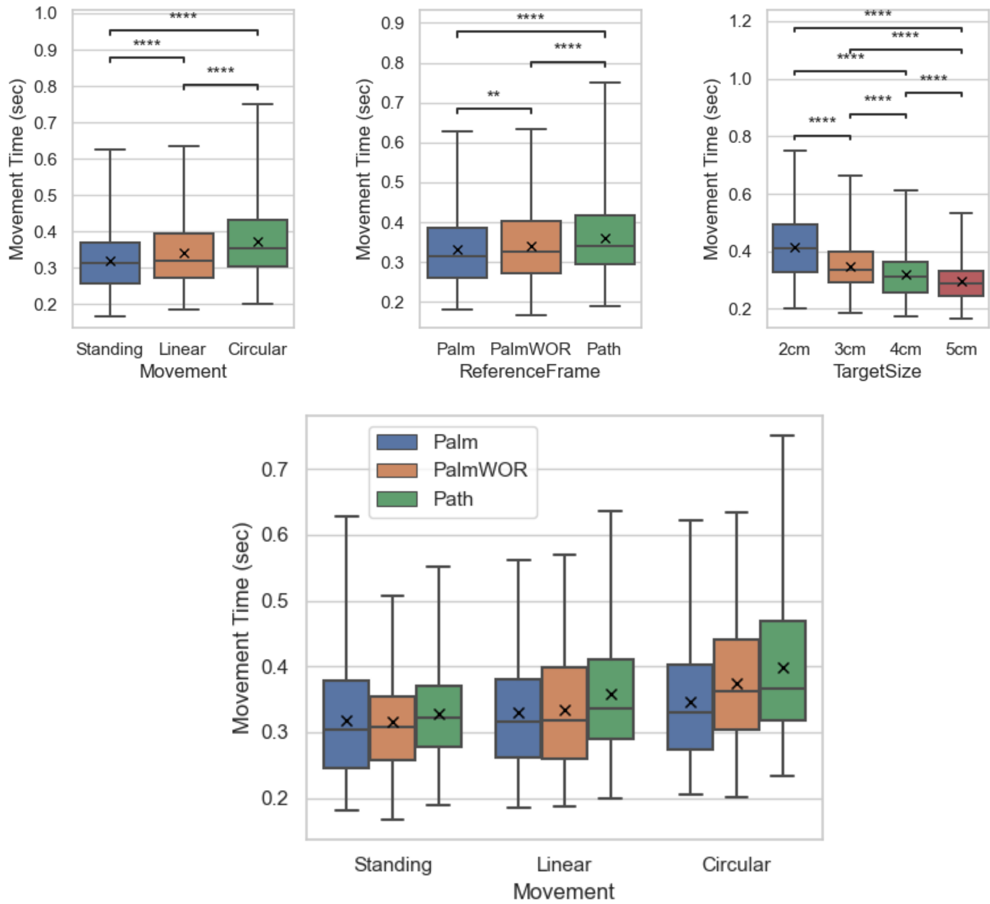


Figure 10: Boxplots depicting movement time (MT), whiskers end on minimum and maximum and box ends on the 25th and 75th percentiles. The X denotes the mean, the line denotes the median. Significance values for $Movement : ReferenceFrame$ can be read in Appendix A

5.3 RQ3: Does having more control over the position of the interface lead to more or less speed and accuracy?

As mentioned, the chosen reference frames, *Path*, *PalmWORotation* and *Palm*, form a continuum of the amount of control a user has over the UI. Since *Palm* has the highest measured throughput and greatest amount of control, while *Path* has the lowest throughput and control, our measurements suggest that increased control correlates with increased performance. The measured throughput of *PalmWORotation* also fell between the two others, which is consistent with this finding. The movement time results also reflect this finding, as *Palm* had the lowest on average, while *Path* had the highest. Since the accuracy did not show significant differences between reference frames, the increase in performance seems to have mainly come from the decrease in movement time.

5.4 RQ4: Is there a reference frame that is suited more for one movement type and another for a different movement?

We found a significant interaction effect between the choice of Reference Frame and Movement in movement time, where it seems *Path* was similar to other reference frames when standing, while worse in walking tasks. *Palm* did not observe a significant increase in movement time from *Standing* to *Linear* or from *Linear* to *Circular* while from *Standing* to *Circular* it did. *PalmWORotation* faired similarly to *Palm*. *Path* on the other hand significantly increased movement time even between

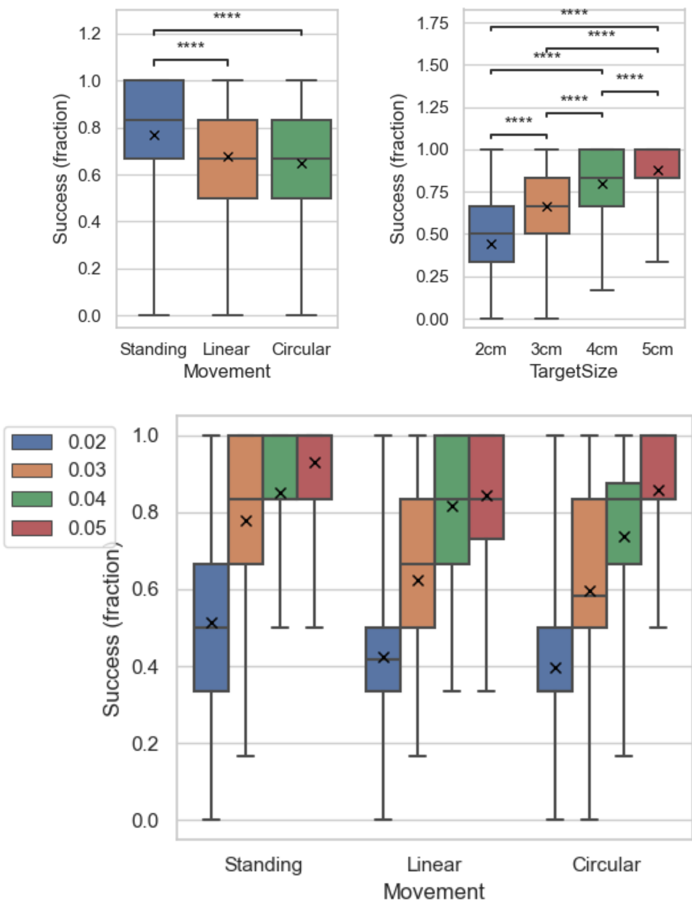


Figure 11: Boxplots depicting success, whiskers end on minimum and maximum and box ends on the 25th and 75th percentiles. The X denotes the mean, the line denotes the median. Significance values for $Movement : TargetSize$ can be read in Appendix A

Standing to *Linear*. We did not find a significant interaction between movement and reference frame choice in other measures.

