The Cost of Virtuality Switching: Searching for Physical and Virtual Targets in Optical-See-Through Augmented Reality

Kangyou Yu* Yunhao Luo† Radha Kumaran‡ Shane Dirksen§ Misha Sra¶ Tobias Höllerer‖

University of California, Santa Barbara

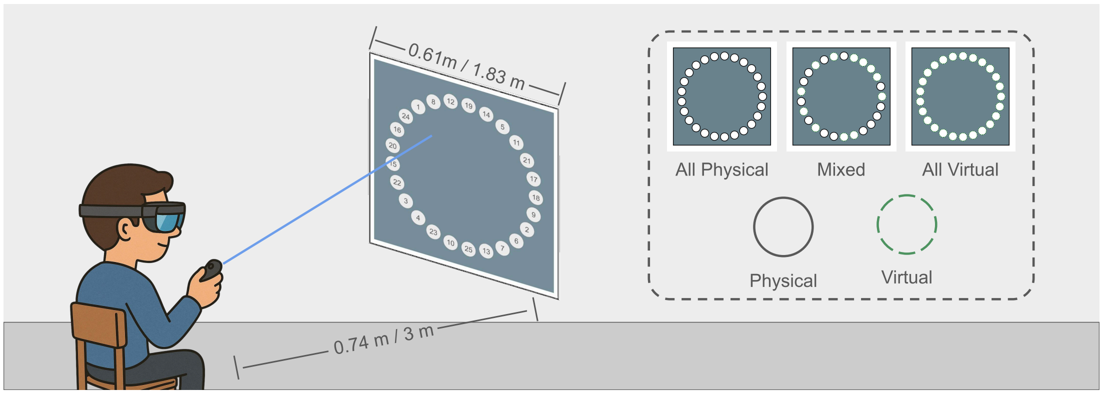


Figure 1: Experimental setup. The subject is positioned at two different distances from the poster on the wall, corresponding to different poster sizes maintaining equal apparent size. Each session involves one of three poster types: all physical, a mix of virtual and physical, or all virtual. *The person in the teaser figure was generated using ChatGPT-4o's image generation capabilities.*

Abstract

As AR applications expand across our daily lives, understanding user interactions within mixed environments—where virtual and physical objects coexist—has become increasingly important. This work investigates human performance and behavior during visual search and selection tasks across three object conditions: (1) virtual objects only, (2) physical objects only, and (3) a combination of virtual and physical objects (Mixed) requiring frequent *virtuality switching*. We also vary the distance to the target plane while maintaining subtended visual angle: a 'near' condition at the headset's focal plane and a 'far' condition at a mid-zone action space distance of 3 meters. Results indicate that, while there are some small effects that can be linked back to established display phenomena such as Vergence-Accommodation Conflict, a main cause for performance differences among the object conditions comes from people adjusting their search and selection behavior to the challenges of virtuality switching, resulting in Mixed conditions requiring significant longer completion times, associated with significantly larger head motion, eye movement, and controller movement. Mixed conditions also resulted in significantly lower accuracy for target selection. Virtual-to-Physical transitions took the longest to complete, followed by Physical-to-Virtual transitions, both significantly longer than transitions to targets within the same virtuality. Participants also reported increased Eye Strain, Fatigue, and Task Load with the Mixed conditions. This work provides insight into the complexities of mixed object interaction and presents quantitative assessments of pronounced virtuality switching, with implications for designing effective AR interfaces.

Keywords: Augmented Reality, Visual Search, Object Selection

1 Introduction

Augmented Reality (AR) and eXtended Reality (XR) technologies have become increasingly used in different domains of our daily lives, such as education [1, 22, 44], healthcare [3, 12, 46], industrial training [10, 41], and entertainment [6, 8]. Several commercially available and research prototype head-mounted displays allow users and researchers to engage with virtual elements overlaid onto the physical environment, both via video-feed-through (e.g. Varjo XR-4 and Aero, Meta Quest Pro and 3, Apple Vision Pro), and optical-see-through head-worn displays (e.g., HoloLens-2, Magic Leap 2, XREAL Air 2 Ultra, Snap Spectacles, Meta Orion).

Considering Physical and Virtual Objects in quick succession in XR is a common occurrence, understanding cognitive and psychophysical demands and behavioral implications of such virtuality switches is important. For example, a user might view a historic building while wearing an AR headset that highlights and explains architectural features like facade ornaments. As the user looks back and forth between the physical structure and the virtual annotations and explanations, various cognitive, perceptual, psychophysical, and even physiological demands play a role in the user's interactions [6, 28].

When users repeatedly shift their attention between physical and virtual elements, cognitive load may increase due to the need for rapid reorientation, perceptual adjustments may become more demanding, and psychophysical stress can accumulate with continued use [7]. Additionally, physiological challenges, particularly Vergence-Accommodation Conflict (VAC), have been widely documented as a significant barrier to effective AR usage, potentially causing user discomfort and decreased task performance [20, 26, 27, 29].

*e-mail: kangyouyu@ucsb.edu
†e-mail:yunhaoluo@ucsb.edu
‡e-mail:rkumaran@ucsb.edu
§e-mail:shanedirksen@ucsb.edu
¶e-mail:sra@cs.ucsb.edu
‖e-mail:holl@cs.ucsb.edu

Recognizing that clear delineation of reality and augmentations can both be a serious practical constraint as well as a design lever for intentional attention guidance, AR designers and practitioners will be interested in the costs of *virtuality switching*, i.e., shifting user focus and attention back and forth between reality and virtual content. In our own informal explorations leading into this work, we noticed that increased occurrence of such switching can lead to changes in AR user behavior, cognitive demand, and perhaps fatigue.

While recent studies have begun to explore performance differences across display modalities and content types—for example, one that investigates visual search with physical and virtual objects in AR and Augmented Virtuality settings [9]—comparatively few have systematically examined how users interact with virtual versus physical objects embedded in the same environment and task structure.

In order to identify the nature and magnitude of the costs in switching attention and focus between objects of different virtualities, we conducted a controlled study involving a search and selection task, which is among the most common and ecologically relevant to human life and behaviors [32, 35].

We employed Optical See-Through (OST) AR in this first controlled study of the cost of virtuality switching because it preserves a natural view of the physical world and OST headsets are well suited for the highly mobile and dynamic AR scenarios where such switching will be particularly frequent.

Our study investigates whether users exhibit different task performance and search behaviors when having to search among physical, virtual, and mixed-reality target sets. Our research aims to answer the following research questions:

RQ1: What is the impact of switching between physical and virtual objects in a task that requires attention/focus in quick succession?

RQ2: What costs of virtuality switches are involved? We are interested in behavioral costs (if and because humans can currently clearly delineate virtual and physical objects and change their scanning/selection behavior in response) as well as subconscious or pre-cognitive costs such as VAC and other causes of eye strain.

RQ3: Do different visual transitions among target types, virtual (v) $\rightarrow$ physical (p) vs. p $\rightarrow$ v impact user performance and experience differently?

To investigate these questions, we conducted a controlled user study (N = 36) following a 2×3 within-subjects design. We varied *Distance* (near vs. far) and search target *Physicality* (all-virtual (AV), all-physical (AP), and half-virtual and half-physical (Mixed)), resulting in six conditions to explore differences in user performance and behavior across object physicalities and interaction distances.

By analyzing performance metrics subjective experience, our findings suggest that effective AR environment design requires a careful balance between virtuality and physicality, as well as the spatial context of user interactions. In particular, we demonstrate:

- Mixed conditions require significant longer completion times, associated with significantly larger head motion, eye movement, and controller movement
- Mixed conditions resulted in significantly lower accuracy for target selection.
- Virtual-to-physical transitions took the longest to complete, followed by physical-to-virtual transitions, and then, at comparable cost, physical-to-physical transitions and virtual-to-virtual transitions.
- In the All-Virtual conditions, trials from the Far viewing distance took longer to complete than from the Near viewing distance.
- There is a significant interaction between Virtuality and Distance regarding Selection Accuracy. In the All-Virtual conditions, we have a higher accuracy in the Near condition than in the Far condition
- Participants exhibit more active and complex eye gaze activity (more glancing over and more dwelling on targets per second) in the Mixed and Far conditions.

Overall, our work presents a controlled study comparing how users perform when interacting with physical, virtual, and mixed objects in the same visual search and selection task. It offers new insights into how spatial distance affects user behavior depending on the type of object being interacted with, shedding light on proximity-related interaction patterns. It identifies specific performance costs associated with pronounced Virtuality Switching, such as longer task completion times and increased movement, suggesting that switching between virtual and physical objects introduces cognitive and perceptual challenges. Finally, it contributes to a deeper understanding of how the “reality" of an object influences user behavior, with implications for designing AR systems that better support seamless interaction across physical and digital elements.

2 Related Work

Existing research highlights significant perceptual differences between physical and virtual content in AR environments, which can affect visual search processes.

2.1 Visual Search Tasks in AR

Virtual environments have been used to replicate experiments from physical settings, revealing comparable outcomes in terms of search speed, accuracy, and cognitive absorption across VR and physical conditions [17]. However, unique factors such as reliance on familiar size cues for object recognition in VR demonstrate the need for strategies tailored to perceptual differences in AR [33]. For example, while physical models excel in conveying object size and spatial accuracy, the scale perception in AR requires careful design to bridge the gap between virtual and real objects [13, 45].

