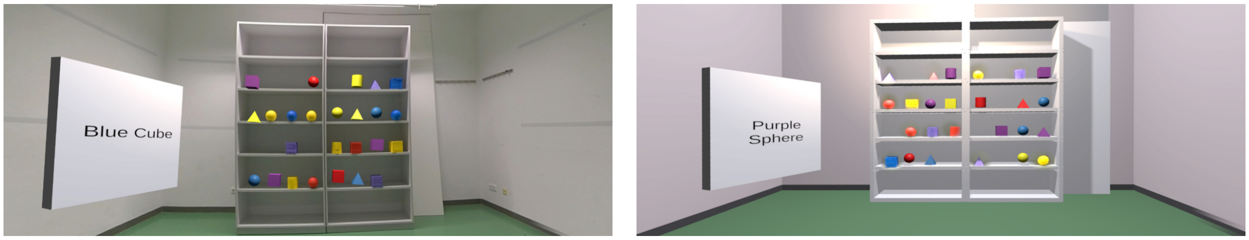
Searching Across Realities: Investigating ERPs and Eye-Tracking Correlates of Visual Search in Mixed Reality

Francesco Chiossi, Ines Trautmannsheimer, Changkun Ou, Uwe Gruenefeld, and Sven Mayer



(a) AR Search Display

(b) AV Search Display

Fig. 1: The visual search displays for AR and AV conditions with physical and virtual targets. In (a), participants have to search for a blue cube. The target object is physical. In (b), participants must search for a purple sphere. The target object is virtual. We did not place objects on the highest and lowest compartments to avoid object occlusion and limited visibility.

Abstract— Mixed Reality allows us to integrate virtual and physical content into users' environments seamlessly. Yet, how this fusion affects perceptual and cognitive resources and our ability to find virtual or physical objects remains uncertain. Displaying virtual and physical information simultaneously might lead to divided attention and increased visual complexity, impacting users' visual processing, performance, and workload. In a visual search task, we asked participants to locate virtual and physical objects in Augmented Reality and Augmented Virtuality to understand the effects on performance. We evaluated search efficiency and attention allocation for virtual and physical objects using event-related potentials, fixation and saccade metrics, and behavioral measures. We found that users were more efficient in identifying objects in Augmented Virtuality, while virtual objects gained saliency in Augmented Virtuality. This suggests that visual fidelity might increase the perceptual load of the scene. Reduced amplitude in distractor positivity ERP, and fixation patterns supported improved distractor suppression and search efficiency in Augmented Virtuality. We discuss design implications for mixed reality adaptive systems based on physiological inputs for interaction.

Index Terms—Mixed Reality, Augmented Reality, Augmented Virtuality, Visual Search, EEG, Eye Tracking, Event-Related Potentials

1 Introduction

Mixed Reality (MR) systems are becoming a part of our daily lives, enriching work and leisure with a blend of virtual and physical information. This will allow us to perform a variety of tasks, such as sketching [50], typing [63], and manipulation [86], where virtual and physical elements blend and either be relevant or distracting. However, the extent to which users can efficiently explore MR environments without compromising attention and cognitive resources remains an open question. Studies showed that navigating MR environments imposes a significant cognitive load, impacting users' ability to effectively focus amidst a multitude of stimuli [46, 49]. This issue becomes particularly pronounced in cluttered visual environments, where users must discern items of high relevance among many distracting stimuli [25, 81]. Thus, while technology has progressed and blending is now possible, it is not yet clear if blending realities will help to support users in their tasks, or if blending is more distracting than helpful.

- *Francesco Chiossi is at LMU Munich, Munich, Germany. E-mail: francesco.chiossi@um.ifi.lmu.de.*
- *Ines Trautmannsheimer is at LMU Munich, Munich, Germany. E-mail: i.trautmannsheimer@campus.lmu.de.*
- *Changkun Ou is at LMU Munich, Munich, Germany. E-mail: research@changkun.de.*
- *Uwe Gruenefeld is at the University of Duisburg-Essen, Germany. E-mail: uwe.gruenefeld@uni-due.de.*
- *Sven Mayer is at LMU Munich, Germany. E-mail: info@sven-mayer.com.*

Manuscript received xx xxx. 2024; accepted xx xxx. 2024. Date of Publication xx xxx. 201x; date of current version xx xxx. 2024. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.2024.xxxxxxx

Previous work in visual search showed processing virtual and physical cues simultaneously is demanding, suggesting the need to account for these perceptual differences in MR design to enhance usability [47, 89]. Thus, in MR, visual search presents varying degrees of difficulty based on their representation. Although users can currently differentiate between physical and virtual objects due to differences in fidelity, anticipated advancements in MR technology aim to merge real and virtual elements even more [1]. This convergence will likely increase the complexity of visual search tasks by making the distinction between real and virtual content less apparent, thereby placing a greater demand on users' visual processing capabilities [43]. Further, there is evidence that AR might introduce detrimental effects like split attention and visual complexity, as shown in contexts such as medical surgery [16]. Moreover, exploring Augmented Virtuality's (AV) implications in visual search is still in its infancy [83].

Despite the body of work investigating visual search within MR, understanding how users process virtual and physical target information in Augmented Reality (AR) and AV remains elusive. Thus, we explore such a research gap by systematically examining visual search tasks across the Reality-Virtuality continuum. We focus on actualities, i.e., the currently experienced reality of a user on the Reality-Virtuality Continuum [1], involving a shared blend of virtual and physical information, whether distracting or target objects. We conducted a within-subjects user study with two different ACTUALITIES: *AR* and *AV*, where participants engaged in searching for two different types of TARGET that are either *Virtual* or *Physical*. Our objectives are threefold: first, to determine the impact of different actualities (AR vs. AV) and the nature of the targets (physical vs. virtual) on participants' performance and perceived workload. Second, we investigate how distractors are suppressed during visual search tasks, as indicated by the event-related potential (ERP) distractor positivity component (PD) [27]. The PD is an ERP component that reflects neural processes involved in the active inhibition of distracting stimuli, facilitating more efficient target detection and attention allocation Finally, we seek to evaluate visual search efficiency across the MR continuum and different target types by analyzing eye-tracking metrics such as fixations, saccades, and index of pupillary activity, which serve as indicators of visual search efficiency and cognitive load. These insights support a joint research direction between physiological computing [22] and adaptive MR systems [56]. By sensing users' attentional and workload states, such systems can dynamically adjust the degree of virtuality of the surrounding environment and objects to support target information processing, regardless of MR actuality.

We found that the ERP correlate of distractor suppression showed a more efficient target processing in AV. In line with this, our ET results showed that the visual search was more scattered and cognitively demanding in AR than in AV, as shown by fixation features and index of pupillary activity (IPA). In addition, users reported increased perceived workload in AR compared to AV. Similar conclusions can also be generalized to Virtual Targets, which were faster to be identified with shorter fixations and led to fewer misses than their Physical counterparts. Our findings imply that the interplay between virtuality and target physicality must be considered when designing interactions in MR. Furthermore, we identified distinct physiological patterns tied to the search for task-relevant information in MR environments. These point to physiological computing applications in adaptive MR systems, which can detect user workload and support visual search capabilities.

2 Related Work

2.1 Visual Search Between Reality and Virtuality

While the term MR is frequently used in research, its meaning remains ambiguous [84]. To avoid confusion, we strictly follow the definition outlined by Milgram and Kishino [64], who define MR as everything that falls between reality and virtuality. Following this definition, every MR experience contains physical and virtual components. Therefore, in MR, the target object of a visual search can be either physical or virtual. Nonetheless, depending on the current manifestation of MR, the amount of physical and virtual content differs. For our holistic investigation of MR, we selected the common manifestations of AR and AV as they are well-known and have opposite physical and virtual content blends.

When exploring visual search performance within MR environments, it is important to consider that perceptual processing might differ significantly between physical and virtual content. Therefore, previous work has proposed virtual environments to replicate experiments conducted in a physical setting [34]. Researchers have conducted comparative studies between physical and virtual settings, showing some evidence that visual search speed, accuracy, workload, and cognitive absorption factors are comparable in VR and physical environments [88]. In VR, participants rely heavily on familiar size for object recognition, irrespective of binocular cues availability [78]. This altered reliance on familiar size cues suggests that strategies for visual search in MR need to account for these perceptual differences. Similarly, the interface of human vision and perception with digital content blended in physical reality highlights the necessity of including human perceptual capabilities in MR design for enhanced visual search and object manipulation [47]. Furthermore, spatial reasoning and scale perception studies reveal that physical models offer more accuracy and quickness in conveying object size than VR systems [17, 95]. This shows how scale perception is relevant when designing virtuality in MR environments [85]. Realistic applications of MR technology emphasize the role of virtual objects as concrete models that facilitate the perception and representation of real objects, highlighting the need for MR technologies to bridge perceptual differences between virtual and real objects for effective visual search strategies [15].

However, visual search processes that span physical and virtual content remain uncertain and have not been explored in previous work. So far, researchers have often limited the search for virtual objects to virtual environments and for physical objects to physical environments (e.g., [88]) . Alternatively, they have focused on one specific manifestation, for example, AR, to investigate the search for both virtual and physical objects [48], providing evidence that physical objects are harder to recall in a physical environment. Consequently, our work focuses on the simultaneous presence of virtual and physical content in different manifestations (AR, AV) of MR.

2.2 Event-Related Potentials in Visual Search

Exploring attentional resource allocation in real-world and MR environments is crucial for visual neuroscience and human-computer interaction (HCI) [20]. Here, EEG's high temporal resolution allows precise investigation of such processes, particularly useful in dynamic MR contexts for assessing interactions and user experience [67].