5.5 RQ5: What is the user experience with the reference frames?

Regarding the user feedback results, we saw a significant increase in physical demand from *Palm* and *PalmWORotation*, compared to *Path*. This means that participants generally experienced more physical demand from reference frames that rely on their non-dominant hand, which is to be expected. Additionally, the participants experienced significantly lower frustration with *Palm* compared to the other two. All other factors in the NASA-TLX questionnaire showed that the choice of reference frame did not have a significant effect on the chosen score. In the end, the preference results did show a significantly lower value for *Palm*, than *PalmWORotation*, with *Path* being in-between. This means that the participants generally preferred *Palm* over any of the other two. The preference standard deviation of the *Path* reference frame was slightly higher than the two other reference frames, which indicates that the participants generally had different opinions of *Path*. Contrary to this, there was a stronger agreement on the preference of *PalmWORotation*, as the standard deviation was lower, and it was the least preferred.

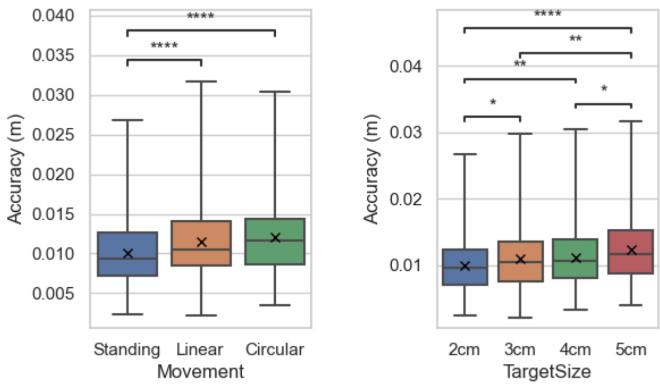


Figure 12: Boxplots depicting accuracy (SD_x), whiskers end on minimum and maximum and box ends on the 25th and 75th percentiles. The X denotes the mean, the line denotes the median.

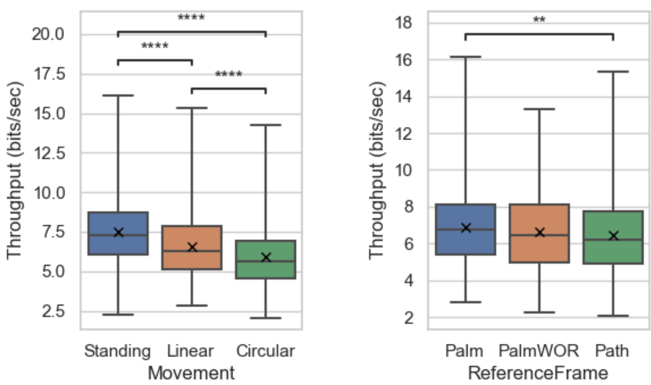


Figure 13: Boxplots depicting throughput (TP), whiskers end on minimum and maximum and box ends on the 25th and 75th percentiles. The X denotes the mean, the line denotes the median.

5.6 User feedback

This subsection covers the qualitative feedback from the final questionnaire. We will highlight important quotes that contextualize the above ordinal and quantitative data.

5.6.1 Physical demand

A common feedback response regarding the reference frames was remarked on the physical demand of specifically the two palm-anchored reference frames, *Palm* and *PalmWORotation*. 4 participants directly addressed the physical demand of these, and described them as e.g. *"exhausting"* (P16) or *"physically demanding"* (P7). Two participants (P17, P14) explicitly reported tiredness, e.g. (P14:

Table 2: User feedback results, statistically tested with the Friedman method

24 participants			
Variable	Effect	Chi-squared	p
Mental demand	RF	2.747	.253
Physical demand	RF	14.439	<.001
Temporal demand	RF	2.882	.237
Performance	RF	1.859	.395
Effort	RF	4.431	.109
Frustration	RF	6.500	.039
Preference	RF	7.750	.020

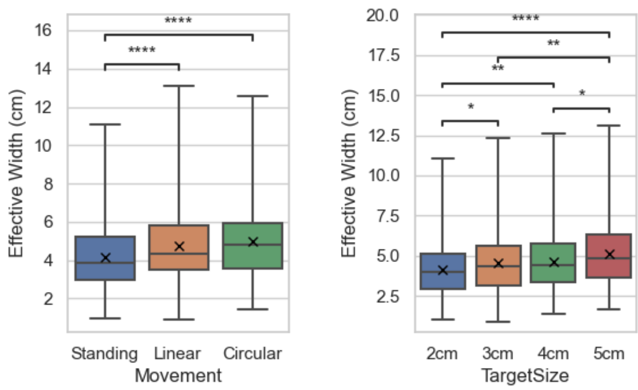


Figure 14: Boxplots depicting effective width (W_e), whiskers end on minimum and maximum and box ends on the 25th and 75th percentiles. The X denotes the mean, the line denotes the median.

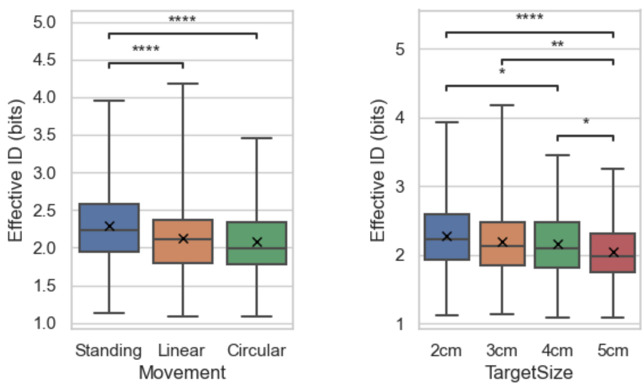


Figure 15: Boxplots depicting effective Index of Difficulty (ID_e), whiskers end on minimum and maximum and box ends on the 25th and 75th percentiles. The X denotes the mean, the line denotes the median.