Despite progress in understanding AR visual search, gaps remain regarding simultaneous interactions with physical and virtual targets. Most studies limit their focus to either physical or virtual objects within specific XR manifestations, like AR or AV [40]. Another study has shown that physical objects can be more difficult to recall in their respective environments [24]. A more recent study conducted by Chiossi et al. [9] investigated the difference between physical objects and virtual objects in the AR and AV setting. They found that AV settings and virtual targets can improve user performance on visual search tasks. Building on prior work, our work investigates user performance and behavior when searching and selecting for all physical, all virtual, and mixed physical and virtual targets that coexist at the same time in AR.

2.2 Switching Attention between Virtual and Physical Content

In AR settings where users must shift attention between virtual and physical information sources, context switching has been shown to incur measurable cognitive and performance costs. Early work by Huckauf et al. [21] examined context switching between AR and physical displays in an industrial setting and reported decreased visual performance as a result. More recently, Gabbard et al. [16] found that switching between virtual and real content (even without focal distance changes) significantly reduced user performance and increased visual fatigue, particularly for distant targets.

Eiberger et al. [15] investigated switching between a smartwatch-like display (0.3 m) and a distant, collimated AR display (3.7 m) using a graphical visual search task, and found that both task completion time and error rates increased under combined context and focal switching conditions. Drouot et al. [14] extended this line of research using a Microsoft HoloLens 2 at 1.5–2.0 meter focal distances. While they observed a negative performance impact from focal distance switching, context switching alone did not significantly affect performance in their setup. Arefin et al. [2] then replicated and extended Gabbard's experiment [16]. Using a custom-built AR Haploscope, they found that context switching in AR environments can increase visual fatigue, although it may not significantly impact task performance. In contrast, larger focal distance switches have been shown to both increase eye fatigue and impair user performance [2, 16]. Syiem et al. [37] have also demonstrated that AR interface design and task complexity can induce attentional tunneling, leading users to over-focus on virtual content and reduce awareness of the physical environment.

Beyond AR-specific context switching, prior work by Wolfe et al. [43] found that people exhibited persistent within-type selection behaviors even when multiple target types were present, suggesting an attentional or cognitive cost in switching.

While prior work has provided valuable insights into the cognitive effects of switching between virtual and physical content in abstract symbolic AR or search tasks, our study presents a controlled, but realistic practical task: visual search and selection among physical/virtual targets and transitions (physical-to-virtual and virtual-to-physical) in a shared spatial layout, reflecting real-world AR interactions.

2.3 Vergence-Accomodation Conflict (VAC) in XR

Previous work [30, 31] found that stereo displays are beneficial for depth-related tasks in the near-field with virtual ray casting techniques [25, 39]. However, pointing throughput is typically well below what users can achieve in 2D tasks [36, 38, 39]. One of the likely factors for the lower performance in the pointing and selection task is that targets are at different depths. Teather and Stuerzlinger [39] showed that the different distance of the targets has a huge impact on user performance and experience. With the design of AR and VR glasses, it is common to have this kind of impact on user performance and experience, due to the mismatch of the focal distance and the vergence distance [11]. In a previous study conducted by Batmaz et al. [4], they showed that users will have worse performance and user experience with the varying VAC conditions in VR systems, which means that the target depth will keep changing back and forth to different depths than the constant VAC and no VAC conditions. Gabbard et al. [16] systematically demonstrated that switching focal distances between virtual and physical content impairs both task accuracy and speed, especially when the focal demands are mismatched or placed at optical infinity. Arefin et al. [2] also suggested larger focal plane switches will lead to higher eye strain and fatigue, and decrease human performance.

Although many studies have investigated the VAC problem in AR and VR systems, most focus on interactions with purely virtual content. However, in AR, users often interact with both physical and virtual objects at varying distances. In this work, we investigate how the distance of either virtual or physical objects impacts user performance and experience during interaction in an AR environment.

3 Study Design

The goal of our study is to investigate user performance during a visual search and selection task for physical and virtual targets in optical-see-through AR using a Magic Leap 2 (see Figure 1). We conducted the study with 36 participants using a 2 x 3 within-subjects design. The two main factors in our study were: Distance (near: 0.74m, the headset's focal plane, vs. far: 3m) and Physicality (all virtual vs. all physical vs. mixed), resulting in six conditions. We used a Latin square design to balance the condition orders to minimize the learning effect from the study and mitigate potential order effects.

- **Physicality** refers to the type of target objects that participants were required to search for and select during the task. To investigate the impact of different object types on user performance and experience, we identified three levels of physicality in the study: (1) All Virtual (AV), (2) All Physical (AP), and (3) Mixed. In the **all virtual** condition, all target objects are virtual and rendered by the AR headset. In the **all physical** condition, all targets are real-world physical objects , printed on a paper and affixed to a square poster. In the **mixed** condition, the target set was split between physical and virtual objects: approximately half were physical and visible to the naked eye, while the remaining half were virtual and visible only through the AR headset.
- **Distance** refers to the distance between the target plane and the user's position of sitting. There are two levels of **Distance**: near (0.74m) and far (3m). Based on the official specification of Magic Leap 2, 0.74 m is the comfortable display zone. For the far distance, we selected 3m based on prior research that examined user performance with different input modalities for object selection and manipulation at a distances ranging from 2.4 to 4.9 m [42].

We had six conditions in our study where we manipulate the distance (near, far) and the physicality (AV, AP, Mixed): (1) Near_AV, (2) Near_AP, (3) Near_Mixed, (4) Far_AV, (5) Far_AP, and (6) Far_Mixed. Each condition combined one level of **distance** and one level of **physicality**.

3.1 Mixed Conditions Design

In the two mixed conditions (Near_Mixed and Far_Mixed), we also wanted to investigate how different types of object transitions affect user performance and experience. Therefore, we defined 4 object transition types: **v2v:** Virtual to Virtual, **v2p:** Virtual to Physical, **p2v:** Physical to Virtual, and **p2p:** Physical to Physical.

These transitions allowed us to analyze behavioral patterns related to "virtuality inertia" which we define as the tendency to continue searching within the same virtuality domain before making a switch.

3.2 Participants

We conducted the within-subjects study with 36 participants (M = 22.19, SD = 2.85). Among them, 25 self-identified as male and 11 as female. In order to ensure the quality of eye-tracking data, we only recruited subjects who self-reported normal or corrected-to-normal vision.

3.3 Setup

We implemented the visual search and selection task using Unity 3D (Version 2022.3.21f1 LTS) and presented all conditions through the Magic Leap 2 headset. We used the MagicLeap SDK for the implementation. We used built-in eye tracking to record real-time eye tracking data frame by frame. The study was conducted in a controlled indoor laboratory environment.

We selected a target size that aligns with common practice in AR studies [5]. Specifically, we used the diameter of a standard ping-pong ball (4cm) at the headset's focal distance as a reference for scaling. To enhance experimental control and consistency, we represented (either rendered in AR or printed on paper) all 25 targets as flat discs at the same scale, rather than volumetric spheres. This flat-disc design also minimized potential depth-related confounds and ensured consistent visual presentation from a fixed participant's position.

The targets were evenly arranged along a circular ring. In addition, we chose to have 25 objects in the study to get enough data for four object transitions (25 objects will have 24 transitions (6 times for v2v, v2p, p2v, and pvp each).

In the near-distance conditions, the target size was set to 4 cm in diameter (consistent with the reference scale), the ring radius was 20 cm, and the poster size was 60.96 cm x 60.96 cm. As for the far-distance condition, to eliminate the effect of visual size on user performance and experience, we ensured all objects appeared

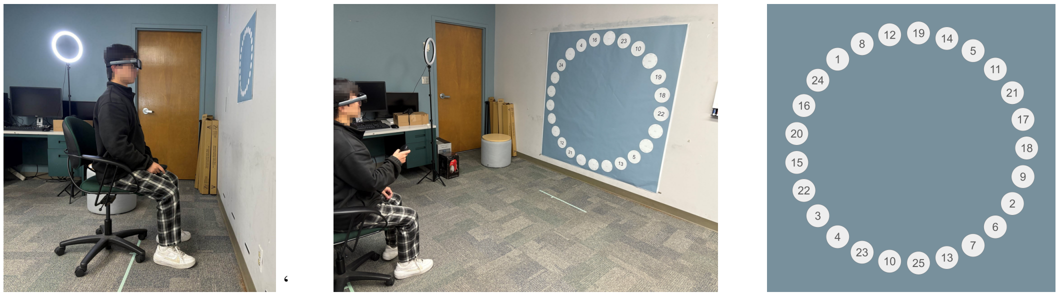


Figure 2: Experimental setup. **Left**: near viewing condition. **Center**: far viewing condition, which also illustrates the mixed condition where half of the objects are physical (discs with physically labeled numbers) and the other half being virtual objects (blank discs whose numbers are rendered through the headset). **Right**: the full ring with 25 numbered target discs arranged in pseudorandom order.

at the same visual size at both near and far distances. Using the near distance as the reference, the target size was adjusted to 16.22 cm in diameter, the ring radius was adjusted to 81.1cm, and the square poster size was adjusted to 182.88 cm x 182.88 cm. Figure 2 shows an example of the layout of a poster we used in the study.