When investigating visual search, two processes are central: visual attention and distraction [58]. Here, Event-Related Potentials (ERPs) are a common technique to investigate visual search tasks. ERPs are specific patterns of brain activity, measured as voltage changes on the scalp, that are triggered in response to particular events or stimuli. A late positive voltage deflection occurs after visual search display onset over the visual cortex, termed PD [38]. PD acts as a mediator for inhibiting distracting stimuli, acting in cognitive control mechanisms that manage potentially distracting information to facilitate improved visual search performance [27]. Several studies have shown how the PD component is involved in feature-search mode, i.e., identifying a specific shape rather than for a singleton [52], resulting in increased PD amplitudes, reflecting enhanced cognitive efforts to process or inhibit distractors [38]. However, investigation of PD and distractor suppression during visual search in MR environments remains mostly unexplored. A starting point is given by Marini et al. [61], who explored the distinct visuo-motor brain dynamics elicited by real-world objects compared to planar images. Their results showed that physical objects trigger stronger and more prolonged activation in neural populations involved in visuomotor action planning. This suggests that objects' physicality and affordance significantly influence the allocation of attentional resources and the inhibition of distracting stimuli. Their findings indicate that in MR environments, where the blending of real and virtual elements can vary the degree of object physicality and affordances, PD responses could differ significantly from those observed in purely physical or image-based settings.

2.3 Eye Tracking Correlates of Visual Search

Eye tracking is a non-invasive method to evaluate visual search behavior and its cognitive and perceptual processes [60]. During visual search, eye movements are guided by cognitive processes that support target identification [20]. Recent advancements highlight eye-tracking integration when searching for target information in MR, allowing usability evaluation and active and passive input controls [71]. Gardony et al. [26] argue that eye-tracking can inform MR design by offering interface interactions and adaptive graphical rendering capabilities. These developments are underpinned by eye-tracking metrics like fixation duration, count, saccade, and pupil patterns, which are informative about visual attention and cognitive load in MR.

Common metrics for attention allocation and information processing are *fixation duration*, *fixation count*, and *pupil diameter* [19]. Longer fixation durations suggest that users invest more cognitive resources to process information in MR [72]. In MR settings, particularly in visual search tasks, fiction metrics reveal correlations between users' eye fixations and various factors such as task-related objects and users' head rotation velocities [39].

Saccades connect fixation points and can inform how visual search patterns are executed in MR [74]. Saccades, play a crucial role in shifting visual attention towards targets, indicating the efficiency of visual searches. Typically, efficient searches are marked by fewer, well-directed saccades, especially in tasks with lower perceptual demands [29]. Environmental variables, such as distinctively colored distractors, can impact saccadic behavior, altering the trajectory and saccadic effectiveness [72].

Pupil diameter emerged as a significant indicator of visual search efficiency, closely tied to cognitive load and task difficulty in MR [7], and recently extended as a near-real-time metric for measuring cognitive load [56]. Recent advancements in MR interfaces leverage dynamic context-aware optimization techniques, significantly reducing manual adjustments and enhancing the user experience by adapting interfaces in real-time based on the user's cognitive load, task, and environment [56] or by inferring users' interaction goals [14]. However, interpreting pupil diameter as a cognitive load measure in MR is not straightforward. External factors such as scene colors, brightness, and movement significantly influence pupil size [69].

3 Mixed Reality Visual Search Environment

First, we designed an MR environment that allows users to perform a visual search task across two actualities of the MR continuum, *AR* and *AV*, where physical objects and virtual objects are presented as search targets. We employed a real-world setup to represent the physical world, objects, and their virtual counterparts to achieve this. This approach embraced an ecological methodology, drawing inspiration from David et al. [14], where participants selected objects placed on shelves, mirroring their real-world analogs. We implemented the visual search task using two different models of the same scene with different levels of virtuality, see Fig. 2a and Fig. 2b.

3.1 Implementation of the Real World

3.1.1 Physical Environment

For the physical environment, we chose a room at our institution with a minimalist aesthetic with white walls, a green floor, and a grey shelf as the focal point for the visual search task, see Fig. 2a. The dimensions of the room are 8.7 meters in length, 4.6 meters in width, and 2.8 meters in height. This setting was selected to minimize visual noise and eliminate extraneous details that could detract from the task, ensuring that participants' attention was drawn primarily to the target objects.

3.1.2 Physical Objects

The stimuli object placed on the shelves set consisted of four physical objects resembling real-life counterparts: a sphere, e.g., a soccer ball; a cylinder, e.g., a soda can; a cube, e.g., a rolling dice; and a pyramid, e.g., building blocks. We fabricated physical objects using extruded polystyrene with a heating wire and foam cutter with angle adjustment. Physical objects are not directly derived from virtual objects but are physical replicas of our virtual objects. We placed them on the shelf in our study room and photographed them from 3 m distance, i.e. participants' chair position. As our virtual environment precisely mimics the study room, we can blend these photographs directly with the virtual environment. We chose photographs over a live video feed to better control the experiment by avoiding time-consuming reconfigurations between the trials and eliminating the potential for human error introduced through the condition assembly on the spot. All objects fit into a cube of 12 cm per dimension (2.75° visual angle) based on [59], i.e., half the shelf compartment height (24 cm) with a volume of 1,728 cm^3, ensuring that all three objects are 10 cm equidistant. Each shelf is 96 cm wide, allowing four (virtual/physical) objects to be equally spaced (12 cm distance from each other) per shelf compartment. To design the Physical Object - Visual Search trials, we took 200 pictures with 13 physical object search displays with an InstaVR 360 Pro2 (7680×4320 pixels, 120 fps) placed in randomized locations to design the trials for the physical objects. The objects did not overlap with the location of the virtual objects on the shelf.

3.2 Implementation of the Virtual World

3.2.1 Virtual Environment

For the virtual world scene, as depicted in Fig. 2b, we employed a systematic modeling approach to recreating a laboratory setting that mirrors its physical counterpart accurately, following previous work [53]. Starting with a low-fidelity model, we constructed the basic geometry to outline all principal features of the room, such as the shelf, walls, and floor, ensuring clear identification without the inclusion of details like door knobs or complex textures. Progressing to a medium-fidelity model, we refined the geometry, adding elements such as detailed window frames, accompanied by low-resolution textures to enhance visual depth. Overall, we aimed for a high-fidelity representation, with increased polygon count for all objects and the addition of all visible minor features. We employed high-resolution textures to achieve a realistic appearance, and we introduced baked lighting to incorporate static shadows. Throughout this process, we meticulously controlled for luminance to ensure consistent lighting and visual perception across different fidelity levels.

3.2.2 Virtual Objects

The set of virtual objects was the same as for the physical objects, see Sect. 3.1.2. For the virtual object color coding, we color-picked the original color from the physical objects, resulting in the following RGB values: red `#BF1818`, blue `#377EB8`, purple `#e92053`, and yellow `#FF0000`. Virtual objects have virtual, opaque color features. We designed that the virtual objects to match the dimensions of the physical objects; thus, we maintained spatial consistency in our visual search environment,. Adhering to the specifications for the physical objects, we rendered each virtual object within a virtual cube of 12 cm per dimension, corresponding to a 2.75° visual angle and occupying a volume of 1,728 cm^3.

3.3 Blending the Worlds

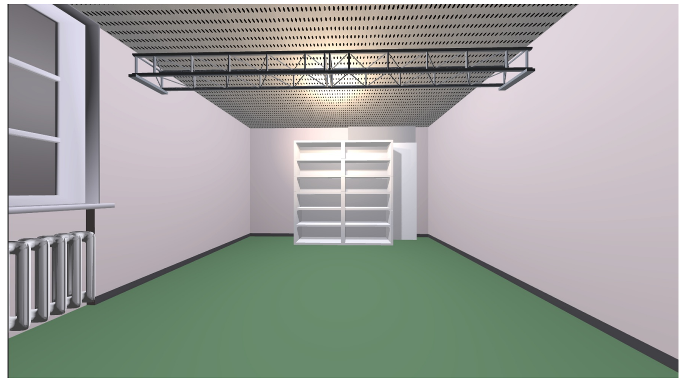
3.3.1 Rendering AR

For the AR condition, which integrates a physical background with physical and virtual objects simultaneously, we utilized a high-resolution image of the physical environment (captured with an Insta360 Pro 2 at 8K resolution) to present the physical items at predetermined shelf spots. Concurrently, we superimposed virtual objects onto specific, generated locations on the shelves, ensuring they did not overlap with their physical counterparts. This was accomplished by capturing images of the physical objects positioned on the shelves, maintaining a consistent distance and luminance, to serve as the backdrop for the subsequent overlay of virtual objects. Following this procedure, we ensured that the placement of physical objects in the real world was consistent, while simultaneously displaying virtual content in a controlled manner. This approach allowed for a controlled integration of physical and virtual elements within the AR environment. For displaying the virtual objects, we opted for simulating a video-see-through AR display within a sphere for its capacity to superimpose digital information directly onto the user's view of the real world [80]. The need for precision and consistency in the presentation of virtual objects drove this choice. Unlike typical AR implementations where objects might possess a degree of transparency, our approach ensured that the virtual objects are solid and visually consistent, thus, avoiding confounders on object saliency due to transparency [41].

3.3.2 Rendering AV

In the AV condition, we rendered the virtual world with virtual and physical objects on the shelves. Here, we presented the virtual objects onto specific, pre-generated locations on the shelves. To display the real objects, we created a Unity shader that renders circular sections from the sphere of the AR condition. This shader not only preserves the texture and depth cues of the physical objects but also maintains their natural lighting and depth, which is crucial for realistic integration. The shader functionality includes rendering circular sections from the sphere of the AR condition, with each section featuring an inner circle at full opacity and an outer circle where opacity decreases with distance from the center, creating a smooth fade-out effect. This technique ensures that the circular cutouts around the physical objects are slightly larger than the objects themselves, allowing for complete coverage regardless of their shape. We applied alpha blending along the borders of these cutouts, as suggested by previous research in AV [63, 91], to achieve seamless integration of physical objects into the virtual environment. The calculations for the circles and their radii are based on the world positions of the pixels, avoiding the use of texture positions that would result in distorted ellipses due to the sphere's curvature. In the rendering sequence within Unity, the virtual cabinet is rendered first (queue: 2000, z-buffer: on), followed by the virtual objects (queue: 2500, z-buffer:

(a) Physical Search Environment



(b) Virtual Search Environment

Fig. 2: MR Visual Search Environments. We situated the visual search task in two environments: a physical one to display AR content and a virtual one to display the physical and virtual objects (AV condition). For the AR scenario, we chose a low-complexity room at our institution to avoid environmental distractions from the task. We modeled the AV environment as closely as possible to its' AR counterpart and controlled for luminance.

on), and finally, the sphere with the custom shader (queue: 3000, z-buffer: off). This order ensures that the physical objects embedded within the virtual shelves are clearly visible and not occluded by other elements.