"My arm became more tired when I had to hold it up and it was jittering a bit, (...)". One participant described the *Palm* as *"annoying"* (P21). All these comments were consistent with the NASA TLX Physical demand results and the *Path* reference frame also got positive feedback that follows this. P11 commented on the increased comfortability, but still experienced less accuracy: *"The most inaccurate from my point of view was the path one, but it was one I liked the most because didn't stress my left hand."*, and P16 too, described it as *"(...) more comfortable"*. P13, although ranking it as least preferred, commended it in the *Standing* context: *"Standing stationary I remember thinking it was nice in the Path one to not have to use both hands"*. Regardless of the physical aspect, most people still preferred *Palm* as it: *"(...) felt the most versatile"*(P5) and *"(...) had the most freedom in terms of the movement"*.

5.6.2 Circle track

Multiple participants raised different issues on the *Circle* track. 5 of the participants specifically raised issues with *Path* on the *Circle* track, with P7 e.g. saying: *"Walking in a circle with the path reference didn't seem symmetric for both directions. I was trying to have the buttons on the right side, so in one direction, I was walking more outside"*. P11 also described it as *"(...) annoying (...)"* because *"(...) the keyboard was placed too much to the left, or to the right, (...)"*. P13 expressed: *"In the path, I remember feeling out of balance when walking in the circle, especially when travelling left from the starting point."*. Opposite to this, P17 said that: @texit"Both palm and palm w/o rotation were distracting me from going in the circle, and I found I could concentrate more on

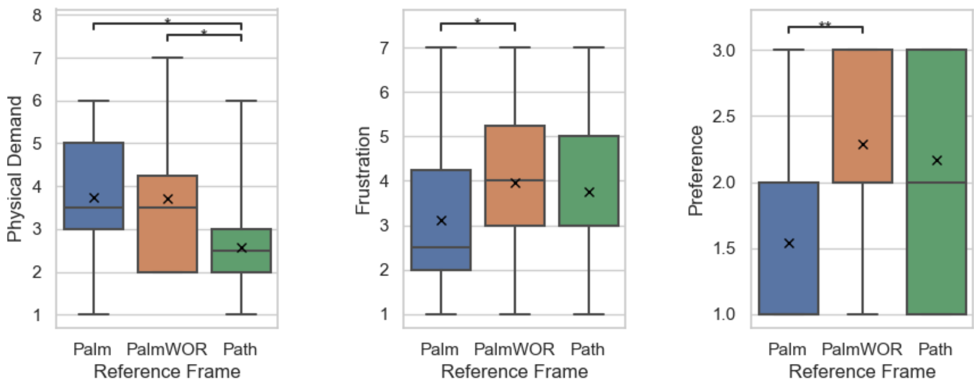


Figure 16: Boxplots depicting physical demand (left) and frustration (middle) and preference (right), whiskers end on minimum and maximum and box ends on the 25th and 75th percentiles. The X denotes the mean, the line denotes the median.

where I was walking in the path condition.". P8 commented on the design of the *PalmWORotation* reference frame, where: *"(...) in one direction the UI would move towards my chest, while in the other direction it would move away from my chest."* In contrast, P20 preferred *Palm* on the *circle* track: *"I also think the palm one was significantly more comfortable once moving in circles"*.

5.6.3 Path reference frame

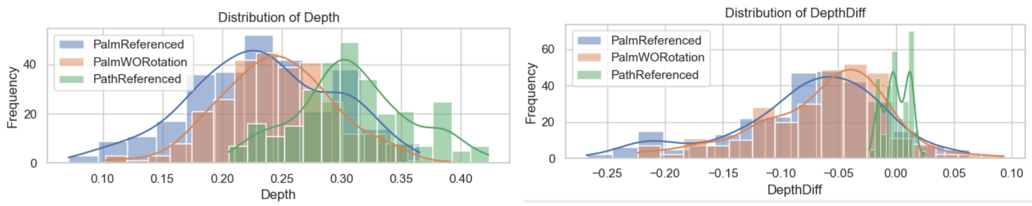
3 participants also mentioned pacing issues in the *Path* reference frame specifically. P18 said: *'The path gave the uncomfortable impression that there was an unnecessary factor of time involved, in an odd way the target felt as if it was going out of reach.'*. P24's only comment on *Path* was: *'path could cause problems with pacing but overall fine'*, and P28 experienced similar issues: *'The path reference frame felt a bit as if I would walk through it which may have affected my pace'*. In contrast to these comments, multiple participants found it better when standing. P23 said that *Path*: *"The Path was easier than other two during standing but more difficult during walking."*. P5 agreed with this: *"Path was quite annoying when I was moving, but it worked very well when I was standing still"*. P19 noted that *Path* was more stable: *"... But it was more comfortable for the eyes when it was stable."*

5.7 Supplementary Metrics

This subsection presents additional results from the logs that describe the placement and rotation of the targets, relative to the user. The metrics are: Depth, Decline and Relative Yaw, which can help shed light on the ergonomics of each reference frame. Depth: The horizontal distance from the head to the targets, along the walking direction axis. Decline: The difference between head height and target height. Relative Yaw: The relative rotation of the targets towards the dominant hand, around the vertical axis. I.e. 0 means that the targets point in the direction of the head, and positive degrees show the amount of rotation towards the dominant hand.

5.7.1 Depth

The average depth of *Path* referenced targets was $0.309m$ (SD:0.047), while the *Palm* and *PalmWORotation* depths were $0.232m$ (SD: 0.06) and $0.246m$ (SD: 0.047) respectively (See Figure 17(a)). What this means is that, on average, participants held the targets in *Palm* $0.076m$ (SD:0.065) closer and *PalmWORotation* $0.062m$ (ST:0.056) closer. This is also shown in Figure 17(b).