In addition, in all levels of **physicality** (AV, AP, and Mixed), we provided consistent visual feedback for selection using a virtual ray projected from the controller and a virtual outline highlight that appeared on the selected target to indicate correctness. These minimal virtual elements were used uniformly across all conditions, including the AP and Mixed conditions, to maintain consistent interaction mechanics and eliminate feedback-related confounds.

3.4 Task

Subjects performed the same visual search and target selection task in each of the six conditions. The task involved a circular ring (shown in Figure 2), consisting of 25 white, flat, numbered discs arranged in a random order.

In each condition, the ring was placed or displayed at a fixed distance in front of the participant—0.74 m in the near condition and 3 m in the far condition. Each disc was labeled with a unique number in its center, referred to as the target object's "label." The label positions were randomly assigned in each block.

Subjects were instructed to scan the ring to find the target object corresponding to the current trial and aim at their chosen object using the controller's ray cast to make a selection. Once the target object was aligned with the ray cast, then they pressed the Magic Leap 2 controller's trigger to confirm their choice for that trial, and then the subject proceeded to the next trial. Subjects held the controller with their dominant hand, and we encouraged them to respond quickly and accurately.

Moreover, subjects were allowed unlimited attempts in each trial. For example, if they selected an incorrect target or did not move the ray precisely with the correct target before pressing the trigger, the application registered the action as a wrong selection. In such cases, an error sound was played to remind participants of the incorrect selection. Conversely, when subjects successfully selected the correct target, the application highlighted the target's outline in green and played a confirmation sound to indicate a correct selection. The selection ray and outline after successful selection are always rendered in virtually, regardless of condition. Because their behavior is identical in every trial and they appear only after a successful selection, they deliver uniform and minimal feedback with no impact on the search and selection process itself. Thus, the sole variable that differs between conditions is the physicality of the target, not the feedback mechanism. Once the target for a trial was correctly selected, subjects were required to immediately proceed to the next trial.

The search and selection followed a sequential numerical order, from 1 to 25. In other words, in the first trial of each block, subjects were required to search for and select the disc labeled "1" by pointing the controller ray at it and pressing the trigger, thereby completing the first trial. In the second trial, the target was the disc labeled "2," and so on, until all 25 trials were completed in order.

3.5 Study Procedure

Upon arrival, subjects were introduced to the study and asked to sign the consent form (protocol #anonymous) and complete the demographic questionnaire to provide basic information if agreed to participate. They were then given a brief tutorial on how to wear the headset and use the controller. In order to reduce the learning effect of participants from the study, a training session was conducted for them to get familiar with the study process, how to perform the task, and how to calibrate the virtual objects to align with the physical objects. During calibration, participants manually aligned unlabeled virtual duplicates of the 25 discs to the physical reference targets on the wall-mounted poster using a keyboard. This step ensured visual consistency across all conditions—including the Mixed condition, where both physical and virtual targets were present—and allowed the highlight outline of targets to function accurately without disrupting the user experience. After the training session, they started the main study if they felt confident enough to perform the task.

In each condition in the main study, subjects first completed a calibration step to ensure that virtual objects aligned with the positions of physical objects on the poster from their visual angle. After calibration, they were able to start the task. After completing task in each condition, participants were required to fill out a post-condition questionnaire.

After completing six conditions and six post-condition questionnaires, they were required to complete a post-study questionnaire. We compensated them at the rate of $20 per hour. In general, completing the entire study took approximately 1 hour for each subject. The entire work flow of the study is shown in Figure 3

3.6 Measures

To evaluate the effects of object physicality, distance, and transition type, we collected both objective performance data and subjective user feedback across all experimental conditions.

3.6.1 Objective Measures

Task performance was evaluated using a set of metrics commonly applied in 3D selection and manipulation studies in AR and VR [5], these included task completion time (in seconds), measured as the total duration required to complete all selections within a condition, and selection deviation (in centimeters), defined as the Euclidean distance between the center of the intended target and the controller's selection point. We also recorded head and controller movement (in meters), quantified as the total translational distance traveled, and head and controller rotation (in degrees), measured as the cumulative angular displacement.

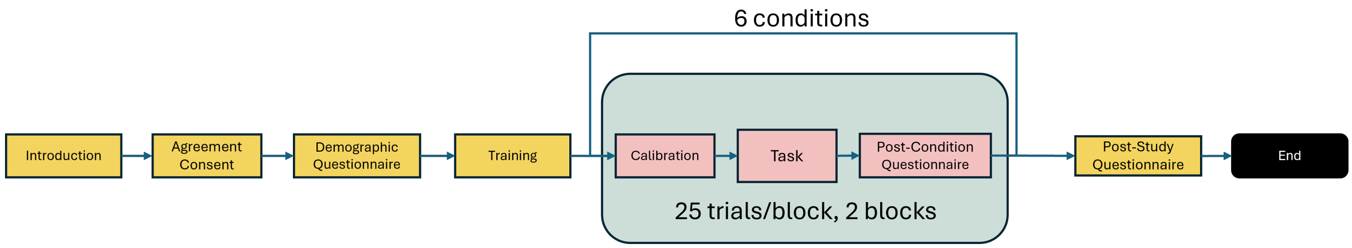


Figure 3: Overall procedure of the user study. Before starting the study, participants first completed a demographic questionnaire and signed a consent form, followed by a training session. During the main study, each participant experienced all six conditions (Near_AV, Near_Mixed, Near_AP, Far_AV, Far_Mixed, and Far_AP). For each condition, they completed two blocks of tasks and then filled out a post-condition questionnaire. After completing all six conditions, participants were asked to complete a final post-study questionnaire.

3.6.2 Subjective Measures

To better understand the user experience, we collected and evaluated subjective feedback for all study conditions using post-condition and post-study questionnaires.

The post-condition questionnaire consists of three parts, (1) the complete NASA-TLX, (2) subscale of Simulator Sickness Questionnaire (SSQ), and (3) customized questions.

We used the NASA-TLX [18] specifically to assess participant perceived workload for each condition. To evaluate motion sickness symptoms, we adapted items from the Simulator Sickness Questionnaire (SSQ) [23], specifically targeting symptoms relevant to our AR context, such as fatigue and eye strain. In addition, we administered a set of custom questions designed to capture participant subjective impressions and experience of the experiment setup and design.

In the post-study questionnaire, we asked subjects to rank the six conditions from most to least preferred based on their overall preference (according to the difficulty of the conditions) and we asked their experience about the different distances and different object-type arrangements in six conditions.

3.7 Eye Tracking and Dwelling

The Magic Leap 2 eye tracker provides eye fixation data in 3D coordinates relative to the Unity world coordinate system. To determine each eye fixation point on the target plane, we calculated the intersection between the line extending from the midpoint between the eyes to the 3D fixation point and the target plane. To filter out unreliable eye-tracking-data, we calculated the average eye tracking confidence provided by Magic Leap 2 between two selections. If the average confidence was less than 0.5 (on a scale from 0 to 1), the corresponding eye-tracking data was discarded. To capture participant eye dwelling patterns, we calculated the shortest distance between each fixation point and the centers of all targets. Due to potential inaccuracies from several sources, including the eye tracker's fixation error (approximately 1 degree of visual angle[1]), the tracking position of the virtual discs, and the possibility that participants were using their peripheral vision during target scanning, we treated a fixation as landing on a target when that distance was less than three times the target's radius. To distinguish true dwelling from rapid scanning across multiple targets, we applied a threshold of 10 consecutive frames on the same target, which was then classified as a dwelling event. The Magic Leap 2 eye tracking samples at 60 frames per second, so 10 consecutive frames is around 0.17 seconds–within the 100-200 ms fixation window identified in prior research [34]. To minimize the influence of eye-tracking noise on the results, eye-tracking analyses were conducted on a subset of participants whose eye tracking data exhibited minimal noise. To ensure balanced condition orders, we randomly selected a subset from this low-noise group with evenly distributed condition sequences, resulting in a final analysis sample of 12 participants. Due to participants adjusting the AR glasses between conditions, systematic shifts or scaling distortions in eye tracking data may occur. To correct for these, we visually inspected the overlaid trajectories and applied consistent translation and scaling adjustment when necessary to realign the data to their correct positions.

[1]https://developer-docs.magicleap.cloud/docs/guides/features/eye-tracking

4 Results

A Shapiro-Wilk test was used to test the normality distribution of our data. As for the performance measurements, we employ a Generalized Linear Mixed Model (GLMM) with a Gamma distribution and log link to investigate differences in the performance measures. To present the model results in a more interpretable and familiar format, we reported ANOVA-style summaries of the GLMMs using Ttpe III Wald chi-square test via car package in R. For subjective experience measures, We applied the Friedman test and the pairwise Wilcoxon test if the data are not normally distributed. Otherwise, we performed a two-way RM ANOVA to evaluate the differences on all measures.