4 User Study

Our study investigates users' performance identifying physical and virtual target information in AR and AV, see Fig. 1. We engaged participants in an adapted visual search task by Chiossi et al. [10] and situated the task in AR and AV while presenting physical and virtual target objects. We used a 2×2 within-participants experimental design with the independent variables ACTUALITY (two levels: *AR*, and *AV*) and TARGET (two levels: *Physical* / *Virtual*). ACTUALITY describes the envisioned the user is immersed in, and TARGET describes the type of objects, while distractors are always presented of both types. Drawing from previous work from visual search, we formulate the following research questions:

RQ1: Do different actualities impact performance and perceived workload differently?

RQ2: Do the MR actuality and target type impact and eye tracking correlates of visual search efficiency (fixations and saccades), and workload (IPA)?

RQ3: How does distractor suppression in a visual search task vary when searching for target and physical objects across the MR actualities, as indexed by Event-Related Distractor Positivity?

4.1 Procedure

Upon arrival, we briefed participants about the study and gave informed consent. Then, we prepped the water-based EEG data acquisition. Next, we asked them to wear the HTC Vive Pro Eye headset and sit on a chair. We seated participants in front of the shelf, not in motion, minimizing the impact of body movements on visual perception. The distance between the shelf and the seating position was the same as between the 360 camera and the shelf (i.e., 3 m). Keeping the spatial distances the same and having participants stationary helps to minimize potential depth perception and distortion issues. They then performed a five-point eye-tracking calibration. The main part started with participants completing a training phase and experiencing all the experimental conditions. This training phase comprised 20 visual search trials, i.e., 5 with physical targets in AR, 5 with virtual targets in AR, 5 with physical targets in AV, and 5 with virtual targets in AV. Participants needed to achieve at least 80% to proceed; failing that, they repeated the training to meet this criterion. Next, we guided participants through the four conditions containing 100 trials. To avoid learning effects, we counterbalanced the order of conditions in a balanced Latin Williams square design with four levels [90]. After each block, participants responded to the raw NASA TLX questionnaire [35]. We administered the questionnaires using the VR Questionnaires Toolkit [23]. The study averaged an hour in duration, which we compensated with 12 €.

4.2 Task

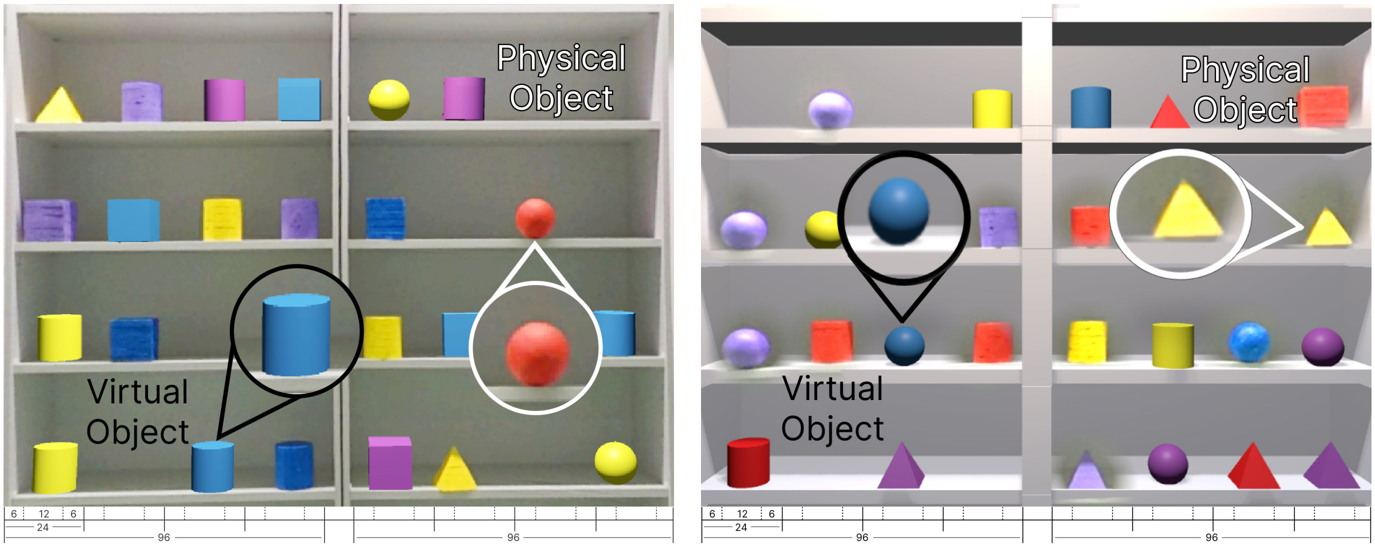
Participants carried out the visual search task in two MR environments: AR and AV. While engaged in one of the two conditions, they were presented with 25 physical and virtual objects placed on a shelf. To select the target item from 24 distractors, they used the trigger button on the VIVE controller. The target object's name was displayed laterally (left or right) in the participant's view to be capable of identifying it. This target display's location (either left or right) was randomly varied across different trials to prevent habituation effects. We chose to display the name and not the picture of the target object to ensure that participants were not biased toward recognizing either the virtual or physical version of the target. Presenting the name rather than the image ensures that both versions of the target are treated equally in the search process, as participants rely on their understanding and interpretation of the name rather than pre-existing visual features from an image. To enforce this, we did not previously inform participants of the physical or virtual nature of the target, which directly required them to identify the object that best matched the description. Participants needed to scan the MR environment visually, aiming at their chosen object using the controller's ray cast to make a selection. Once the target object was aligned with the ray cast, they pressed the VIVE controller's trigger button to confirm their choice. Participants held the controller with their dominant hand, and we encouraged them to respond quickly and accurately.

4.2.1 Trial Structure

We designed our trial based on Forschack et al. [24], with a real-world visual search task approach in mind [93]. The structure of the task, was as follows: (1) we asked participants to fixate a red fixation cross (+) with a pseudorandom duration (1250, 1500, or 1750 ms) at the center of the target display, see Fig. 4. (2) Participants visually searched for the target object; the objects disappeared after selection. (3) After 5000ms, an inter-stimulus interval (ISI) of 1000ms was presented with no cross or objects presented to reset the neural and attentional reserve and avoid fatigue effects [94]. Participants had 5000 ms after visual search display onset to select the target among distractors. For a trial visualization, refer to Fig. 4.

4.2.2 Stimuli

The stimulus set included four objects, both in virtual and physical forms, designed to mirror everyday items: a sphere (resembling a soccer



(a) AR (b) AV

Fig. 3: Objects display in AR and AV conditions. We displayed virtual and physical objects across conditions over two different compartments on the shelves in rows. Each shelf is 96 cm wide, allowing for 4 (virtual/physical) objects equally spaced (12 cm distance from each other) per shelf compartment. We randomly generated 400 unique object placement arrangements in which all 25 virtual and physical objects have a specific location on the shelf. We seated participants 3 m away from the shelves. In this figure, we display the distribution of the virtual-physical objects in a trial and report the schematic of object distances. Distances are reported in cm.

ball), a cylinder (similar to a soda can), a cube (akin to rolling dice), and a pyramid (comparable to building blocks). These stimuli were presented in one of four colors: red `#BF1818`, blue `#0000FF`, purple `#984EA3`, and yellow `#FF0000`. Those objects were placed over a shelf of four rows in two compartments. Each compartment has sixteen possible locations. Thus, objects are spawned for 32 positions. In those 32 positions, we spawn 25 objects. Overall, 1 is the target, 12 are virtual distractors, and 12 are physical distractors. Of the 24 distractors, a third shares the shape of the target (8), a third shares the color of the target (8), and a third has a different shape and color from the target (8). In this way, we control for feature confounders contributing to the overall perceptual load of the objects that could impact visual search [77]. All objects fit into a cube of 12 cm per dimension (2.75°visual angle) based on [59], i.e., half the shelf compartment height (24 cm) with a volume of 1,728 cm ^{3}, ensuring that all three objects are 10 cm equidistant, see Fig. 3b for the AV condition and Fig. 3a for the AR condition.

4.3 Measurements

We collected a set of multi-modal variables: search accuracy, reaction times, missed search trials, PD peak amplitude, eye tracking features (IPA, last fixation duration, fixation count, saccade frequency), and perceived workload (raw NASA TLX [35]). Search accuracy is the percentage of trials where participants identify the target within 5 seconds. Missed trials are those that expire without a selection. Reaction times are measured from the onset of the visual search display until the participant clicks to select the target. The last fixation duration is the duration of the final fixation on the target object before the trial ends.

4.4 Apparatus

We implemented the visual search task using Unity (Version 2022.3.21f1 LTS) and presented the AR and AV conditions through an HTC VIVE Eye Pro headset, with a display resolution of 2880 × 1600 pixels and a 110-degree field of view. We used the MR toolkit *VRception* [33] for the implementation. The environment tracking employed two HTC Vive 2.0 lighthouses. We used a LiveAmp (BrainProducts, Germany) amplifier to record EEG signals at 500 Hz for EEG recording. We acquired eye-tracking data at 120 Hz using the HTC Vive Pro Eye headset. We employed the Lab Streaming Layer (LSL) framework to integrate and stream physiological data within our Unity VR setup. We recorded the data on a Windows 10 PC (Intel Core i7-11700K, 3.60 GHz, 16GB RAM).