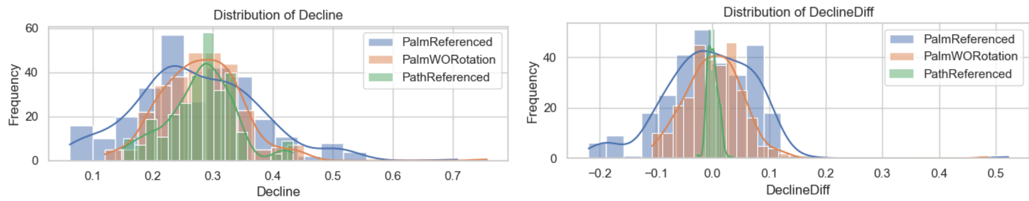


(a) (b)

Figure 17: Frequency of per-trial averages of: (a) Depth (b) The depth difference in path trials and the average target placement in other trials within the same participant

5.7.2 Decline

The average decline of *Palm* and *PalmWORotation* fell within $1cm$ of *Path*. This can be seen in Figure 18, where the average decline in each trial was $0.280m$(SD:0.056) for *Path*, $0.283m$(SD:0.069) for *PalmWORotation*, and $0.273m$(SD:0.103) for *Palm*.

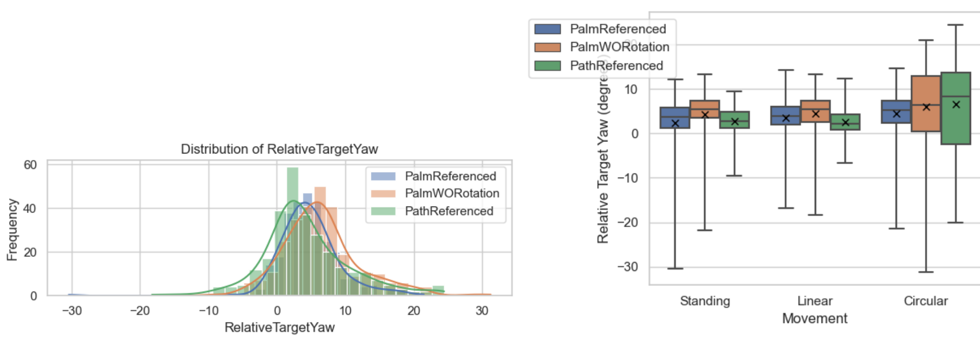


(a) (b)

Figure 18: Frequency of per-trial averages of: (a) Decline: Vertical distance from head height to targets (b) The difference between average target decline in path trials and the average target decline in other trials within the same participant

5.7.3 Relative target yaw

The targets were found to be rotated slightly towards the dominant hand in all reference frames. On average *Path* was rotated by $4.3°$ (SD: 6.39), *PalmWORotation* by $6.42°$ (SD:5.79) and *Palm* by $4.82°$ (SD:5.79) (See Figure 19(a)). The conditions that contributed the most to this spread were the *Circular+PalmWORotation*(SD: $7.84°$) and *Circular+Path*(SD: $9.98°$), which can be seen in Figure 19(b).



(a) (b)

Figure 19: (a) The relative rotation of the targets with 0 being pointing directly at the head, and positive degrees being rotation towards the dominant hand (b) Relative Yaw in each *Movement*

6 Discussion

Where to place user interfaces in Augmented Reality during locomotion? There seems to be one clear answer based on the results. *Palm* proved to be the least frustrating, most functional reference frame in our study. This could be caused by multiple factors, but we conjecture that it might be due to ergonomic reasons. As shown in section 5.7.1, the participants would hold the targets 6-7 cm closer on average in non-*Path* trials, even though they were given the opportunity to reposition the *Path* referenced targets after the first training. This, as well as the relatively large spread in Figure 17(b) and 18(b) suggest that participants would actively utilize the extra mobility of the hands to increase ergonomics between trials. Intuitively, hands are meant to be reference frames, as we use our mobile phones with our hands when walking. One thing to note is that most people use their smartphones with one hand, and we focused on two-handed input in this thesis. Nevertheless, the ergonomics of holding a reference frame and being able to position it with our palm is probably translated into the higher throughput and preference for *Palm*. The results seem to be in line with the results of Weiss et al. [34], who found that a handheld device could achieve better target acquisition than a *world-referenced* AR interface. This could have closed the gap.

One problem participants had with our locked rotation reference frames was that it was looking away from them. As seen in Figure 19(b), participants would position themselves or the targets with a consistent angle towards the dominant hand, except on the *Circular* track with *PalmWORotation* or *Path*. The high spread in these two conditions confirms the experience of the targets looking away and shows that participants were not able to point the targets towards the dominant hand in a consistent manner. The general problem with the *Path* reference frame is that it forces the user to follow one particular path to be optimally positioned, which is particularly challenging on the *Circular* track. This is in some part just an implementation detail of the reference frame, which could be optimized in future work and potentially generalized to real-world walking paths. However, enabling the participant to move freely would need a more advanced algorithm to calculate its position. It would need to know the user's path in the future and whether one movement is part of the path or just a random sideways motion. On the *Circular* track, the relative yaw of *Palm* reference targets was found to be almost consistent with the two other *Movement* types, indicating better ergonomics. Additionally, while not significant, *Palm* was also found to perform better than *Path* in standing and significantly better in linear walking, leading to the conclusion that *Palm* might just be more ergonomic overall.

However, there were exceptions, for some people, the *Path* seemed much more usable, as it provided them a stable way to interact with the UI. While they were clearly in the minority, we propose that reference frames in the future should be customizable, and switching between them should be easy. The physical demand that *Palm* needed would be draining for users over a longer period, this is shown by the subjective results.

While the combination of stability and control intuitively makes sense, the reference frame in-between, *PalmWORotation*, performed the worst. The frustration levels users experienced were as high as *Path*, but the physical demand was as high as *Palm*. Throughput was not as good as *Palm*, but this difference was not significant. While there were some who performed best on this reference frame, they were outliers, and when presented with the data, they would also be surprised.

During walking and especially during turning, we lose throughput. This is similar to what other research has found [29, 3, 18]. This is useful because it means that in the future, interfaces could change button sizes according to the environment to preserve the Index of Difficulty [12]. In other metrics, like success and effective width, the difference between linear and circular walking was negligible, which could be because the act of walking influences performance more than turning, at normal speeds. It is also possible that the radius of turning was not small enough for more significant differences.