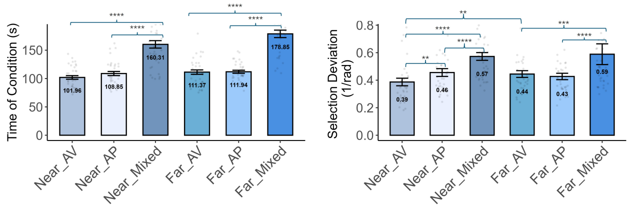
Here we only discuss significant results in detail. A p-value of 0.05 was used for all statistical analysis.

4.1 Overall Performance Measurements

All performance measurements were first collected for each participant across all six conditions. To control for the effect of spatial distance on performance, we calculated the subtended angle between between successive target selections, we then normalized the time metric using the logarithm of the subtended angle (log(angle + 1)), while other measures were normalized by the subtended angle itself. Unless otherwise stated, the results reported in this section reflect post-normalization values.

- Task completion time (s). We first analyzed the overall task completion time for each condition, see Figure 4a. The model revealed a significant main effect of **Physicality** ($\chi^2(2) = 233.4299$, $p < 0.0001$), indicating that physicality type influenced task completion time. There was no significant main effect of **Distance** ($\chi^2(1) = 1.1773$, $p = 0.2779$), nor a significant interaction effect between Distance and Physicality. Post-hoc pairwise comparisons for Physicality showed Mixed conditions took significantly more time than AV ($p < 0.0001$) and AP ($p < 0.0001$) conditions.
- Selection deviation (1/rad). We analyzed the overall selection deviation in conditions, see Figure 4b. There is a significant main effect of **Physicality** on selection accuracy ($\chi^2(2) = 28.2638$, $p < 0.0001$). There is also a significant interaction effect between **Distance** and **Physicality** ($\chi^2(2) = 9.0420$, $p = 0.0109$). Post-hoc comparisons with Physicality showed that participants were significantly less accurate under the Far_AV condition than under the Near_AV ($p < 0.005$). Post-hoc comparisons with Distance showed there are significant differences between the Mixed condition and AV ($p < 0.0001$)

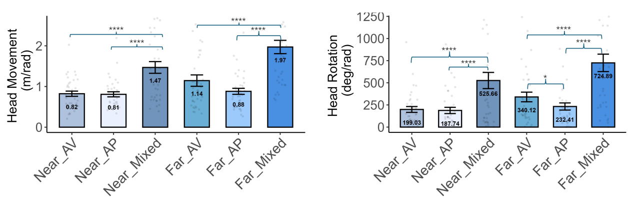
and AP (p < 0.0001) conditions at far distance. In addition, Mixed condition is significantly less accurate than AV (p < 0.0001) and AP (p < 0.0001) at near distance.



(a) Total time spent for each condition. (b) Average selection deviation across conditions.

Figure 4: Task completion time and average selection deviation across conditions.

- Head movement (m/rad). Results revealed a significant main effect of **Physicality** on HeadMovement ($\chi^2(2) = 83.9241$, $p < 0.0001$), see Figure 5a. There was no significant interaction effect between Distance and Physicality. Post-hoc pairwise comparison for Physicality showed that the Mixed condition led to significantly greater head movement compared to AV (p < 0.0001) and AP (p < 0.0001) conditions.
- Head rotation (deg/rad). The results revealed a significant main effect of **Physicality** on head rotation ($\chi^2(2) = 78.6813$, $p < 0.0001$). There was no significant interaction effect between Distance and Physicality. Post-hoc comparisons for physicality showed that Mixed conditions led to significantly more head rotation than AV (p < 0.001) and AP (p < 0.001) conditions at both near and far distance, see Figure 5b.

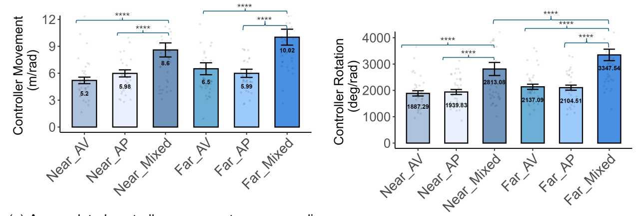


(a) Accumulated head movement across conditions. (b) Accumulated head rotation across conditions.

Figure 5: Accumulated head pose across conditions.

- Controller movement (m/rad). There was a significant main effect of **Physicality** on controller movement ($\chi^2(2) = 78.6810$, $p < 0.0001$), and no interaction effect was found. Post-hoc comparisons for Physicality revealed that participants in the Mixed condition showed significantly greater controller movement than in the: AV (p < 0.0001) and AP (p < 0.0001) conditions, see Figure 6a.
- Controller rotation (deg/rad). We found a significant main effect of **Distance** ($\chi^2(1) = 3.8603$, $p < 0.05$) on controller rotation. There was also a significant main effect of **Physicality** ($\chi^2(2) = 113.4051$, $p < 0.0001$). After post-hoc pairwise comparison for Distance, it showed that the far distance resulted in more controller rotation than Near distance (p < 0.0001). Post-hoc comparisons for Physicality showed that participants exhibited significantly more controller rotation under the mixed condition than under the AV (p < 0.0001), and AP (p < 0.0001) conditions, see Figure 6b.

To investigate potential learning effects across the experiment, we analyzed both block order (across 12 blocks) and condition order (across 6 conditions) on key performance metrics, including task completion time, selection deviation, head movement, head rotation, controller movement, and controller rotation. Results revealed no significant effects of block order or condition order on any of the performance metrics (all p > .15), indicating that participants' performance was not systematically influenced by the order in which



(a) Accumulated controller movement across conditions. (b) Accumulated controller rotation across conditions.

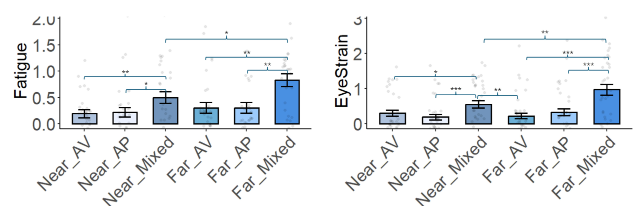
Figure 6: Accumulated controller pose across conditions.

they experienced the conditions or blocks. Full statistical results are included in Appendix A.

4.2 Subjective Measurements

We only report the significant and most relevant results here, more statistical results can be found in the Appendix.

- Fatigue. Fatigue was measured using the selected item (Fatigue) from the SSQ, rated on a 4-point Likert scale. Figure 7a shows the overall level of perceived fatigue in each condition during the study. Friedman test results show that there is a statistically significant difference in fatigue depending on which condition was running, $\chi^2(5) = 31.00498, p < 0.001$. Post-hoc analysis showed the fatigue level is significantly higher at Near_Mixed condition than Near_AV (p = 0.008) and Near_AP (p = 0.014) conditions, perceived fatigue level is also significant higher at Far_Mixed than Far_AV (p < 0.005) and Far_AP (p < 0.005).
- Eye strain. Eye strain was measured using the selected item (Eye strain) from the SSQ, rated on a 4-point Likert scale, see Figure 7b. Friedman test showed that there is significant difference in Eye strain on conditions ($\chi^2(36) = 44.18261, p < 0.0001$). The post-hoc results revealed that participants perceived higher level of eye strain at Near_Mixed condition than Near_AV (p < 0.05) and Near_AP (p < 0.01) conditions. Similarly, participants will have higher eye strain at Far_Mixed condition than Far_AV (p < 0.0005) and Far_AP (p < 0.0005) conditions.

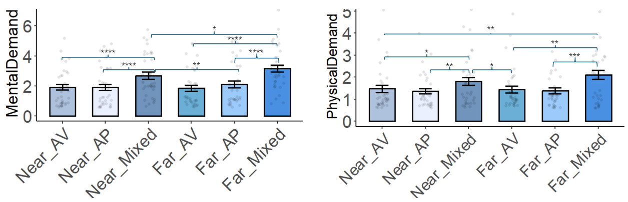


(a) Perceived level of fatigue across conditions (b) Perceived level of eye strain across conditions

Figure 7: Participants' perceived level of (a) fatigue and (b) eye strain across conditions.

- Mental demand. Mental demand was measured by the NASA-TLX with a 7-point Likert question, see Figure 8a. Friedman test results show that there is a statistically significant difference in fatigue depending on which condition was running, $\chi^2(5) = 75.14706, p < 0.001$. Post-hoc analysis showed the mental demand level is significantly higher at Near_Mixed condition than Near_AV (p = 0.008) and Near_AP (p < 0.0001) conditions, perceived mental demand is also significant higher at Far_Mixed than Far_AV (p < 0.0001) and Far_AP (p < 0.0001). In addition, participants had higher mental load at Far_Mixed condition than Near_Mixed condition (p < 0.05).
- Physical demand. Physical demand was measured by the NASA-TLX with a 7-point Likert question, see Figure 8b. Friedman test results show that there is a statistically significant difference in fatigue depending on which condition

was running, $\chi^2(5) = 35.66055, p < 0.001$. Post-hoc analysis showed the physical demand level is significantly higher at Near_Mixed condition than Near_AV (p < 0.05) and Near_AP (p < 0.01) conditions, perceived physical demand is also significant higher at Far_Mixed than Far_AV (p < 0.005) and Far_AP (p < 0.001).



(a) Perceived level of mental demand in conditions (b) Perceived level of physical demand in conditions

Figure 8: Participants' perceived level of (a) mental demand and (b) physical demand across conditions.