We measured luminance across conditions to validate our scenes. We measured the luminance inside the headset using a lux meter sensor (LT300, Extech, USA). Using 50 measurements per condition, we found an average luminance for the AR environment of 194.02 (SD=5.36) and 192.05 (SD=15.58) lux in AV. Those values align with luminance guidelines (below 200 nits) based on eye-tracking best practices to avoid confounders for pupil size computation [6, 62].

4.4.1 EEG Recording & Preprocessing

We acquired EEG data (sampling rate = 500 Hz) via LiveAmp amplifier from 32 water-based electrodes from the R-Net elastic cap (Fp1, F3, F7, F9, FC5, FC1, C3, T7, CP5, CP1, Pz, P3, P7, P9, O1, Oz, O2, P10, P8, P4, CP2, CP6, T8, C4, FC2, FC6, F10, F8, F4, Fp2, Fz). We kept impedance levels below $\leq$20 kΩ. We set the reference at FCz during the recording, while FPz was used as ground. We placed the electrodes using the International 10-20 layout. For time synchronization with the MR environment, we employed the Lab Streaming Layer Framework, while for preprocessing and analysis, we used the MNE-Python Toolbox [32]. We first automatically detected bad or outliers channels via random sample consensus (RANSAC) method [4] of spherical splines for estimating scalp potential based on algorithms proposed by Perrin [70]. We then applied a notch filter (50 Hz) and band-passed the signal between (1-15 Hz) to remove high and low-frequency noise. We then re-referenced to the common average reference (CAR). We applied an Independent Component Analysis (ICA) for artifact detection and correction with extended Infomax algorithm [54]. We automatized the labeling and rejection process of ICA components via the MNE plugin "ICLabel" [55]. We rejected epochs that showed blinks, eye movement, muscle, or single-channel artifacts in any of the electrodes. On average, we removed .33 ($SD = .353$) independent components within each participant.

4.4.2 ERP Analysis

We segmented continuous signals between 200 ms before and 1000 ms after the search display onset, removing a 200 ms baseline before stimulus onset. The Pd component was quantified as positive average peak amplitudes in the 300 – 900 ms. This window is centered upon the

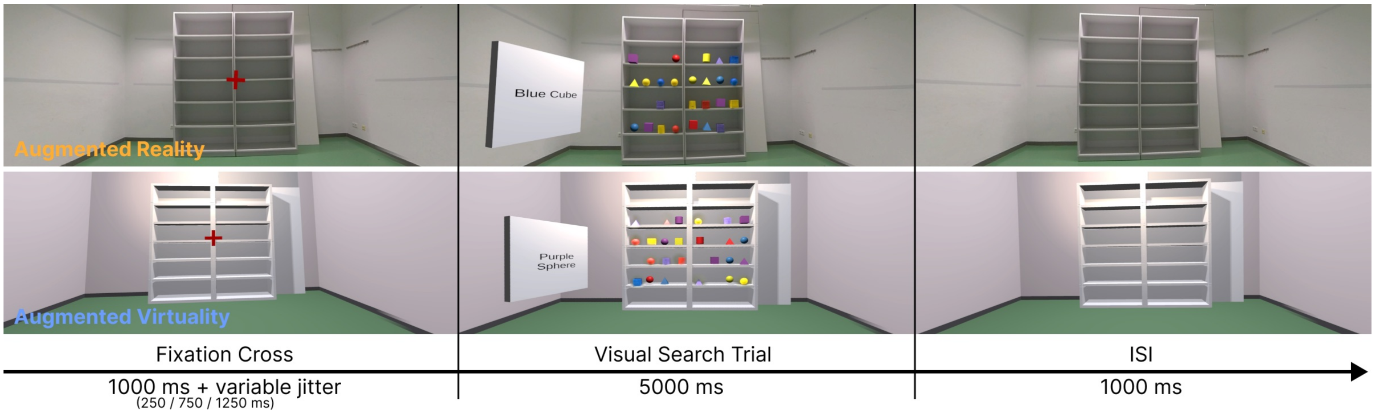


Fig. 4: Trial Structure in the two Conditions. The visual search trial comprised three stages: Initially, a fixation cross was presented for a baseline of 1000 ms, supplemented by a random jitter of 250 ms, 750 ms, or 1250 ms, leading to a total fixation cross duration of 1250 ms to 2250 ms per trial. Following this, participants had 5000 ms to discern the target from among the distractors, and this was succeeded by a 1000 ms interstimulus interval (ISI). Each participant completed 100 trials per experimental condition.

peak latency of each component in the grand average waveforms [76]. For ERP computation, we chose electrodes Oz, O1, and O2 based on previous work [65]. We excluded trials where the last fixation was on a distractor or error trials, i.e., wrong selection, to ensure data captured cognitive processes linked to target recognition [5]. We calculated the overall participant's head movement trajectory as $M = 10.5cm$ ($SD = 4.2cm$) per trial. Moreover, 95% of all HMD positions fell within 5 cm from the trial's starting point.

4.4.3 Eye Tracking Recording & Preprocessing

We acquired three-dimensional head position and orientation data from Unity and recorded eye-tracking metrics using the HTC Vive Pro Eye headset's integrated eye tracker (120 Hz). This data capture utilized the SRanipal eye tracking SDK, which provided vectors indicating eye direction relative to the head and the world. We removed blinks and related artifacts. We defined blinks as missing data points from the eye tracker, with durations ranging from 75ms to 500ms. We removed data 200ms before and after the blinks [2]. We linearly interpolated removed data and smoothed with a 6th-order Butterworth filter whose cutoff frequency was set at .15 Hz [92]. For analysis, these vectors—head direction, eye-in-world direction, and eye-in-head direction—were initially translated into two-dimensional Fick angles. This process was based on the Fick-gimbal method described by Haslwanter [36]. The transformation involved two rotational movements: one around the vertical axis and another around the horizontal axis within the former, enabling us to determine the vectors' positions precisely. We then utilized these 2D Fick angles representing the eye and head orientations as the foundation for further investigation.

4.4.4 Fixation-Saccade Analysis

We analyzed fixation and saccade data using pymovements [51]. For identifying fixations, we utilized pymovements' application of the ID-T algorithm [79], setting the fixation thresholds to a minimum duration of 83ms and a maximum dispersion of 1.8 degrees, in line with prior research [3, 87]. This approach enabled us to derive key metrics related to fixation, including total and average fixation duration, number of fixations, and the interval from the visual search's onset to its final fixation. In analyzing saccades, we applied the microsaccade algorithm offered by pymovements [21], which facilitated the measurement of saccade amplitude—the angular distance between the start and end points of a saccade—and saccade frequency, calculated by the total number of saccades over the trial length.

4.4.5 Index of Pupillary Activity Analysis

We employed the implementation by Duchowski et [18] for computing IPA. Thus, we utilized discrete wavelet transforms (DWT) to analyze pupil diameter signals, starting with a two-level DWT to break down the signal and explore its variability. We normalized the wavelet coefficients to ensure a uniform analysis and identified key peaks in the signal to mark significant changes in pupil diameter. We then applied a universal threshold to filter out noise.

4.5 Participants

This study engaged 20 volunteers ($M = 24.85$, $SD = 4.67$; comprising 11 females, 9 males, none diverse), recruited through institutional email lists and convenience sampling methods. Overall, we did not exclude any participants. The participants' familiarity with AR, AV, and VR technologies was assessed, following previous work [8]. All participants had prior exposure to AR ($M = 2.76$, $SD = 1.56$), AV ($M = 2.88$, $SD = 1.97$), and VR ($M = 3.8$, $SD = 1.83$) technologies, rated on a familiarity scale ranging from 1 (not familiar at all) to 7 (extremely familiar). Exclusion criteria for participants included a medical history of psychological or neurological disorders, color blindness, and visual impairments.

5 Results

In this section, we first present the results of our multimodal evaluation. We employ a Generalized Linear Mixed Model (GLMM) to investigate differences in the behavioral measures, ERPs, and eye-tracking features. We performed model selection based on the Bayesian information criterion (BIC). Details of this process can be found in Supplementary Materials, see Sect. 8. We determined perceived workload scores using the Shapiro-Wilk test, t-tests, or paired samples of the Wilcoxon test.

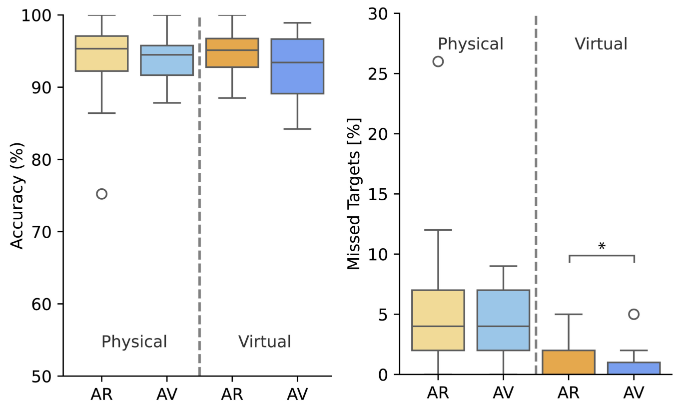
5.1 Behavioral Data

5.1.1 Accuracy

First, we analyzed the overall accuracy, see Fig. 5a. Within this model, the effect of Actuality at the *AV* level was negative but not significant ($\beta = -.12$, 95% CI [-2.46, 2.22], $t(78) = -.10$, $p = .921$; Standardized $\beta = -.03$, 95% CI [-0.58, 0.53]). Similarly, Target at the *Virtual* level was positive but without statistical significance ($\beta = .85$, 95% CI [-1.49, 3.19], $t(78) = 0.72$, $p = .472$; Standardized $\beta = .20$, 95% CI [-0.35, 0.76]). Additionally, the interaction effect between Actuality *AV* and Target *Virtual* was negative, yet not significant ($\beta = -1.84$, 95% CI [-5.14, 1.47], $t(78) = -1.11$, $p = .272$; Standardized $\beta = -.44$, 95% CI [-1.22, .35]). Participants maintained a consistent level of performance regardless of the Actuality or Target.