Hand input on the go? More than anything, our research shows that Fitts' law is also applicable in AR environments as target size did influence results accordingly. This is promising because this could mean that the error caused by sub-par hand tracking was low enough to not cause random outcomes. For results between reference frames, we can see that a hand-controlled reference frame performed best, which illustrates the hand tracking precision. However, we used controllers to track hands, precisely because we observed that if we use optical hand tracking, the shakiness increases to a level that might influence results. While on the right path, hand tracking can still improve and this is especially true on the go.

6.1 Limitations and future work

One of the biggest differences we had from an ideal world was that we had to use a controller to track the palm. This is problematic because it adds extra weight to the hands and constantly touches them, which could be distracting. However, adding weight to the hand might have had a stabilizing effect. It would certainly be interesting to see whether adding more or less weight to the hands affects stability and in effect accuracy. The extra weight is also a source of dissatisfaction as the results show palm-referenced frames have higher physical demand, this might be alleviated by free hands.

Frustration was one of the factors *PalmWORotation* reference frame was worse in than *Palm*. This is most likely because it was rotating in a way the users were not used to, as also indicated by Figure 19(b). This might also have affected *Path* even if it was not significantly more frustrating. Exploring further ways to position the reference frames could bridge the gap between this and *Palm*. One main way we see to do that is to make the reference frame free-moving and not constrained to a predefined path, as that would alleviate the problem that people don't walk in the middle of predefined paths. It would also ensure that the study has more external validity, as the current setup is very restricting.

Another limitation is that we did not use any obstacles and one rarely walks in a perfect circle or straight line. The participants also only had one task, to follow the track, and if the task was to go from A to B, we would have observed a lot more different outcomes. It would be important to see whether our numbers hold up in an uncontrolled environment. Walking also includes acceleration and deceleration, our participants walked at a constant speed to a constant frequency of steps. In the future, if everyone is using AR glasses to interact with their digital life, a simple stop light for a zebra crossing should not completely stop the interaction.

Another limitation of our study is that our participants are rather homogeneous. It could be interesting to see whether it would have any impact on the results if we included people of all ages, as they could have less experience with AR or smartphones specifically.

We discussed the option of including jogging in our study, but we decided against it because the headset was too heavy. In the future when this becomes more feasible, it would be interesting to see whether hand-stabilised reference frames still perform better than world-stabilized. Additionally, some participants reported that the headset was too heavy, even after fifteen minutes of use, to make the study more externally valid, repeating it with a lighter device could change the results further.

Lastly, the standard we used to compare the reference frames, ISO 9241-9, while good for a comprehensive and objective generation of a number, might not be too generalisable to real-world results. Developing more complex applications to work on these reference frames, and then comparing them qualitatively might give us more insight into their performance, especially in the eyes of the users. Our task was boring to multiple participants and competitive for others, if we want results that mimic the attitude of a generic user, we might make the task more complex.

7 Conclusion

In this study, we investigate how the choice of spatial reference frames in AR, affects the speed and accuracy of selecting targets with one's hand, during locomotion on linear and circular paths. We present a Fitts' law study that compares 3 different reference frames, which span a design space of increasing amount of control, in 3 different movement types: standing, linear walking and circular walking. We show empirical evidence that the choice of reference does affect the hand input performance when walking, and that the performance is positively correlated in general, to the amount of control the user has over the targets. This is especially true for interaction during movement. This finding is consistent with the fact that the reference frame with the most amount of control is the most preferred. Additionally, we examine the relevant dependent variables which gives detailed insights into the factors that are affected the most. Finally, this work fills the knowledge gap of hand input on the go and provides a foundation for further research or reference frame design for applications used during locomotion.

8 Acknowledgements

We would like to thank our supervisors for their unwavering support during the semester, in feedback, in helping us design the study and specifically Pavel for helping us in conducting studies and the analysis of our results. We would also like to thank our 28 participants who dedicated around two hours to our study.

References

[1] Ferran Argelaguet and Carlos Andujar. A survey of 3d object selection techniques for virtual environments. *Computers Graphics*, 37(3):121–136, 2013.

[2] Ronald T Azuma. A Survey of Augmented Reality. 1997.

[3] Joanna Bergstrom-Lehtovirta, Antti Oulasvirta, and Stephen Brewster. 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*, MobileHCI '11, page 143–146, New York, NY, USA, 2011. Association for Computing Machinery.

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

[5] Volkert Buchmann, Stephen Violich, Mark Billinghurst, and Andy Cockburn. Fingartips: gesture based direct manipulation in augmented reality. In *Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, GRAPHITE '04, page 212–221, New York, NY, USA, 2004. Association for Computing Machinery.

[6] Zoe Chan, Aislinn MacPhail, Ivan Au, Janet Zhang-Lea, Ben Lam, Reed Ferber, and Roy Cheung. Walking with head-mounted virtual and augmented reality devices: Effects on position control and gait biomechanics. *PLOS ONE*, 14:e0225972, 12 2019.

[7] James Crowley, François Bérard, and Joëlle Coutaz. Finger Tracking as an Input Device for Augmented Reality. May 1999.

[8] Barrett Ens, Juan David Hincapié-Ramos, and Pourang Irani. 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*, SUI '14, page 2–12, New York, NY, USA, 2014. Association for Computing Machinery.

[9] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. Windows on the world: 2d windows for 3d augmented reality. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, UIST '93, page 145–155, New York, NY, USA, 1993. Association for Computing Machinery.

[10] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139–183. North-Holland, 1988.

[11] ISO ISO. 9241-9 ergonomic requirements for office work with visual display terminals (vdts)-part 9: Requirements for non-keyboard input devices (fdis-final draft international standard), 2000. *International Organization for Standardization*, 2000.

[12] Shaun Kane, Jacob Wobbrock, and Ian Smith. Getting off the treadmill: evaluating walking user interfaces for mobile devices in public spaces. pages 109–118, 09 2008.

[13] Mohamed Khamis, Carl Oechsner, Florian Alt, and Andreas Bulling. Vrpursuits: interaction in virtual reality using smooth pursuit eye movements. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, AVI '18, New York, NY, USA, 2018. Association for Computing Machinery.

[14] Elisa Maria Klose, Nils Adrian Mack, Jens Hegenberg, and Ludger Schmidt. Text presentation for augmented reality applications in dual-task situations. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 636–644, 2019.

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

[16] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. Chapter 3 - experimental design. In Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser, editors, *Research Methods in Human Computer Interaction (Second Edition)*, pages 45–69. Morgan Kaufmann, Boston, second edition edition, 2017.