- Condition preference. Rankings were converted into preference scores using an inverse scoring method: the most preferred condition received a score of 6, the second 5, and so on, down to a score of 1 for the least preferred condition. Figure 9 shows the general preference of 36 participants for six conditions. Apparently, participants like the Near_AV condition the most, which means they think Near_AV is the easiest condition (because they were asked to rank preference based on difficulty), then the Far_AV, Near_AP, Far_AP, Near_Mixed and the Far_Mixed condition was ranked as the hardest condition.

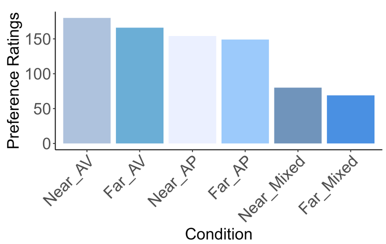
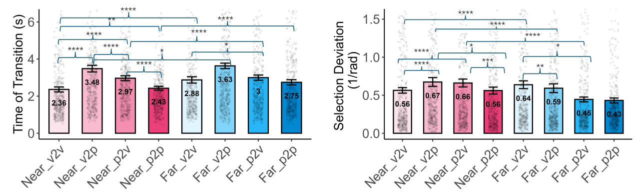


Figure 9: Overall preference rating for all conditions.

4.3 Performance Measurements in Transitions

Similarly, all performance measurements were first collected for each selection, then normalized by subtended angle method mentioned in Section 4.1. While our study recorded both head-based (movement and rotation) and controller-based (movement and rotation) data, we chose to report only controller-related metrics here, since we examined eye-tracking data, which already provides insight into the head orientation and rotation.

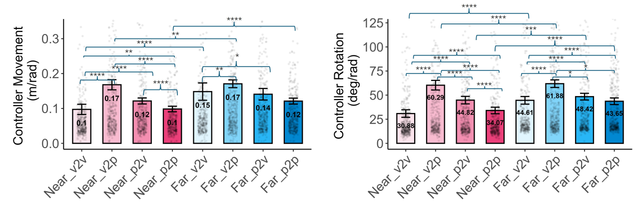
- Time per selection (s). We analyze the time for each selection under mixed conditions; see Figure 10a. The results showed a significant main effect of **Distance** on timer per selection ($\chi^2(1) = 27.9615, p < 0.001$), there was also a significant main effect **Transition** on time per selection $\chi^2(3) = 9.2847, p < 0.05$. Furthermore, a significant interaction was found between distance and transition ($\chi^2(3) = 54.6698, p < 0.0001$). Post-hoc comparisons showed that participants took significantly more time at far distance than near distance on p2p (p < 0.0001), v2p (p < 0.05), and v2v (p < 0.0001) transitions. It also showed that v2p transition will take significantly longer time than v2v (p < 0.0001) and p2p (p < 0.0001) at near distance.
- Selection deviation (1/rad). The results showed a significant main effect of **Transition** on selection deviation ($\chi^2(3) = 12.7364, p < 0.0001$). Additionally, a significant interaction effect between distance and transition was found ($\chi^2(3) = 58.3612, p < 0.0001$), see Figure 10b. Post-hoc tests showed significantly higher selection deviation at the far distance than near distance on p2v (p < 0.0001), v2v (p < 0.0001), and v2p (p < 0.0001) transitions. In addition, p2v (p < 0.05) and v2p (p < 0.01) transitions resulted in higher selection deviation than v2v at far distance.



(a) Average time for each transition type (b) Selection deviation for each transition type

Figure 10: Selection time and deviation across transition types.

- Controller movement (m/rad). The results showed a significant main effect of **Distance** on controller movement ($\chi^2(1) = 44.6987, p < 0.0001$). Additionally, a significant interaction effect between distance and transition was found ($\chi^2(3) = 75.3138, p < 0.0001$), see Fiigure 11a. Post-hoc tests showed significantly greater controller movement at the far distance than near distance on p2p (p < 0.0001), v2p (p < 0.01) and v2v (p < 0.0001) transitions. In addition, p2v and v2p transitions result in significantly higher controller movement than p2p (p < 0.0001) and v2v (p < 0.0001) transitions at near distance.
- Controller rotation (deg/rad). The results showed a significant main effect of **Distance** on controller rotation ($\chi^2(1) = 2340.2719, p < 0.0001$). There was also a main effect of **Transition** ($\chi^2(3) = 243.9881, p < 0.0001$). Additionally, a significant interaction effect between distance and transition was found ($\chi^2(3) = 4735.4299, p < 0.0001$) , see Figure 11b. Post-hoc tests showed participants had significantly more controller rotation at the far distance than near distance on p2p (p < 0.001), v2p (p < 0.001), p2v (p < 0.005) and v2v (p < 0.001) transitions. In addition, p2v and v2p transitions result in significantly higher controller rotation than p2p (p2v > p2p: p < 0.0001, v2p > p2p: p < 0.0001) and v2v (p2v > v2v: p < 0.0001, v2p > v2v: p < 0.0001) transitions at both near and far distances.



(a) Accumulated controller movement across transition types. (b) Accumulated controller rotation across transition types

Figure 11: Accumulated controller pose across condition types.

4.4 Eye Tracking and Dwelling

Saccade length Saccade lengths were measured by the total travel distance of eye-tracking fixations. An ANOVA style GLMM analysis reveals a significant main effect of condition on distance ($\chi^2(5) = 70.28$, $p < 0.0001$) and physicality ($\chi^2(5) = 7.79$, $p = 0.02$). Saccade lengths in mixed conditions were significantly greater than those in both AV and AP conditions at the same distance.

Scan Amount A Friedman test and pairwise Wilcoxon test show that in near conditions, participants' scans (see Sec. 3.7) touched on objects more times in Near_Mixed than both Near_AV ($p = 0.02$) and Near_AP ($p = 0.02$). In the far conditions, participants' scans in Far_Mixed touched on objects more times than in

Far_AV ($p = 0.003$) and Far_AP ($p = 0.02$). Participants' scans in Far_Mixed also touched on objects more times than in Near_AP ($p < 0.001$) and Near_AV ($p < 0.001$).

See Appendix C for more information, charts, and trajectory visualizations.

4.5 Self-reported Subjective Feedback

To gain qualitative insights, we designed questions on participants' subjective experience across different main factors, overall conditions and experiment design. The feedback provided additional perspectives on search strategies/behaviors and the study design.

Search strategies and behaviors. We designed a question to ask about behavior and search strategies during the tasks. Seven participants mentioned that they preferred to search for one type of object first, then go through the other type until they found the target. For example, P25 stated: "I would scan the virtual ones first if the previous number was virtual, because those were more top of mind, and then I would look at the real ones", P27 also said something similar: "I scanned through either physical or virtual circles first before switching to the latter". One participant (P26) said they kept doing a linear search every time: "I ended up doing a linear search every time I couldn't find a previous number, which still caused me to miss numbers during that linear search". However, there are also some participants (P33, 34, 35) who mentioned that they tried to remember other targets positions that were close to the current target.

Experience of Distance. As the **Distance** is one of the main factors in our study, so we asked participants about their experience of tasks at different distances. Many of the participants stated that they preferred the near distance over the far distance, because the far distance would made the target harder to search and select. For instance, P5 said "When I'm in far distance, it is harder than near to find the targets." P20 remarked, "When I was further away, I felt it was harder to find the next number." P23 also commented that near distance will improve the searching: "Closer distance is better. I can scan through the circle faster." However, there are some participants who said they liked the far distance conditions better because they perceived a lower eye strain and it was easier for them to find the object from the far distance.

Experience of Physicality. As the other main factor that we want to investigate in this study, we asked participants how they liked the different physicalities and different transitions in the mixed condition. P20 said "The Mixed condition was the hardest for me because my mind was going back and forth between physical and virtual numbers. Virtual felt the easiest because I was wearing a virtual headset so it was easier to see than the physical numbers." P25 thought the mixed condition was more challenging: "The all-physical targets were by far the easiest to complete. The mixed condition was more challenging because it was harder to find the next number." P26 also said the mixed condition is the hardest: "I think the mixed was easily the hardest since I had to switch my brain a fair amount. The virtual was nice since it was very lined up with the rendered graphics, but the physical was nicer to look at since it was actually real." P30 perceived more mental load in mixed condition: "I definitely had to pay more mental cost when trying to switch between physical and virtual numbers. Some of the virtual numbers did not appear as clearly as the physical numbers, which also required additional effort."

5 Discussion

Our study investigated the impact of distance (*near* and *far*) and object type (*all virtual*, *all physical* and *mixed*) on user performance and experience in augmented reality (AR) environments. A key focus point was the cost of switching back and forth between virtual and physical objects.

Before our studies, we had formed the following hypotheses:

H1 (regarding RQ1): Mixed conditions that involve switching (rapidly) among virtual and physical objects will result in longer task times and more required participant effort.

H2 (regarding RQ2): VAC should be a measurable cost (we designed the two distance conditions to partially control for this).

H3 (regarding RQ2): We expected behavioral components leading to increased cost when a lot of *virtuality switching* was required.

H4 (regarding RQ3): Transitions from Virtual to Physical and from Physical to Virtual will both be more costly than staying within one virtuality.