5.1.2 Missed Targets

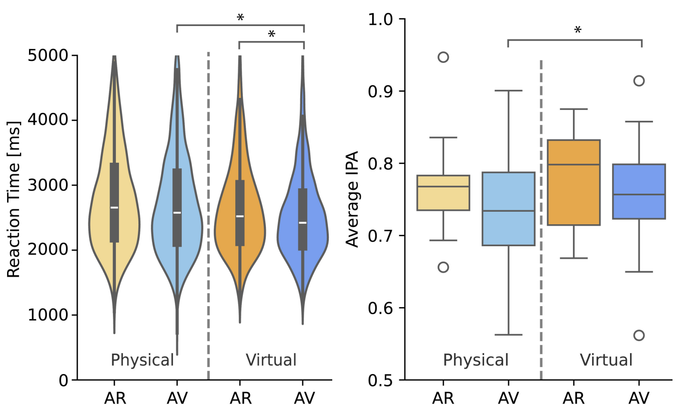
We analyzed the targets participants missed to select within the 5 seconds of the task, see Fig. 5b. Within this analytical framework, the effect of Actuality at the *AV* level was found to be negative, but not significant ($\beta = -1.24$, 95% CI [-3.05, .58], $t(78) = -1.36$, $p = .178$;



(a) Accuracy

(b) Missed Targets

Fig. 5: Accuracy and Missed Targets for the Visual Search Task. Participants performed with comparable accuracy levels across conditions. However, when inspecting the Missed Targets, i.e., when spending the entire trial duration searching for the target with no selection, we found that VIRTUAL TARGETS showed the lowest amount of Misses.



(a) Reaction Time

(b) Index of Pupillary Activity (IPA)

Fig. 6: Reaction times and IPA results. For reaction times, we only computed correct trials. Here, participants showed faster reaction times when searching for objects in AV searching for VIRTUAL OBJECTS. When inspecting IPA, participants showed decreased IPA, indicative of workload in the AV condition.

Standardized $\beta = -0.32$, 95% CI [-.80, .15]). Conversely, the effect of TARGET at the *Virtual* level on misses was both significant and negative ($\beta = -4.43$, 95% CI [-6.24, -2.61], $t(78) = -4.86$, $p < .001$; Standardized $\beta = -1.16$, 95% CI [-1.63, -.68]), indicating a reduction in misses for tasks involving virtual targets. The interaction between ACTUALITY *AV* and TARGET *Virtual* was positive but did not reach statistical significance ($\beta = 0.81$, 95% CI [-1.76, 3.38], $t(78) = 0.63$, $p = .532$; Standardized $\beta = .21$, 95% CI [-.46, .88]).

5.1.3 Reaction Times

We analyzed the reaction times, depicted in Fig. 6a. The model showed a significant negative effect of ACTUALITY ($\beta = -106.11$, 95% CI [-149.34, -62.88], $t(8294) = -4.81$, $p < .001$; Standardized $\beta = -0.13$, 95% CI [-0.18, -0.08]. The effect of TARGET [Virtual] is statistically significant and negative ($\beta = -239.21$, 95% CI [-284.40, -194.02], $t(8294) = -10.38$, $p < .001$; Standardized $\beta = -0.29$, 95% CI [-0.35, -0.24]), suggesting faster reaction times when finding virtual targets. The interaction effect of Actuality [AV] × Target [Virtual] is statistically non-significant and negative ($\beta = -7.78$, 95% CI [-72.23, 56.67], $t(8294) = -0.24$, $p = 0.813$; Standardized $\beta = -9.45e-03$, 95% CI [-0.09, 0.07]).

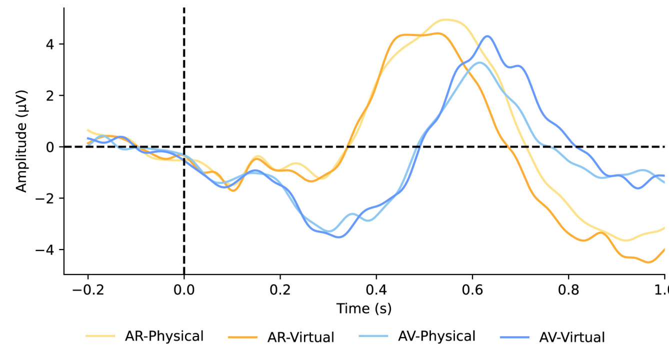
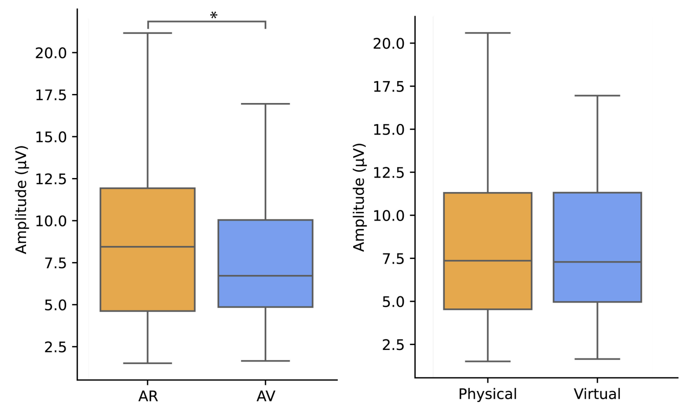


Fig. 7: Grand Average ERP event-locked to visual search display onset. Data reflect the results obtained from occipital ROI for each ACTUALITY and TARGET condition. The plot suggests a pronounced decrease in peak amplitude on PD, with marked variations between AV and AR. We found no effects of TARGET.



(a) Actuality

(b) Target

Fig. 8: PD Amplitude. (a) The effect of ACTUALITY reveals a significant reduction in Amplitude when comparing AR to AV, showing a lower Amplitude in AV. An asterisk marks this difference to denote statistical significance. (b) We found no effect of TARGET.

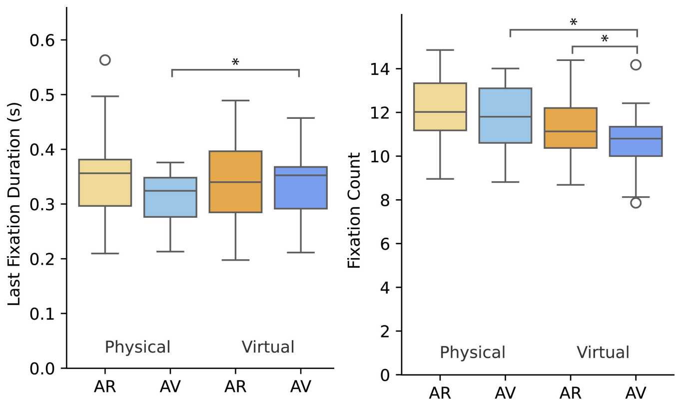
5.2 ERP - Distractor Positivity

Analysis revealed a main effect of ACTUALITY, for the *AV* condition, significantly reducing Amplitude, evidenced by $\beta = -3.64$, 95% CI [-6.33, -.95], $t(74) = -2.7$, $p = .009$, with a moderate standardized beta of -.32. Conversely, no significant effect of TARGET on Amplitude was observed ($\beta = -.96$, 95% CI [-3.65, 1.73], $t(74) = -.71$, $p = .478$), with a small effect size (standardized beta = -.08). Additionally, the interaction between ACTUALITY and TARGET was positive but not significant ($\beta = 2.05$, 95% CI [-1.76, 5.85], $t(74) = 1.07$, $p = .287$), with a standardized beta of .18. ERP grand average is visualized in Fig. 7, and boxplot visualizations are provided in Fig. 8.

5.3 Eye Tracking Data

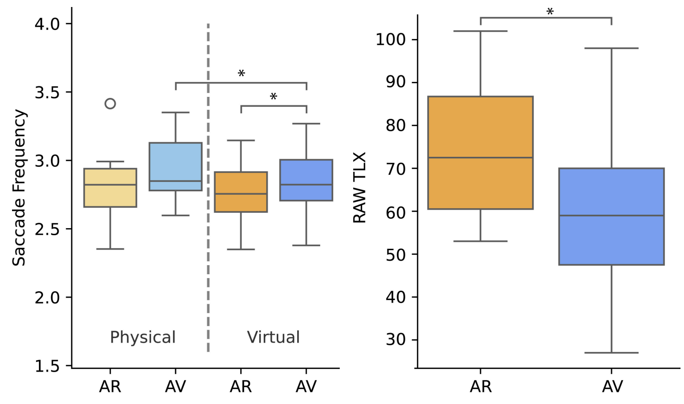
Index of Pupillary Activity (IPA) We found a significant decrease in IPA with ACTUALITY AV ($\beta = -0.03$, $95\%CI[-0.05, -0.01]$, $t(7093) = -3.34$, $p < .001$), suggesting a reduction in cognitive load within AV environments, see Fig. 6b. Conversely, we found no significant changes associated with TARGET Virtual ($\beta = .01$, $95\%CI[-.009, .03]$, $t(7093) = 1.09$, $p = .275$), and the interaction between ACTUALITY AV and TARGET Virtual also did not significantly affect IPA ($\beta = .008$, $95\%CI[-.02, .04]$, $t(7093) = .55$, $p = .582$).