[17] Juyoung Lee, Sang Ahn, and Jae-In Hwang. A walking-in-place method for virtual reality using position and orientation tracking. *Sensors*, 18:2832, 08 2018.

[18] Yang Li, Juan Liu, Jin Huang, Yang Zhang, Xiaolan Peng, Yulong Bian, and Feng Tian. Evaluating the Effects of User Motion and Viewing Mode on Target Selection in Augmented Reality, July 2023.

[19] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. Glanceable ar: Evaluating information access methods for head-worn augmented reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 930–939. IEEE, 2020.

[20] Yiqin Lu, Bingjian Huang, Chun Yu, Guahong Liu, and Yuanchun Shi. Designing and evaluating hand-to-hand gestures with dual commodity wrist-worn devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(1), mar 2020.

[21] Yujun Lu, BoYu Gao, Huawei Tu, Huiyue Wu, Weiqiang Xin, Hui Cui, Weiqi Luo, and Henry Been-Lirn Duh. Effects of physical walking on eyes-engaged target selection with ray-casting pointing in virtual reality. *Virtual Reality*, August 2022.

[22] Andrés Lucero and Akos Vetek. Notifeye: using interactive glasses to deal with notifications while walking in public. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, ACE '14, New York, NY, USA, 2014. Association for Computing Machinery.

[23] I. Scott MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact.*, 7(1):91–139, mar 1992.

[24] I. Scott MacKenzie. *Fitts' Law*, chapter 17, pages 347–370. John Wiley Sons, Ltd, 2018.

[25] Pavel Manakhov, Ludwig Sidenmark, Ken Pfeuffer, and Hans Gellersen. 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, New York, NY, USA, 2024. Association for Computing Machinery.

[26] Steve Mann and James Fung. 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*, pages 48–55, 2001.

[27] Florian Müller, Martin Schmitz, Daniel Schmitt, Sebastian Günther, Markus Funk, and Max Mühlhäuser. Walk the line: Leveraging lateral shifts of the walking path as an input modality for head-mounted displays. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–15, New York, NY, USA, 2020. Association for Computing Machinery.

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

[29] Bastian Schildbach and Enrico Rukzio. 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*, MobileHCI '10, page 93–102, New York, NY, USA, 2010. Association for Computing Machinery.

[30] R. William Soukoreff and I. Scott MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61(6):751–789, December 2004.

[31] Ivan E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, AFIPS '68 (Fall, part I), page 757–764, New York, NY, USA, 1968. Association for Computing Machinery.

[32] Takumi Toyama, Daniel Sonntag, Jason Orlosky, and Kiyoshi Kiyokawa. Attention engagement and cognitive state analysis for augmented reality text display functions. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI '15, page 322–332, New York, NY, USA, 2015. Association for Computing Machinery.

[33] Uta Wagner, Mathias N. Lystbæk, Pavel Manakhov, Jens Emil Sloth Grønbæk, Ken Pfeuffer, and Hans Gellersen. 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*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery.

[34] Hannah Weiss, Jianyang Tang, Connor Williams, and Leia Stirling. Performance on a target acquisition task differs between augmented reality and touch screen displays. *Applied Ergonomics*, 116:104185, April 2024.

[35] Qiushi Zhou, Difeng Yu, Martin N Reinoso, Joshua Newn, Jorge Goncalves, and Eduardo Velloso. Eyes-free target acquisition during walking in immersive mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3423–3433, 2020.

[36] Qiushi Zhou, Difeng Yu, Martin N Reinoso, Joshua Newn, Jorge Goncalves, and Eduardo Velloso. Eyes-free target acquisition during walking in immersive mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3423–3433, 2020.

Appendices

A Significance Tables

A.1 Movement time ($Movement : ReferenceFame$)

Factor-pair 1		Factor-pair 2		Significance
Movement	ReferenceFrame	Movement	ReferenceFrame	
Circle	PalmReferenced	Circle	PalmWORotation	0.003
Circle	PalmReferenced	Circle	PathReferenced	<0.001
Circle	PalmReferenced	Standing	PalmReferenced	0.006
Circle	PalmReferenced	Standing	PalmWORotation	0.007
Circle	PalmReferenced	Standing	PathReferenced	1.0
Circle	PalmReferenced	Walking	PalmReferenced	1.0
Circle	PalmReferenced	Walking	PalmWORotation	1.0
Circle	PalmReferenced	Walking	PathReferenced	1.0
Circle	PalmWORotation	Circle	PathReferenced	0.587
Circle	PalmWORotation	Standing	PalmReferenced	<0.001
Circle	PalmWORotation	Standing	PalmWORotation	<0.001
Circle	PalmWORotation	Standing	PathReferenced	<0.001
Circle	PalmWORotation	Walking	PalmReferenced	<0.001
Circle	PalmWORotation	Walking	PalmWORotation	<0.001
Circle	PalmWORotation	Walking	PathReferenced	1.0
Circle	PathReferenced	Standing	PalmReferenced	<0.001
Circle	PathReferenced	Standing	PalmWORotation	<0.001
Circle	PathReferenced	Standing	PathReferenced	<0.001
Circle	PathReferenced	Walking	PalmReferenced	<0.001
Circle	PathReferenced	Walking	PalmWORotation	<0.001
Circle	PathReferenced	Walking	PathReferenced	0.001
Standing	PalmReferenced	Standing	PalmWORotation	1.0
Standing	PalmReferenced	Standing	PathReferenced	0.982
Standing	PalmReferenced	Walking	PalmReferenced	1.0
Standing	PalmReferenced	Walking	PalmWORotation	0.311
Standing	PalmReferenced	Walking	PathReferenced	<0.001
Standing	PalmWORotation	Standing	PathReferenced	1.0
Standing	PalmWORotation	Walking	PalmReferenced	1.0
Standing	PalmWORotation	Walking	PalmWORotation	0.371
Standing	PalmWORotation	Walking	PathReferenced	<0.001
Standing	PathReferenced	Walking	PalmReferenced	1.0
Standing	PathReferenced	Walking	PalmWORotation	1.0
Standing	PathReferenced	Walking	PathReferenced	0.007
Walking	PalmReferenced	Walking	PalmWORotation	1.0
Walking	PalmReferenced	Walking	PathReferenced	0.001
Walking	PalmWORotation	Walking	PathReferenced	0.035