The results provide several insights into how these factors influence user interaction and highlight areas for further exploration.

Impact of Physicality (Object Type) on User Performance and Experience Our first research question (RQ1) explores the influence of object type (all virtual, all physical, and mixed) on user performance and experience. The results revealed significant differences between these conditions, with the AV condition and AP condition outperforming the mixed conditions in terms of time, accuracy and perceived fatigue. While AV condition also showed slightly better performance than the AP condition on several measurements, these differences were not statistically significant. These findings align with previous work suggesting that virtual objects, being more visually distinct and predictable, facilitate efficient search and selection [9,17,24].

In contrast, mixed conditions were consistently rated the most challenging, particularly at far distances. This difficulty likely arises from the need to distinguish and transition between different types of objects, such as virtual to physical objects and physical to virtual objects, which imposes higher cognitive and motor demands. Such transitions disrupt visual and motor coordination, increasing task complexity and perceived workload. These findings extend previous research on the challenges of simultaneous interaction with virtual and physical targets [19,40] by highlighting the compounded difficulty when these interactions are performed at varying distances.

Interestingly, subjective feedback indicated that participants found virtual objects more engaging and easier to interact with than physical objects, which aligns with the finding of Kim et al. [24]. In contrast, physical objects would be affected by many real-world constraints such as lighting variability and physical imperfections, introduce additional complexity that can hinder performance. These insights suggest a trade-off between physical fidelity and interaction efficiency, with virtual objects offering advantages in environments where speed and accuracy are critical.

Overall, the results provided strong evidence confirming H1.

Impact of Vergence-Accommodation Conflict (VAC) on User Performance and Experience As part of our second research question (RQ2), we investigated the costs of virtuality switching and whether distance (near vs. far) impacts user performance and experience during visual search and selection tasks. The results showed a clear impact of distance on task performance metrics, including time, accuracy, and subjective measures such as fatigue and search difficulty. Consistent with prior findings that tasks at farther distances impose higher cognitive and motor demands due to reduced spatial resolution and depth cue challenges [11,39], participants performed significantly better in near-distance conditions.

Our study design purposefully placed the Near conditions at the focal distance of the XR headset, which eliminated VAC effects as a contributing factor for these conditions, whereas the Far conditions were done at a distance that would trigger VAC.

Overall, the impact of VAC emerged in the results of our study, for example, by the Near_AV condition consistently outperforming the Far_AV condition in terms of selection accuracy and Near_AV reached less completion time than the Far_AV condition, highlighting the challenges users face when interacting with distant objects in AR environments, which also aligns with the findings from a previous paper on VAC related challenges [4]. VAC issues becomes more pronounced at greater distances, adding to the cognitive load and contributing to lower performance.

Subjective feedback further supports these findings, with participants reporting higher levels of fatigue and difficulty concentrating in far-distance tasks, particularly in mixed-object settings. These results align with prior research emphasizing the role of VAC in AR and VR systems, which can adversely affect depth perception and interaction efficiency [4, 30].

Our findings also suggest that the challenges of interacting with far-distance targets in AR environments are exacerbated when mixed object types are present. Mixed conditions inherently require rapid switching between virtual and physical objects, increasing cognitive load and disrupting task flow. These results underscore the importance of designing AR systems with careful consideration of target distance and object types, particularly in scenarios requiring precise and frequent interactions.

However, VAC was clearly not a dominant factor compared to the large time costs of the Mixed Conditions overall. So, H2, while reasonably confirmed, is also downplayed in terms of importance.

Behavior Modifications Regarding RQ3, let's examine possible main causes for the large performance penalties for our Mixed conditions:

We posit two (non-exclusive) explanations why transitions involving Virtuality Switching were more costly than within-virtuality target transitions (confirming H4). Both of these are behavior adaptations (providing some evidence, but no firm confirmation of, H3):

First, we noticed from eye tracking playback and analysis that people dwelled (paused) more when they were forced to look back and forth between virtual and physical items. Moreover, in far conditions, participants generally dwelled on more distinct targets per second compared to near conditions. Within both near and far conditions, the highest number of dwells on distinct targets per second were from mixed conditions. This increased dwelling time suggests that the presence of both virtual and physical objects in this search task prompted participants to scan at a faster pace. However, this higher scan speed did not lead to faster completion of the search; on the contrary, participants generally took longer to complete selections in mixed conditions.

A second main cause for performance differences among the object conditions appeared to come from people's tendency in mixed-virtuality search tasks to continue searching for the next item in the same virtuality domain: if the current item is virtual, people tend to check the virtual search space before switching attention to physical candidates and vice versa, which aligns with the finding of Wolfe et al. [43]. We'd like to refer to this phenomenon as *Virtuality Inertia*.

Eye Scan Pattern Between Selections We found some indications for this behavior from our eye tracking data: Regarding encountered Virtual and Physical targets (objects that were "touched by" eye gaze): In mixed conditions, participants' scanning behavior involved significantly more target touches compared to both all virtual and all physical conditions. Similarly, participants' scanning trajectories were significantly longer–measured by saccade length–in mixed condition than all virtual and all physical conditions at the same distance. In other words, participants re- and over-scanned the targets are simultaneously present. Overall, we witnessed more active eye gaze activity (more glancing over and more dwelling on targets per second) in the Mixed conditions.

Our data indicate that in mixed conditions at far distances, participants scanned more targets compared to other far conditions. A higher level of scanning effort required in the mixed conditions implies that the switches between physicality led to more scanning before successfully identifying and select the correct target.

6 Limitations and Future Work

This study serves as a foundational step toward understanding how users interact with both physical and virtual targets in OST AR. We focused on a core interaction task, visual search and selection using a controller-based input method. While our study successfully manipulated different types of target objects, several limitations remain that point to directions for future work.

Despite our best efforts to make the physical objects closely resemble their virtual counterparts, slight differences remained due to hardware constraints of the Magic Leap 2 headset. In addition, the simplified task environment, designed to maximize experimental control, does not fully reflect the complexity of real-world scenarios. Although the experiment was conducted in a controlled laboratory setting, we were unable to perfectly replicate the ambient lighting and brightness consistently across all participants, which may have influenced participant perceptions of the objects. Furthermore, the task itself was intentionally simple, requiring participants to use a controller to select target objects. Although this approach minimized complexity, it may have exaggerated user perception of the differences between virtual and physical objects.

Future work will extend this investigation to include more natural and diverse interaction techniques, such as hand tracking, gaze-based selection, and multimodal inputs (e.g., gaze+pinch), to better understand their impact on user experience in hybrid environments. The search task itself could be expanded from 2D targets to 3D objects or mixed 2D/3D scenes. Increasing the task and environmental complexity could improve ecological validity and better reflect real-world AR use cases. Moreover, we believe there is significant value in replicating the similar experiment on other AR platforms, such as mobile AR and projection-based AR. Future work can assess the generalizability of our findings across devices and uncover potential platform-specific differences. These efforts will help bridge the gap between controlled experimental insights and applied AR system design.

7 Conclusion

In this study, we investigated user performance and behavior in visual search and target selection tasks across different target types (all virtual, all physical, and mixed) and distances (near vs. far) within AR environments. By employing a systematic experimental design, we evaluated key performance metrics such as time, selection deviation and head movement, along with subjective measures such as perceived fatigue and eye strain. Our results showed that the target type and distance significantly affect performance and experience. Tasks with all virtual objects consistently outperformed those with all physical or mixed objects, particularly at far distances. The role of distance was equally critical, with near-distance tasks outperforming far-distance tasks across nearly all performance and subjective measures. This work provided some of the first detailed quantification for the cost of switching between physical and virtual targets, demonstrating the strong impact of behavior modifications (here: regarding the scanning of upcoming targets) based on perceived differences in virtual and physical object appearance, reporting increased scanning and dwelling and *virtuality inertia* as example reactions to being faced with the challenge of back-and-forth focus and attention switching between virtual and physical targets. It also demonstrated a small detrimental effect of VAC on selection accuracy and user comfort. These results are a first step towards more comprehensive modeling of user cognitive and psychophysiological phenomena associated with *Virtuality Switching*.

Acknowledgments

Support for this work was provided in part by ONR grant N00014-23-1-2118 as well as NSF grant IIS-2211784.

REFERENCES

[1] M. Akçayır and G. Akçayır. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, 20:1–11, 2017. doi: 10.1016/j.edurev.2016.11.002

[2] M. S. Arefin, N. Phillips, A. Plopski, J. L. Gabbard, and J. E. Swan. The effect of context switching, focal switching distance, binocular and monocular viewing, and transient focal blur on human performance in optical see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2014–2025, 2022. doi: 10.1109/TVCG.2022.3150503

[3] E. Z. Barsom, M. Graafland, and M. P. Schijven. Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical endoscopy*, 30:4174–4183, 2016.

[4] A. U. Batmaz, R. Turkmen, M. Sarac, M. D. Barrera Machuca, and W. Stuerzlinger. Re-investigating the effect of the vergence-accommodation conflict on 3d pointing. In *Proceedings of the 29th ACM symposium on virtual reality software and technology*, pp. 1–10, 2023.

[5] J. Bergström, T.-S. Dalsgaard, J. Alexander, and K. Hornbæk. How to evaluate object selection and manipulation in vr? guidelines from 20 years of studies. In *proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–20, 2021.