Last Fixation Duration The model reported that ACTUALITY [AV] significantly reduces the duration of the last fixation ($\beta = -.04$, $95\%CI[-.05, -.03]$, $t(7093) = -6.08$, $p < .001$; Standardized $\beta = -.19$, $95\%CI[-.25, -.13]$), suggesting a shorter engagement period



(a) Last Fixation Duration (b) Fixation Count

Fig. 9: Last Fixation Duration and Fixation Count results. We found a negative significant main effect for AV in Last Fixation Duration (left). For Fixation Duration, we found that participants perform faster fixation in AV and when searching for VIRTUAL TARGETS.



(a) Saccade Frequency (b) Perceived Workload

Fig. 10: The Saccade Frequency and Perceived Workload from the raw NASA TLX results. Saccade frequencies were higher in the AV condition, reflecting a more active search process, but lower with VIRTUAL TARGETS indexing fast target processing. Results from Raw NASA-TLX show how participants perceived the AR scenario as more demanding than the AV scenario.

in AV conditions than AR, see Fig. 9a. Conversely, the effect TARGET [Virtual] showed a non-significant reduction in fixation duration ($\beta = -.01$, $95\%CI[-.02, 2.54e-03]$, $t(7093) = -1.59$, $p = .113$; Standardized $\beta = -.05$, $95\%CI[-.11, .01]$). Moreover, a significant negative effect was observed in the interaction between ACTUALITY [AV] and TARGET [Virtual] ($\beta = .03$, $95\%CI[.01, .05]$, $t(7093) = 3.40$, $p < .001$; Standardized $\beta = .15$, $95\%CI[.07, .24]$).

Fixation Count The model output showed that ACTUALITY [AV] yielded a significant decrease in fixation count ($\beta = -.35$, $95\%CI[-.57, -.13]$, $t(7093) = -3.09$, $p = .002$). This reduction was paralleled by a significant negative impact of TARGET [Virtual] on fixation counts ($\beta = -.92$, $95\%CI[-1.15, -.69]$, $t(7093) = -7.80$, $p < .001$). However, the interaction effect between ACTUALITY [AV] and TARGET [Virtual] did not significantly alter fixation counts ($\beta = -.22$, $95\%CI[-.56, .11]$, $t(7093) = -1.33$, $p = .183$).

Saccade Frequency Analysis revealed that ACTUALITY [AV] significantly increased saccade frequency ($\beta = .12$, $95\%CI[.08, .15]$, $t(7093) = 6.62$, $p < .001$; Standardized $\beta = .20$, $95\%CI[.14, .26]$) Conversely, the introduction of a TARGET [Virtual] was associated with a significant reduction in saccade frequency ($\beta = -.04$, $95\%CI[-.07, -.0034]$, $t(7093) = -2.15$, $p = .032$; Standardized $\beta = -.07$, $95\%CI[-.13, -.0059]$). The interaction between ACTUALITY [AV] and TARGET [Virtual], however, did not significantly influence saccade frequency ($\beta = -.05$, $95\%CI[-.10, .005]$, $t(7093) = -1.75$, $p = .08$; Standardized $\beta = -.08$, $95\%CI[-.17, .009]$).

5.4 Perceived Workload

As data showed a not-normal distribution ($W = .923$, $p = .009$), a Wilcoxon signed-rank test was conducted to compare the NASA-TLX scores between AR and AV. When comparing NASA-TLX scores in AR ($M = 74.7$, $SD = 15.28$) to the scores for AV ($M = 58.75$, $SD = 17.85$), we found a significant difference in the NASA-TLX scores ($V = 164$, $p = .029$) with higher workload in the AR condition, see Fig. 10b.

6 DISCUSSION

We evaluated the impact of different MR actualities (*AR* and *AV*), and targets (*Physical* and *Virtual*), on behavioral and physiological correlates of visual search efficiency, distractor suppression, and workload.

6.1 Impact of Actualities on Visual Search Performance

With our first research question (**RQ1**), we investigate if the ACTUALITY affected the user's performance. The overall accuracy shows a comparable outcome across conditions. On the other hand, we see that the missed targets are significantly higher in the *Physical* conditions. Combining this with our trial design in which the maximum time to search the target is 5000ms, we argue that we see ceiling effects on the overall accuracy. The higher errors in the *Physical* conditions align with slower reaction time results. Here, we found a negative main effect of both ACTUALITY and TARGET, showing that participants were faster in target identification in AV and with Virtual Targets. Those results are consistently supported by our results on perceived workload, where participants reported the AV environment to be less demanding. Here, the AR environment's physical fidelity [30] emerges as a potential factor influencing user performance. The inherent visual noise in AR settings potentially distracts and overloads users' cognitive processing capabilities [73], impacting their ability to swiftly and accurately identify targets. In more ecological settings, where controlling the density and arrangement of visual elements is less feasible than in laboratory conditions, the implications of our findings become even more critical. The design of MR environments must carefully consider how visual fidelity, object complexity, and spatial arrangements affect user performance and cognitive load.

6.2 Eye Tracking Correlates of Visual Search Efficiency

Eye tracking correlates of visual search efficiency were the basis for our **RQ2**. Here, we investigate cognitive workload, indexed by IPA, and search patterns by fixations and saccade frequency. We found that cognitive load, as indicated by a decrease in IPA, was reduced in AV suggesting that virtual surroundings facilitate a more efficient cognitive processing experience compared to AR. Furthermore, search patterns, as evidenced by fixation counts and saccade frequency, showed that AV environments and virtual targets support target identification. Specifically, we observed a significant reduction in fixation counts in AV conditions and when interacting with virtual targets, indicating a streamlined search process with fewer distractions. Conversely, saccade frequency increased in AV settings, reflecting more active visual exploration. However, it decreased for virtual targets, implying that once these targets are detected, they require less extensive scanning to process.

Interpreting our results within Guided Search Theory [93], which suggests both top-down and bottom-up mechanisms drive visual search, we can derive that AV may facilitate these mechanisms more effectively than AR. As indicated by IPA, a decrease in cognitive load in AV suggests less demanding attention resource management [45]. Additionally, fewer fixations and adjusted saccade rates in AV and virtual objects index an efficient attention allocation, likely due to more distinct targets enhancing search efficiency [26]. The lowered saccade frequency for virtual targets highlights their ability to draw and keep attention quickly, a sign of effective bottom-up processing [42]. This efficiency could result from the features of virtual elements, which are simple and less noisy than those in AR, mirroring Guided Search Theory's focus on how stimulus traits direct attention. This suggests AV optimizes visual attributes to engage users better and streamline search tasks.

The findings suggest a trade-off between physical fidelity and optimizing for visual search efficiency. More simple objects, which are easier to identify due to reduced visual noise, align with the principles of effective bottom-up processing by enhancing the saliency of targets. However, simplifying object design can lead to spatial constraints, such as making objects perceptually closer, potentially complicating the visual search task as the perceptual load increases. This interpretation highlights the need to carefully balance the design of MR environments and objects, considering both physical fidelity and the spatial arrangement of stimuli. Thus, this result has implications for the design of MR tasks where visual search efficiency is crucial, such as in training simulations [40], education [28], or navigation [75]. In scenarios where quick identification and interaction with virtual elements are essential, designing environments and tasks with lower fidelity and diminished visual noise, such as in our AV, could enhance performance. For example, in educational MR applications, presenting virtual objects that are salient and easily distinguishable from the surrounding environment could facilitate learning and information retention. Similarly, in navigational aids, ensuring that virtual indicators or paths are designed to stand out against the real-world backdrop could support wayfinding.

6.3 Distractor Suppression in MR Visual Search

In **RQ3** we investigated if distractor suppression varies across ACTUALITIES and TARGETS. We focused on PD, an ERP component that reflects a suppressive process towards distracting information after visuals search display onset [27]. Our results showed that in AV, the effect of actuality on PD was significantly negative, suggesting an efficient allocation of cognitive resources leading to diminished distractor processing. This aligns with the functional significance of PD in visual search: a decreased amplitude usually indicates efficient distractor suppression, resulting in improved behavioral performance through improved target focus [37]. Conversely, in AR, we found an increased PD indexing increased distractor saliency [82].

Top-down and bottom-up attention mechanisms may explain the observed variations in PD amplitudes. In AV settings, top-down control likely facilitates distractor suppression through goal-directed attention processes, enhancing target detection efficiency [66]. Conversely, in AR settings, heightened distractor saliency might engage bottom-up mechanisms, leading to increased PD amplitudes. These variations also relate to sustained attention. Higher PD amplitudes in AR settings indicate prolonged cognitive resource allocation for managing distractors [44], reflecting sustained attention. In contrast, lower PD amplitudes are associated with lower attentional demands, reflected by better focus on targets as shown by the results on missed targets.

Integrating ERP and eye-tracking results, where lower IPA denoted higher cognitive efficiency, supports the fact that participants in AV environments experienced a streamlined visual search. Smaller PD amplitudes suggest efficient distractor processing, are aligned with fewer fixations, and have reduced IPA, indicating early suppression of irrelevant stimuli by top-down control mechanisms. The relationship between lower IPA and efficient distractor suppression suggests that when cognitive load is reduced, as indicated by lower IPA, the brain can more effectively allocate resources to suppress distractors. This efficient suppression likely frees up attentional resources, facilitating faster target identification and reaction times. This early suppression facilitates improved target attention allocation, resulting in faster reaction times, consistent with previous work [37].

6.4 Towards Adaptive Mixed Reality

Our multimodal evaluation allows us to better understand visual search efficiency in MR and consider our metrics as input for an adaptive MR system that can be aware of users' context [13]. Adaptive MR systems, informed by eye-tracking features and PD, can be the foundation for hybrid Brain-Computer Interfaces (BCIs) responsive to workload and attention fluctuations [11, 12, 57]. These systems can dynamically modulate the visual nature of stimuli or introduce virtual aids to augment user performance in visual search tasks [8, 56]. By inputting gaze features and ERP components indicative of cognitive effort towards distractors, MR adaptive interfaces can infer interaction intent and future actions in real-time, i.e., identify exploratory visual search behavior or when their attention is diverted from target information.