A.2 Success ($Movement : TargetSize$)

Factor-pair 1		Factor-pair 2		Significance
Movement	TargetSize	Movement	TargetSize	
Circle	0.02	Circle	0.03	<0.001
Circle	0.02	Circle	0.04	<0.001
Circle	0.02	Circle	0.05	<0.001
Circle	0.02	Standing	0.02	0.063
Circle	0.02	Standing	0.03	<0.001
Circle	0.02	Standing	0.04	<0.001
Circle	0.02	Standing	0.05	<0.001
Circle	0.02	Walking	0.02	1.0
Circle	0.02	Walking	0.03	<0.001
Circle	0.02	Walking	0.04	<0.001
Circle	0.02	Walking	0.05	<0.001
Circle	0.03	Circle	0.04	<0.001
Circle	0.03	Circle	0.05	<0.001
Circle	0.03	Standing	0.02	0.827
Circle	0.03	Standing	0.03	<0.001
Circle	0.03	Standing	0.04	<0.001
Circle	0.03	Standing	0.05	<0.001
Circle	0.03	Walking	0.02	<0.001
Circle	0.03	Walking	0.03	1.0
Circle	0.03	Walking	0.04	<0.001
Circle	0.03	Walking	0.05	<0.001
Circle	0.04	Circle	0.05	0.001
Circle	0.04	Standing	0.02	<0.001
Circle	0.04	Standing	0.03	1.0
Circle	0.04	Standing	0.04	0.004
Circle	0.04	Standing	0.05	<0.001
Circle	0.04	Walking	0.02	<0.001
Circle	0.04	Walking	0.03	<0.001
Circle	0.04	Walking	0.04	1.0
Circle	0.04	Walking	0.05	0.107
Circle	0.05	Standing	0.02	<0.001
Circle	0.05	Standing	0.03	1.0
Circle	0.05	Standing	0.04	1.0
Circle	0.05	Standing	0.05	0.224
Circle	0.05	Walking	0.02	<0.001
Circle	0.05	Walking	0.03	<0.001
Circle	0.05	Walking	0.04	1.0
Circle	0.05	Walking	0.05	1.0

Factor-pair 1		Factor-pair 2		Significance
Movement	TargetSize	Movement	TargetSize	
Standing	0.02	Standing	0.03	<0.001
Standing	0.02	Standing	0.04	<0.001
Standing	0.02	Standing	0.05	<0.001
Standing	0.02	Walking	0.02	0.219
Standing	0.02	Walking	0.03	0.302
Standing	0.02	Walking	0.04	<0.001
Standing	0.02	Walking	0.05	<0.001
Standing	0.03	Standing	0.04	1.0
Standing	0.03	Standing	0.05	<0.001
Standing	0.03	Walking	0.02	<0.001
Standing	0.03	Walking	0.03	<0.001
Standing	0.03	Walking	0.04	1.0
Standing	0.03	Walking	0.05	1.0
Standing	0.04	Standing	0.05	0.115
Standing	0.04	Walking	0.02	<0.001
Standing	0.04	Walking	0.03	<0.001
Standing	0.04	Walking	0.04	1.0
Standing	0.04	Walking	0.05	1.0
Standing	0.05	Walking	0.02	<0.001
Standing	0.05	Walking	0.03	<0.001
Standing	0.05	Walking	0.04	<0.001
Standing	0.05	Walking	0.05	0.004
Walking	0.02	Walking	0.03	<0.001
Walking	0.02	Walking	0.04	<0.001
Walking	0.02	Walking	0.05	<0.001
Walking	0.03	Walking	0.04	<0.001
Walking	0.03	Walking	0.05	<0.001
Walking	0.04	Walking	0.05	1.0

B Study procedure

- Book the room well before.
- Before the study:
  - Clean the headset's lenses with a fibre cloth.
  - Charge the headset.
  - Charge the headset.
  - Check that you have the charging cable for your laptop.
  - Print out the information duty informed consent forms.
- Put the "Study is in progress" signs on both doors/close them.
- Close the curtains turn on all the lights.
- Make sure that the environment and controllers are aligned with their physical counterparts.
  - In order to align VR controllers you gotta clear physical space boundaries.
- Meeting greeting
  - If a participant doesn't know their way around the campus, meet them at the entrance and walk to the room; make sure they're comfortable relaxed, establish rapport by talking about anything except the experiment.
- Signing the GDPR and informed consent forms demographic questionnaire.

  - Ask them to sign both copies of the informed consent for participation.
  - Give them the information duty form (they don't need to sign it, it's enough for them to just have it).
  - If they are eager to begin the study and don't want to read them, just summarize their content verbally before asking them to sign it.
  - Briefly explain what's going to happen now.
- Onboarding acclimatization
  - Check the volume level on the headset launch the prototype.
  - Don't forget to start mirroring the headset view for training only.
  - Headset fitting for comfort (explain how strap and IPD adjustments work *before* putting the headset on).
  - Don't strap participants in the glove right away, give them time to get accustomed to the pass-through, especially those who had no previous experience with XR.
  - Ask the participant to go around and touch the walls and objects in order for them to feel safe.
  - Without taking off the headset, put the glove on while explaining why we do that.
- Main study
  - Regardless of the order of the conditions, the training of the selection task while standing is always first; the instructions for the training are as follows:
    * Select targets as quickly as possible but at the same time as accurately as possible. Practice until you feel that you've mastered the task and you've found an appropriate balance between speed and accuracy for all target sizes.
    * The participant must feel confident that they have mastered the task before proceeding (you'll see it with your own eyes how good they are).
    * Also, don't forget to mention that their selection strategy can be adapted to the target size, i.e. they can be faster with bigger targets.
  - Before first walking and jogging conditions: do *the walking training under the metronome sound* and the full training that combines selecting targets with physical movement (if movement first).
  - For the Path RF: whichever movement condition comes first for this reference frame, we set up the height and depth of the target individually for each participant. It's important to ask them after they've done several training trials how comfortable the position is and reset it if it's not before control trials begin
  - Explain the importance of verbally reporting all issues.
  - Whenever going from one movement condition to another (e.g., from linear to curvilinear walking), ask participants to walk under the metronome sound a couple of times, as I observed that changing paces is hard for some people.
  - Wait for the participant to stop before validating or invalidating the trial.
  - It's prohibited to talk to the participant while they are selecting targets during a controlled trial.
  - Hitting targets together with the sound of the metronome should be discouraged.
  - At least one short break (5 min) should be taken after the first reference frame and every time when the participant asks for it. After a break, the headset should be restarted.
- Final questionnaire debriefing
  - Administer the final questionnaire.
  - Explain in your own words why it was valuable for us that the person took part in our study. Answer any questions they have and walk them out, if necessary.
  - Remind them to take their copies of the documents.
- After the study
  - Check the validity of the data.
    This was copied from the shared file we used, written by Pavel Manakhov.