[6] M. Billinghurst, A. Clark, and G. Lee. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015. doi: 10.1561/1100000049

[7] J. Buchner, K. Buntins, and M. Kerres. The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, 38(1):285–303, 2022.

[8] A. D. Cheok, K. H. Goh, W. Liu, F. Farbiz, S. L. Teo, H. S. Teo, S. P. Lee, Y. Li, S. W. Fong, and X. Yang. Human pacman: a mobile wide-area entertainment system based on physical, social, and ubiquitous computing. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '04, 2 pages, p. 360–361. Association for Computing Machinery, New York, NY, USA, 2004. doi: 10.1145/1067343.1067402

[9] F. Chiossi, I. Trautmannsheimer, C. Ou, U. Gruenefeld, and S. Mayer. Searching across realities: Investigating erps and eye-tracking correlates of visual search in mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 2024.

[10] S. Choi, K. Jung, and S. D. Noh. Virtual reality applications in manufacturing industries: Past research, present findings, and future directions. *Concurrent Engineering*, 23(1):40–63, 2015. doi: 10.1177/1063293X14568814

[11] T. Clark. A review of the technology and consumer opinion relating to stereoscopic 3d. *Displays*, 2014.

[12] C. J. Dede, J. Jacobson, and J. Richards. *Introduction: Virtual, augmented, and mixed realities in education*. Springer, 2017.

[13] T. Drey, M. Montag, A. Vogt, N. Rixen, T. Seufert, S. Zander, M. Rietzler, and E. Rukzio. Investigating the effects of individual spatial abilities on virtual reality object manipulation. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–24, 2023.

[14] M. Drouot, N. Le Bigot, J.-L. de Bougrenet, and V. Nourrit. Effect of context and distance switching on visual performances in augmented reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 476–477, 2021. doi: 10.1109/VRW52623.2021.00120

[15] A. Eiberger, P. O. Kristensson, S. Mayr, M. Kranz, and J. Grubert. Effects of depth layer switching between an optical see-through head-mounted display and a body-proximate display. In *Symposium on Spatial User Interaction*, SUI '19, article no. 15, 9 pages. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3357251.3357588

[16] J. L. Gabbard, D. G. Mehra, and J. E. Swan. Effects of ar display context switching and focal distance switching on human performance. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2228–2241, 2019. doi: 10.1109/TVCG.2018.2832633

[17] J. Hadnett-Hunter, E. O'Neill, and M. J. Proulx. Contributed session ii: Visual search in virtual reality (vsvr): A visual search toolbox for virtual reality. *Journal of Vision*, 22(3):19–19, 2022.

[18] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. doi: 10.1177/154193120605000909

[19] A. Hettiarachchi and D. Wigdor. Annexing reality: Enabling opportunistic use of everyday objects as tangible proxies in augmented reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1957–1967, 2016.

[20] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):33–33, 2008.

[21] A. Huckauf, M. H. Urbina, J. Grubert, I. Böckelmann, F. Doil, L. Schega, J. Tümler, and R. Mecke. Perceptual issues in optical-see-through displays. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, APGV '10, 8 pages, p. 41–48. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1836248.1836255

[22] M.-B. Ibáñez and C. Delgado-Kloos. Augmented reality for stem learning: A systematic review. *Computers Education*, 123:109–123, 2018. doi: 10.1016/j.compedu.2018.05.002

[23] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[24] Y.-J. Kim, R. Kumaran, E. Sayyad, A. Milner, T. Bullock, B. Giesbrecht, and T. Höllerer. Investigating search among physical and virtual objects under different lighting conditions. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3788–3798, 2022.

[25] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, 68(10):603–615, 2010.

[26] G.-A. Koulieris, B. Bui, M. S. Banks, and G. Drettakis. Accommodation and comfort in head-mounted displays. *ACM Transactions on Graphics (TOG)*, 36(4):1–11, 2017.

[27] G. Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics*, 22(7):1912–1931, 2015.

[28] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE international symposium on mixed and augmented reality*, pp. 3–12. IEEE, 2010.

[29] R. Kumaran, Y.-J. Kim, A. E. Milner, T. Bullock, B. Giesbrecht, and T. Höllerer. The impact of navigation aids on search performance and object recall in wide-area augmented reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2023.

[30] C. J. Lin and B. H. Woldegiorgis. Interaction and visual performance in stereoscopic displays: A review. *Journal of the Society for Information Display*, 23(7):319–332, 2015.

[31] J. P. McIntire, P. R. Havig, and E. E. Geiselman. Stereoscopic 3d displays and human performance: A comprehensive review. *Displays*, 35(1):18–26, 2014.

[32] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: a unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98, 10 pages, p. 179–188. Association for Computing Machinery, New York, NY, USA, 1998. doi: 10.1145/280814.280861

[33] A. M. Rzepka, K. J. Hussey, M. V. Maltz, K. Babin, L. M. Wilcox, and J. C. Culham. Familiar size affects perception differently in virtual reality and the real world. *Philosophical Transactions of the Royal Society B*, 378(1869):20210464, 2023.

[34] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pp. 71–78, 2000.

[35] Y. E. Song, P. Kovacs, M. Niitsuma, and H. Hashimoto. Spatial memory for augmented personal working environments. *J. Adv. Comput. Intell. Intell. Informatics*, 16(2):349–357, 2012.

[36] R. W. Soukoreff and I. S. 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, 2004.

[37] B. V. Syiem, R. M. Kelly, J. Goncalves, E. Velloso, and T. Dingler. Impact of task on attentional tunneling in handheld augmented reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, article no. 193, 14 pages. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445580

[38] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie. Effects of tracking technology, latency, and spatial jitter on object movement. In *2009 IEEE symposium on 3D user interfaces*, pp. 43–50. IEEE, 2009.

[39] R. J. Teather and W. Stuerzlinger. Pointing at 3d targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 87–94. IEEE, 2011.

[40] F. van den Oever, V. Gorobets, B. Saetrevik, M. Fjeld, and A. Kunz. Comparing visual search between physical environments and vr. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 411–416. IEEE, 2022.

[41] X. Wang, S. K. Ong, and A. Y. Nee. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing*, 4:1–22, 2016.

[42] M. Whitlock, E. Harnner, J. R. Brubaker, S. Kane, and D. A. Szafir. Interacting with distant objects in augmented reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 41–48. IEEE, 2018.

[43] J. M. Wolfe, A. M. Aizenman, S. E. Boettcher, and M. S. Cain. Hybrid foraging search: Searching for multiple instances of multiple types of target. *Vision Research*, 119:50–59, 2016. doi: 10.1016/j.visres.2015.12.006

[44] S. C.-Y. Yuen, G. Yaoyuneyong, and E. Johnson. Augmented reality: An overview and five directions for ar in education. *Journal of Educational Technology Development and Exchange (JETDE)*, 4(1):11, 2011.

[45] J. Zhang, Z. Dong, R. Lindeman, and T. Piumsomboon. Spatial scale perception for design tasks in virtual reality. In *Proceedings of the 2020 ACM Symposium on Spatial User Interaction*, pp. 1–3, 2020.

[46] E. Zhu, A. Hadadgar, I. Masiello, and N. Zary. Augmented reality in healthcare education: an integrative review. *PeerJ*, 2:e469, 2014.

Supplementary Analyses for The Cost of Virtuality Switching: Searching for Physical and Virtual Targets in Optical-See-Through Augmented Reality

A Analysis of Learning Effects

The following two tables show the statistical results regarding potential learning effects from condition orders and block orders. An ANOVA style GLMM analysis revealed that neither condition number (1-6) nor block number (1-12) had any significant effect on any of our dependent variables: Time, Selection Deviation, Head Movement, Head Rotation, Controller Movement, or Controller Rotation.

Table 1: Learning Effect Analysis: Condition Order Influence on Performance Metrics

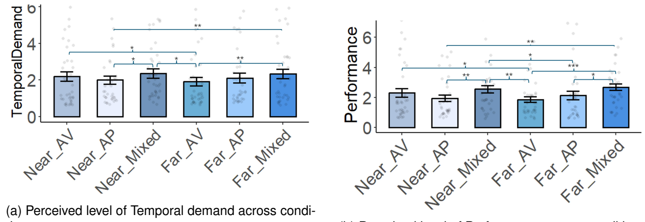
Metric-Condition order	χ^2	p-value	Significance
Time	0.9465	0.9668	ns
Selection Deviation	5.9082	0.3153	ns
Head Movement	5.8105	0.3251	ns
Head Rotation	6.8205	0.2343	ns
Controller Movement	2.9763	0.7036	ns
Controller Rotation	6.7705	0.2383	ns

Table 2: Learning Effect Analysis: Block Order Influence on Performance Metrics

Metric-Block order	χ^2	p-value	Significance
Time	4.53	0.952	ns
Selection Deviation	7.92	0.721	ns
Head Movement	13.63	0.254	ns
Head Rotation	14.59	0.202	ns
Controller Movement	8.65	0.655	ns
Controller Rotation	15.64	0.155	ns