Here, we envision interfaces that could dynamically vary the saliency of distracting single elements by dimming, blurring, or otherwise de-emphasizing non-essential visual elements [9]. Conversely, task-relevant information can be highlighted through increased visual saliency or contextual highlighting, thus facilitating target detection. This approach aligns with Cheng et al. [8] exploration into leveraging virtual-physical semantic connections to optimize MR layout designs, where the virtual content's placement and appearance are adapted based on contextual relevance and user workload.

6.5 Limitations and Future Work

While we manipulated MR actualities and target objects, we acknowledge the limitation of utilizing simplified objects and environments. Our controlled setup, with stimuli resembling the shape of simple objects. However, our stimuli set and environments do not fully capture the complexity of real-world scenarios. Recognizing this, we propose a replication study to determine if the observed effects persist with real-world objects and their virtual counterparts. Moreover, we controlled the presented stimuli and environments to ensure reliable eye-tracking results and consistent luminance. However, with an ecological scenario in mind, we propose investigating varying luminance levels to evaluate the effect on IPA for adaptive MR systems.

Lastly, head movements during the experiment could confound the ERP analysis. Even small movements can introduce variability in EEG signals [68], affecting the reliability of ERP components. These movements can alter the spatial relationship between electrodes and the scalp, impacting signal quality and accurate neural localization. This variability could obscure subtle ERP effects or introduce noise that masks true neural responses. Future studies should control or account for head movements using motion correction algorithms [31] or designing tasks that minimize head movement.

7 Conclusion

We investigated target detection across the reality-virtuality continuum in a visual search task. We used behavioral, subjective, and physiological (ERP and eye tracking) measures to evaluate workload, distractor suppression, and visual search efficiency. Our findings show that AV settings and virtual targets improve visual search performance, with AV environments reducing cognitive workload, as indicated by lower IPA, fewer fixations, and increased saccade frequency. Additionally, AV environments negatively affected PD, indicating enhanced distractor suppression and reduced cognitive demands. Identifying these physiological patterns supports the development of MR systems based on physiological computing, laying the groundwork for adaptive MR interfaces.

8 Open Science

Our experimental setup, collected datasets, and analysis scripts are available on the Open Science Framework[1].

Acknowledgments

Francesco Chiossi was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 251654672-TRR 161 (C06).

References

[1] J. Auda, U. Gruenefeld, S. Faltaous, S. Mayer, and S. Schneegass. A scoping survey on cross-reality systems. *ACM Comput. Surv.*, 56(4), 2023.

[1] `https://osf.io/zdpty/`

[2] T. Bafna, J. P. P. Hansen, and P. Baekgaard. Cognitive load during eye-typing. In *Symposium on Eye Tracking Research and Applications*, ETRA '20. ACM, 2020.
[3] J. A. Beintema, E. M. van Loon, and A. V. van den Berg. Manipulating saccadic decision-rate distributions in visual search. *Journal of vision*, 2005.
[4] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, and K. A. Robbins. The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Frontiers in neuroinformatics*, 9, 2015.
[5] A.-M. Brouwer, B. Reuderink, J. Vincent, M. A. van Gerven, and J. B. van Erp. Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *Journal of vision*, 2013.
[6] B. T. Carter and S. G. Luke. Best practices in eye tracking research. *International Journal of Psychophysiology*, 155, 2020.
[7] R. L. Charles and J. Nixon. Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, 74, 2019.
[8] Y. Cheng, Y. Yan, X. Yi, Y. Shi, and D. Lindlbauer. Semanticadapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, 2021.
[9] Y. F. Cheng, H. Yin, Y. Yan, J. Gugenheimer, and D. Lindlbauer. Towards understanding diminished reality. In *Proc. of the Conference on Human Factors in Computing Systems*, CHI '22. ACM, 2022.
[10] F. Chiossi, U. Gruenefeld, B. J. Hou, J. Newn, C. Ou, R. Liao, R. Welsch, and S. Mayer. Understanding the impact of the reality-virtuality continuum on visual search using fixation-related potentials and eye tracking features. *Proc. ACM Hum.-Comput. Interact.*, (MHCI), 2024.
[11] F. Chiossi, C. Ou, C. Gerhardt, F. Putze, and S. Mayer. Designing and evaluating an adaptive virtual reality system using eeg frequencies to balance internal and external attention states. *arXiv preprint arXiv:2311.10447*, 2023.
[12] F. Chiossi, C. Ou, and S. Mayer. Optimizing visual complexity for physiologically-adaptive vr systems: Evaluating a multimodal dataset using eda, ecg and eeg features. In *Proceedings of the 2024 International Conference on Advanced Visual Interfaces*, AVI '24. Association for Computing Machinery, New York, NY, USA, 2024.
[13] F. Chiossi, J. Zagermann, J. Karolus, N. Rodrigues, P. Balestrucci, D. Weiskopf, B. Ehinger, T. Feuchtner, H. Reiterer, L. L. Chuang, M. Ernst, A. Bulling, S. Mayer, and A. Schmidt. Adapting visualizations and interfaces to the user. *it - Information Technology*, 2022.
[14] B. David-John, C. Peacock, T. Zhang, T. S. Murdison, H. Benko, and T. R. Jonker. Towards gaze-based prediction of the intent to interact in virtual reality. In *Symposium on Eye Tracking Research and Applications*, 2021.
[15] J. Dilworth. Realistic virtual reality and perception. *Philosophical Psychology*, 23(1), 2010.
[16] B. J. Dixon, M. J. Daly, H. Chan, A. D. Vescan, I. J. Witterick, and J. C. Irish. Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical endoscopy*, 27, 2013.
[17] 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 *Proc. CHI '23*, 2023.
[18] A. T. Duchowski, K. Krejtz, I. Krejtz, C. Biele, A. Niedzielska, P. Kiefer, M. Raubal, and I. Giannopoulos. The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proc. of the Conference on Human Factors in Computing Systems*, CHI '18. ACM, 2018.
[19] C. M. Durugbo. Eye tracking for work-related visual search: a cognitive task analysis. *Ergonomics*, 64(2), 2021.
[20] M. P. Eckstein. Visual search: A retrospective. *Journal of vision*, 11(5), 2011.
[21] R. Engbert and R. Kliegl. Microsaccades uncover the orientation of covert attention. *Vision research*, 43(9), 2003.
[22] S. H. Fairclough. Fundamentals of physiological computing. *Interacting with computers*, 21(1-2), 2009.
[23] M. Feick, N. Kleer, A. Tang, and A. Krüger. The virtual reality questionnaire toolkit. In *Adjunct Prof. of the Symposium on User Interface Software and Technology*, UIST '20 Adjunct. ACM, 2020.
[24] N. Forschack, C. Gundlach, S. Hillyard, and M. M. Müller. Dynamics of attentional allocation to targets and distractors during visual search. *NeuroImage*, 264, 2022.
[25] C. Furmanski, R. Azuma, and M. Daily. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In *International Symposium on Mixed and Augmented Reality*. IEEE, 2002.
[26] A. L. Gardony, R. W. Lindeman, and T. T. Brunyé. Eye-tracking for human-centered mixed reality: promises and challenges. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR)*, vol. 11310. SPIE, 2020.
[27] N. Gaspelin, D. Lamy, H. E. Egeth, H. R. Liesefeld, D. Kerzel, A. Mandal, M. M. Müller, J. D. Schall, A. Schubö, H. A. Slagter, B. T. Stilwell, and D. van Moorselaar. The distractor positivity component and the inhibition of distracting stimuli. *Journal of Cognitive Neuroscience*, 35(11), 2023.
[28] J. Gerup, C. B. Soerensen, and P. Dieckmann. Augmented reality and mixed reality for healthcare education beyond surgery: an integrative review. *Int. journal of medical education*, 11, 2020.
[29] J. H. Goldberg, M. J. Stimson, M. Lewenstein, N. Scott, and A. M. Wichansky. Eye tracking in web search tasks: Design implications. In *Proc. of the Symposium on Eye Tracking Research & Applications*, ETRA '02. ACM, 2002.
[30] G. Gonçalves, H. Coelho, P. Monteiro, M. Melo, and M. Bessa. Systematic review of comparative studies of the impact of realism in immersive virtual experiences. *ACM Computing Surveys*, 55(6), 2022.
[31] D. Gorjan, K. Gramann, K. De Pauw, and U. Marusic. Removal of movement-induced eeg artifacts: current state of the art and guidelines. *Journal of neural engineering*, 19(1), 2022.
[32] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, and L. Parkkonen. Meg and eeg data analysis with mne-python. *Frontiers in neuroscience*, 7, 2013.
[33] U. Gruenefeld, J. Auda, F. Mathis, S. Schneegass, M. Khamis, J. Gugenheimer, and S. Mayer. Vrception: Rapid prototyping of cross-reality systems in virtual reality. In *Proc. of the Conference on Human Factors in Computing Systems*, CHI'22, 2022.
[34] 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), 2022.
[35] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proc. of the human factors and ergonomics society annual meeting*, vol. 50. Sage publications Sage CA: Los Angeles, CA, 2006.
[36] T. Haslwanter. Mathematics of three-dimensional eye rotations. *Vision Research*, 35(12), 1995.
[37] A. Heuer and A. Schubö. Cueing distraction: Electrophysiological evidence for anticipatory active suppression of distractor location. *Psychological research*, 84(8), 2020.
[38] C. Hickey, V. Di Lollo, and J. J. McDonald. Electrophysiological indices of target and distractor processing in visual search. *Journal of cognitive neuroscience*, 21(4), 2009.
[39] Z. Hu, A. Bulling, S. Li, and G. Wang. Fixationnet: Forecasting eye fixations in task-oriented virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5), 2021.
[40] C. E. Hughes, C. B. Stapleton, D. E. Hughes, and E. M. Smith. Mixed reality in education, entertainment, and training. *Computer Graphics and Applications*, 25(6), 2005.
[41] M. Hussain and J. Park. Effect of transparency levels and real-world backgrounds on the user interface in augmented reality environments. *International Journal of Human–Computer Interaction*, 2023.
[42] L. Itti. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition*, 12(6), 2005.
[43] K. M. Jackson, T. H. Shaw, and W. S. Helton. The effects of dual-task interference on visual search and verbal memory. *Ergonomics*, 66(1), 2023.
[44] C. Janssens, E. De Loof, C. N. Boehler, G. Pourtois, and T. Verguts. Occipital alpha power reveals fast attentional inhibition of incongruent distractors. *Psychophysiology*, 55(3), 2018.
[45] G. Jayawardena. Raemap: Real-time advanced eye movements analysis pipeline. In *ACM ETRA*, 2020.
[46] F. Ke, S. Lee, and X. Xu. Teaching training in a mixed-reality integrated learning environment. *Computers in Human Behavior*, 62, 2016.
[47] R. V. Kenyon and S. R. Ellis. Vision, perception, and object manipulation in virtual environments. In *Virtual reality for physical and motor rehabilitation*. Springer, 2014.
[48] 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), 2022.
[49] R. Kockord and O. Bodensiek. Cognitive load during first contact with mixed reality learning environments. In *Prof. of the Mensch und Computer*