C Logging

C.1 Selections logger

- ParticipantID
- SelectionID
- Movement
- CircleDirection
- ReferenceFrame
- TargetSize
- DominantHand
- RealtimeSinceStartupMs
- SystemClockTimestampMs
- ActiveTargetIndex
- AbsoluteTargetPositionX
- AbsoluteTargetPositionY
- AbsoluteSelectionPositionX
- AbsoluteSelectionPositionY
- LocalSelectionPositionX
- LocalSelectionPositionY
- Success
- SelectionDuration

C.2 High-frequency logger

- ParticipantID
- MeasurementID
- Movement
- CircleDirection
- ReferenceFrame
- TargetSize
- DominantHand
- RealtimeSinceStartupMs
- SystemClockTimestampMs
- TrackPositionX
- TrackPositionY
- TrackPositionZ
- TrackForwardX
- TrackForwardY
- TrackForwardZ
- TrackUpX
- TrackUpY
- TrackUpZ
- TrackQuaternionX
- TrackQuaternionY
- TrackQuaternionZ
- TrackQuaternionW
- WalkingDirectionPositionX
- WalkingDirectionPositionY
- WalkingDirectionPositionZ
- WalkingDirectionForwardX
- WalkingDirectionForwardY
- WalkingDirectionForwardZ
- WalkingDirectionUpX
- WalkingDirectionUpY
- WalkingDirectionUpZ
- WalkingDirectionQuaternionX
- WalkingDirectionQuaternionY
- WalkingDirectionQuaternionZ
- WalkingDirectionQuaternionW
- HeadPositionX
- HeadPositionY
- HeadPositionZ
- HeadForwardX
- HeadForwardY
- HeadForwardZ
- HeadUpX
- HeadUpY
- HeadUpZ
- HeadQuaternionX
- HeadQuaternionY
- HeadQuaternionZ
- HeadQuaternionW
- DominantPalmCenterPositionX
- DominantPalmCenterPositionY
- DominantPalmCenterPositionZ
- DominantPalmCenterForwardX
- DominantPalmCenterForwardY
- DominantPalmCenterForwardZ
- DominantPalmCenterUpX
- DominantPalmCenterUpY
- DominantPalmCenterUpZ
- DominantPalmCenterQuaternionX
- DominantPalmCenterQuaternionY
- DominantPalmCenterQuaternionZ
- DominantPalmCenterQuaternionW
- DominantIndexTipPositionX
- DominantIndexTipPositionY
- DominantIndexTipPositionZ
- DominantIndexTipForwardX
- DominantIndexTipForwardY
- DominantIndexTipForwardZ
- DominantIndexTipUpX
- DominantIndexTipUpY

- DominantIndexTipUpZ
- DominantIndexTipQuaternionX
- DominantIndexTipQuaternionY
- DominantIndexTipQuaternionZ
- DominantIndexTipQuaternionW
- ControllerPositionX
- ControllerPositionY
- ControllerPositionZ
- ControllerForwardX
- ControllerForwardY
- ControllerForwardZ
- ControllerUpX
- ControllerUpY
- ControllerUpZ
- ControllerQuaternionX
- ControllerQuaternionY
- ControllerQuaternionZ
- ControllerQuaternionW
- AllTargetsPositionX
- AllTargetsPositionY
- AllTargetsPositionZ
- AllTargetsForwardX
- AllTargetsForwardY
- AllTargetsForwardZ
- AllTargetsUpX
- AllTargetsUpY
- AllTargetsUpZ
- AllTargetsQuaternionX
- AllTargetsQuaternionY
- AllTargetsQuaternionZ
- AllTargetsQuaternionW
- ActiveTargetPositionX
- ActiveTargetPositionY
- ActiveTargetPositionZ
- ActiveTargetForwardX
- ActiveTargetForwardY
- ActiveTargetForwardZ
- ActiveTargetUpX
- ActiveTargetUpY
- ActiveTargetUpZ
- ActiveTargetQuaternionX
- ActiveTargetQuaternionY
- ActiveTargetQuaternionZ
- ActiveTargetQuaternionW
- SelectorProjectionOntoAllTargetsX
- SelectorProjectionOntoAllTargetsY
- ActiveTargetIndex
- ActiveTargetInsideAllTargetsX
- ActiveTargetInsideAllTargetsY
- IsSelectorInsideCollider
- DistanceFromSelectorToAllTargetsOXYPlane

D Questionnaire

Question	Answer type
Mental Demand - Palm	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Mental Demand - PalmWORotation	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Mental Demand - Path	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Physical Demand - Palm	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Physical Demand - PalmWORotation	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Physical Demand - Path	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Temporal Demand - Palm	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Temporal Demand - PalmWORotation	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Temporal Demand - Path	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Performance - Palm	1 - Perfect, 2, 3, 4, 5, 6, 7 - Failure
Performance - PalmWORotation	1 - Perfect, 2, 3, 4, 5, 6, 7 - Failure
Performance - Path	1 - Perfect, 2, 3, 4, 5, 6, 7 - Failure
Effort - Palm	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Effort - PalmWORotation	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Effort - Path	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Frustration - Palm	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Frustration - PalmWORotation	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Frustration - Path	1 - Very Low, 2, 3, 4, 5, 6, 7 - Very High
Please arrange the following reference frames in order of preference (the higher in the list the better)	Sort the three reference frames
What are your overall impressions regarding the reference frames you've experienced?	Textbox

E Source Code

The source code is available on GitHub.