B Subjective Measurements-NASA TLX

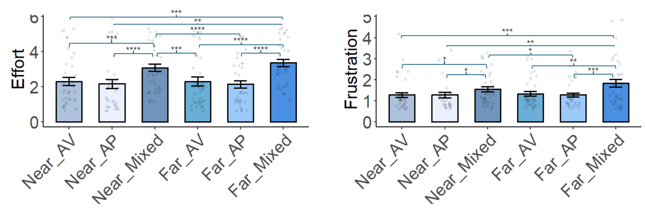
- Temporal demand. Temporal demand was measured using the NASA-TLX with a 7-point Likert scale (see Figure B.1a). A Friedman test revealed a significant main effect of condition on perceived temporal demand ($\chi^2(5) = 17.28704$, $p < 0.01$). Post-hoc pairwise comparisons using Holm correction showed that participants reported significantly higher temporal demand in the Far_Mixed condition compared to Far_AV ($p < 0.01$) and Near_AP ($p = 0.02$), and significantly higher demand in Near_Mixed compared to Near_AP ($p < 0.05$) and Far_AV ($p < 0.05$). These findings indicate that mixed-reality conditions, particularly at farther distances, were perceived as more temporally demanding.
- Performance. Performance was measured using the NASA-TLX with a 7-point Likert scale (see Figure B.1b). A Friedman test revealed a significant main effect of condition on performance, $\chi^2(5) = 31.74172$, $p < 0.0001$. Post-hoc analysis with Holm correction showed that participants performed significantly worse in the Near_Mixed condition compared to Near_AP ($p = 0.008$), and in the Far_Mixed condition compared to Near_AP ($p = 0.004$), Near_Mixed ($p = 0.004$), and Far_AV ($p = 0.001$). These results suggest that mixed physicality, especially at farther distances, negatively affected selection performance.
- Effort. Effort was measured using the NASA-TLX with a 7-point Likert scale (see Figure B.2a). A Friedman test revealed a significant main effect of condition on perceived effort, $\chi^2(5) = 65.88812$, $p < 0.0001$. Post-hoc pairwise comparisons with Holm correction showed that effort was significantly higher in the Near_Mixed condition compared to Near_AV ($p < 0.001$), Near_AP ($p < 0.0001$), and Far_AV ($p < 0.001$); and significantly higher in the Far_Mixed condition compared to all other conditions, including Far_AV ($p < 0.001$) and Far_AP ($p < 0.0001$). These results suggest that mixed-reality conditions—particularly at far distances—demanded the greatest effort from participants during the task.
- Frustration. Frustration was measured using the NASA-TLX with a 7-point Likert scale (see Figure B.2b). A Friedman test revealed a significant effect of condition on frustration levels, $\chi^2(5) = 28.93$, $p < 0.0001$. Post-hoc pairwise comparisons with Holm correction showed that frustration was significantly higher in the Near_Mixed condition compared to Near_AV ($p = 0.021$) and Near_AP ($p = 0.042$), and in the Far_Mixed condition compared to Near_AV ($p = 0.001$), Near_AP ($p = 0.004$), and Far_AP ($p = 0.002$). These results indicate that mixed-reality conditions, particularly at far distances, contributed to increased user frustration.



(a) Perceived level of Temporal demand across conditions

(b) Perceived level of Performance across conditions

Figure B.1: Participants' perceived level of (a) Temporal and (b) Performance across conditions.



(a) Perceived level of Effort across conditions

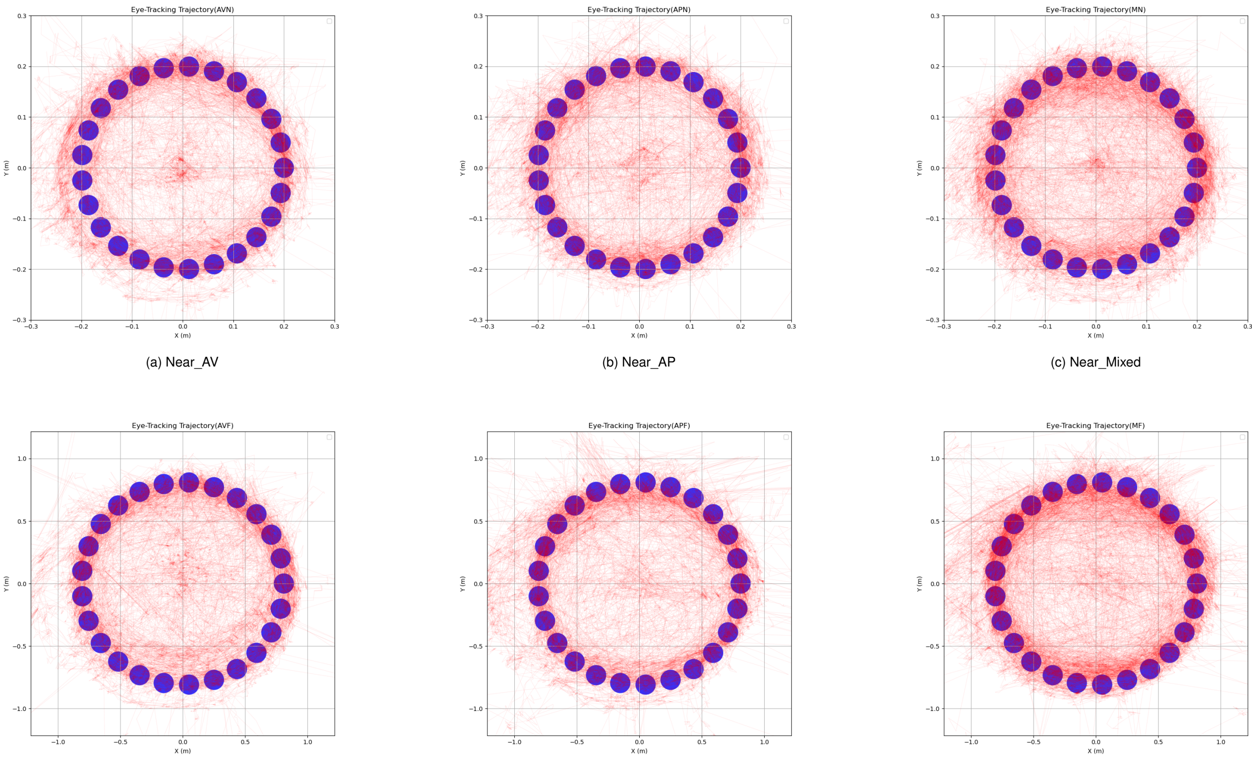
(b) Perceived level of Frustration across conditions

Figure B.2: Participants' perceived level of (a) Effort and (b) Frustration across conditions.

C Eye Tracking Visualization

In order to produce the clearest and most meaningful visuals, eye-tracking trajectories were visualized using data from the 12 participants with the least amount of eye tracking problems and noise (see Figure C.1). As far as we can tell, eye tracking problems occurred randomly, and the visualization stemming from this third of the participants is representative for comparing eye tracking behavior across conditions. Mixed conditions at both distances show denser trajectories than AV and AP conditions, indicating more complex scanning behavior.

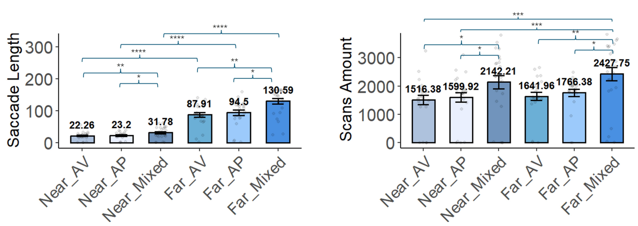
- Scan Amount (see Figure C.1). Scan amount was calculated by the fixation coming within the vicinity of the target center by a threshold of three times the radius of the target. A Friedman test and pairwise Wilcoxon test show that in near conditions, participants' scans touched on objects more times in Near_Mixed than both Near_AV ($p = 0.02$) and Near_AP ($p = 0.02$). In the far conditions, participants'



(a) Near_AV (b) Near_AP (c) Near_Mixed

(d) Far_AV (e) Far_AP (f) Far_Mixed

Figure C.1: Accumulated eye tracking trajectories for a subset of 12 participants exhibiting minimal eye-tracking noise.



(a) Saccade length for a subset of participants exhibiting minimal eye-tracking noise

(b) Number of scan targets for a subset of participants exhibiting minimal eye-tracking noise

Figure C.2: Saccade length and number of scan targets for a subset of 12 participants exhibiting minimal eye-tracking noise.

scans in Far_Mixed touched on objects more times than in Far_AV ($p = 0.003$) and Far_AP ($p = 0.02$). Participants' scans in Far_Mixed also touched on objects more times than in Near_AP ($p < 0.001$) and Near_AV ($p < 0.001$). This suggests that mixed conditions caused participants to scan and rescan the targets more than in the AV and AP conditions.

- Saccade length. Saccade lengths were measured by the total travel distance of eye-tracking fixations (see Figure C.2a). An ANOVA style GLMM analysis reveals a significant main effect of condition on distance $\chi^2(5) = 70.28$, $p < 0.0001$ and physicality $\chi^2(5) = 7.79$, $p = 0.02$. Saccade lengths in mixed conditions were significantly greater than those in both AV and AP conditions at the same distance. This suggests that in mixed conditions, participants had to scan and rescan the targets more than in the AV and AP conditions before making successful selections.