2021, MuC'21. ACM, 2021.
[50] B. I. Kovacs, I. Erb, H. Kaufmann, and P. Ferschin. Mr. sketch. immediate 3d sketching via mixed reality drawing canvases. In *International Symposium on Mixed and Augmented Reality*. IEEE, 2023.
[51] D. G. Krakowczyk, D. R. Reich, J. Chwastek, D. N. Jakobi, P. Prasse, A. Süss, O. Turuta, P. Kasprowski, and L. A. Jäger. Pymovements: A python package for eye movement data processing. In *Proc. of the Symposium on Eye Tracking Research and Applications*, ETRA '23. ACM, 2023.
[52] D. Lamy and H. E. Egeth. Attentional capture in singleton-detection and feature-search modes. *Journal of Experimental Psychology: Human Perception and Performance*, 29(5), 2003.
[53] C. Lee, G. A. Rincon, G. Meyer, T. Höllerer, and D. A. Bowman. The effects of visual realism on search tasks in mixed reality simulation. *transactions on visualization and computer graphics*, 19(4), 2013.
[54] T.-W. Lee, M. Girolami, and T. J. Sejnowski. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation*, 11(2), 1999.
[55] A. Li, J. Feitelberg, A. P. Saini, R. Höchenberger, and M. Scheltienne. Mne-icalabel: Automatically annotating ica components with iclabel in python. *Journal of Open Source Software*, 7(76), 2022.
[56] D. Lindlbauer, A. M. Feit, and O. Hilliges. Context-aware online adaptation of mixed reality interfaces. In *Proc. of the Symposium on User Interface Software and Technology*, 2019.
[57] X. Long, S. Mayer, and F. Chiossi. Multimodal detection of external and internal attention in virtual reality using eeg and eye tracking features. In *Proc. of Mensch und Computer 2024*. ACM, 2024.
[58] S. J. Luck and S. A. Hillyard. Spatial filtering during visual search: evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 1994.
[59] W. J. MacInnes, Ó. I. Jóhannesson, A. Chetverikov, and Á. Kristjánsson. No advantage for separating overt and covert attention in visual search. *Vision*, 4(2), 2020.
[60] B. Mahanama, Y. Jayawardana, S. Rengarajan, G. Jayawardena, L. Chukoskie, J. Snider, and S. Jayarathna. Eye movement and pupil measures: A review. *Front. in Computer Science*, 3, 2022.
[61] F. Marini, K. A. Breeding, and J. C. Snow. Distinct visuo-motor brain dynamics for real-world objects versus planar images. *Neuroimage*, 2019.
[62] N. Matsuda, A. Chapiro, Y. Zhao, C. Smith, R. Bachy, and D. Lanman. Realistic luminance in vr. In *SIGGRAPH Asia*, 2022.
[63] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster. A dose of reality: Overcoming usability challenges in vr head-mounted displays. In *Proc. of the Conference on Human Factors in Computing Systems*, 2015.
[64] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77(12), 1994.
[65] M. Naderi, T. Pladere, R. Alksnis, and G. Krumina. Brain activity underlying visual search in depth when viewing volumetric multiplanar images. *Scientific Reports*, 13(1), 2023.
[66] M. P. Noonan, N. Adamian, A. Pike, F. Printzlau, B. M. Crittenden, and M. G. Stokes. Distinct mechanisms for distractor suppression and target facilitation. *Journal of Neuroscience*, 36(6), 2016.
[67] C. Nwagu, A. AlSlaity, and R. Orji. Eeg-based brain-computer interactions in immersive virtual and augmented reality: A systematic review. *Proc. ACM Hum.-Comput. Interact.*, 7(EICS), 2023.
[68] K. Onikura and K. Iramina. Evaluation of a head movement artifact removal method for eeg considering real-time prosessing. In *8th Biomedical Engineering International Conference*. IEEE, 2015.
[69] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proc. of ETRA '10*. ACM, 2010.
[70] F. Perrin, J. Pernier, O. Bertrand, and J. F. Echallier. Spherical splines for scalp potential and current density mapping. *Electroencephalography and clinical neurophysiology*, 72(2), 1989.
[71] A. Plopski, T. Hirzle, N. Norouzi, L. Qian, G. Bruder, and T. Langlotz. The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Computing Surveys*, 55(3), 2022.
[72] M. Pomplun, T. W. Garaas, and M. Carrasco. The effects of task difficulty on visual search strategy in virtual 3D displays. *Journal of Vision*, 13(3), 2013.
[73] E. D. Ragan, D. A. Bowman, R. Kopper, C. Stinson, S. Scerbo, and R. P. McMahan. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Transactions on Visualization and Computer Graphics*, 21(7), 2015.
[74] N. A. Rappa, S. Ledger, T. Teo, K. Wai Wong, B. Power, and B. Hilliard. The use of eye tracking technology to explore learning and performance within virtual reality and mixed reality settings: a scoping review. *Interactive Learning Environments*, 30(7), 2022.
[75] U. Rehman and S. Cao. Augmented-reality-based indoor navigation: A comparative analysis of handheld devices versus google glass. *Transactions on Human-Machine Systems*, 47(1), 2016.
[76] R. A. Roche, H. Garavan, J. J. Foxe, and S. M. O'Mara. Individual differences discriminate event-related potentials but not performance during response inhibition. *Experimental brain research*, 160, 2005.
[77] Z. J. Roper, J. D. Cosman, and S. P. Vecera. Perceptual load corresponds with factors known to influence visual search. *Journal of experimental psychology: human perception and performance*, 39(5), 2013.
[78] 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), 2023.
[79] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proc. of the Symposium on Eye tracking Research & Applications*, 2000.
[80] A. Samini, K. L. Palmerius, and P. Ljung. A review of current, complete augmented reality solutions. In *2021 International Conference on Cyberworlds (CW)*. IEEE, 2021.
[81] M. Satkowski and R. Dachselt. Investigating the impact of real-world environments on the perception of 2d visualizations in augmented reality. In *Proc. of the Conference on Human Factors in Computing Systems*, CHI'21. ACM, 2021.
[82] R. Sawaki, J. J. Geng, and S. J. Luck. A common neural mechanism for preventing and terminating the allocation of attention. *Journal of Neuroscience*, 32(31), 2012.
[83] R. H. C. e. Souza and E. L. M. Naves. Attention detection in virtual environments using eeg signals: A scoping review. *frontiers in physiology*, 12, 2021.
[84] M. Speicher, B. D. Hall, and M. Nebeling. What is mixed reality? In *Proc. of the Conference on Human Factors in Computing Systems*, CHI '19. ACM, 2019.
[85] L. Sun, T. Fukuda, T. Tokuhara, and N. Yabuki. Difference between a physical model and a virtual environment as regards perception of scale. In *Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2013)*, vol. 457, 2013.
[86] A. Tang, C. Owen, F. Biocca, and W. Mou. Experimental evaluation of augmented reality in object assembly task. In *International symposium on mixed and augmented reality*. IEEE, 2002.
[87] A. Van den Berg and E. Van Loon. An invariant for timing of saccades during visual search. *Vision Research*, 45(12), 2005.
[88] F. van den Oever, V. Gorobets, M. Fjeld, A. Kunz, et al. Comparing visual search between physical environments and vr. In *International Symposium on Mixed and Augmented Reality Adjunct*. IEEE, 2022.
[89] F. van den Oever, V. Gorobets, B. Saetrevik, M. Fjeld, and A. Kunz. Comparing visual search between physical environments and vr. In *International Symposium on Mixed and Augmented Reality Adjunct*, 2022.
[90] B.-S. Wang, X.-J. Wang, and L.-K. Gong. The construction of a williams design and randomization in cross-over clinical trials using sas. *Journal of statistical software*, 29, 2009.
[91] C.-H. Wang, B.-Y. Chen, and L. Chan. Realitylens: A user interface for blending customized physical world view into virtual reality. In *Prof. of the Symposium on User Interface Software and Technology*, UIST '22, 2022.
[92] R. I. Wang, B. Pelfrey, A. T. Duchowski, and D. H. House. Online gaze disparity via bioncular eye tracking on stereoscopic displays. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 2012.
[93] J. M. Wolfe. Visual search: How do we find what we are looking for? *Annual review of vision science*, 6, 2020.
[94] G. F. Woodman and S. J. Luck. Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, 400(6747), 1999.
[95] J. Zhang, Z. Dong, R. Lindeman, and T. Piumsomboon. Spatial scale perception for design tasks in virtual reality. In *Proc. of SUI'20*. ACM, 2020.