Understanding the Impact of the Reality-Virtuality Continuum on Visual Search Using Fixation-Related Potentials and Eye Tracking Features

FRANCESCO CHIOSSI, LMU Munich, Germany
UWE GRUENEFELD, University of Duisburg-Essen, Germany
BAOSHENG JAMES HOU, Lancaster University, United Kingdom
JOSHUA NEWN, Lancaster University, United Kingdom
CHANGKUN OU, LMU Munich, Germany
RULU LIAO, LMU Munich, Germany
ROBIN WELSCH, Aalto University, Finland
SVEN MAYER, LMU Munich, and Munich Center for Machine Learning, Germany

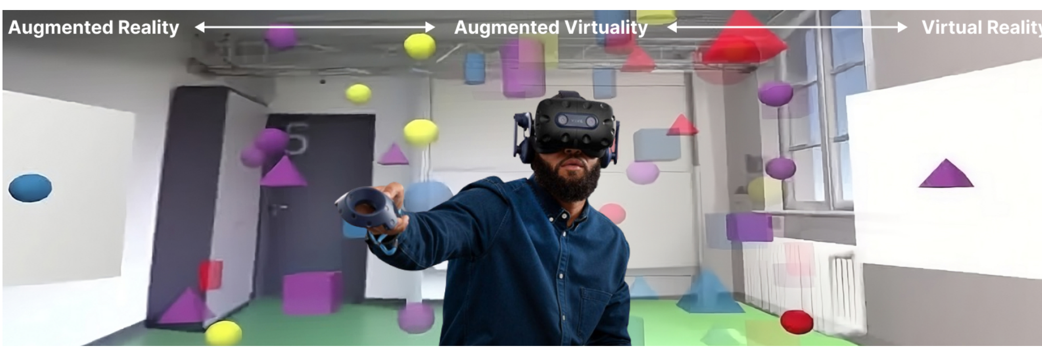


Fig. 1. Mixed reality (MR) will become increasingly embedded in peoples' everyday lives, ranging from productivity to entertainment to health settings. However, different levels of either visual complexity or MR manifestations might place differing demands on the attention allocation of users.

While Mixed Reality allows the seamless blending of digital content in users' surroundings, it is unclear if its fusion with physical information impacts users' perceptual and cognitive resources differently. While the fusion of digital and physical objects provides numerous opportunities to present additional information, it also introduces undesirable side effects, such as split attention and increased visual complexity. We conducted a visual search study in three manifestations of mixed reality (Augmented Reality, Augmented Virtuality, Virtual Reality) to understand the effects of the environment on visual search behavior. We conducted a multimodal evaluation measuring Fixation-Related Potentials (FRPs), alongside eye tracking to assess search

Authors' Contact Information: Francesco Chiossi, LMU Munich, Germany, francesco.chiossi@um.ifi.lmu.de; Uwe Gruenefeld, University of Duisburg-Essen, Essen, Germany, uwe.gruenefeld@uni-due.de; Baosheng James Hou, Lancaster University, United Kingdom, b.hou2@lancaster.ac.uk; Joshua Newn, Lancaster University, United Kingdom, j.newn@lancaster.ac.uk; Changkun Ou, LMU Munich, Germany, research@changkun.de; Rulu Liao, LMU Munich, Munich, Germany, Rulu.Liao@campus.lmu.de; Robin Welsch, Aalto University, Helsinki, Finland, robin.welsch@aalto.fi; Sven Mayer, LMU Munich, and Munich Center for Machine Learning, Munich, Germany, info@sven-mayer.com.

ACM 2573-0142/2024/9-ART281
https://doi.org/10.1145/3676528

efficiency, attention allocation, and behavioral measures. Our findings indicate distinct patterns in FRPs and eye-tracking data that reflect varying cognitive demands across environments. Specifically, AR environments were associated with increased workload, as indicated by decreased FRP - P3 amplitudes and more scattered eye movement patterns, impairing users' ability to identify target information efficiently. Participants reported AR as the most demanding and distracting environment. These insights inform design implications for MR adaptive systems, emphasizing the need for interfaces that dynamically respond to user cognitive load based on physiological inputs.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**.

Additional Key Words and Phrases: Mixed Reality, Visual Search, EEG, Fixation-Related Potentials, Eye Tracking, Physiological Computing

ACM Reference Format:

Francesco Chiossi, Uwe Gruenefeld, Baosheng James Hou, Joshua Newn, Changkun Ou, Rulu Liao, Robin Welsch, and Sven Mayer. 2024. Understanding the Impact of the Reality-Virtuality Continuum on Visual Search Using Fixation-Related Potentials and Eye Tracking Features . *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 281 (September 2024), 33 pages. https://doi.org/10.1145/3676528

1 Introduction

As Mixed Reality (MR) systems become more prevalent in our everyday lives, overlayed digital information will increase the amount and variety of stimuli we perceive and process. Thus, additional information also means more cluttered visual scenes in which task-relevant and task-irrelevant information from both the physical and virtual worlds will populate our spaces. The quality of interaction in such visually rich environments highly depends on how the user perceives, processes, and understands visual information [5]. While MR environments present significant opportunities for performing diverse tasks, we need to understand the cognitive and visual differences between manifestations of the reality–virtuality continuum, such as Augmented Reality (AR), Augmented Virtuality (AV), and Virtual Reality (VR). While there is early evidence that the manifestation of the continuum affects how people gather information in these systems [148], we lack a complete understanding of how the manifestations influence the users' task performance.

This work focuses on a visual search task, which is an integral component of everyday real-world activities, including learning, office work, and gaming. For visual search, users rely on visual cues to quickly scan and identify relevant objects [18]. In MR, users can search for virtual objects, but objects may be harder to locate depending on the manifestation. While both physical and virtual background environments in current MR systems offer varying fidelity, thus remaining distinguishable for users, future systems are expected to blur the lines between real and virtual content, amplifying the task load on visual search. To understand how different MR manifestations influence task load in a visual search task, we utilize Fixation-Related Potentials (FRPs) and Eye Tracking (ET) as main measures. FRPs help us understand cognitive resource allocation under varying complexities in MR, crucial for designing adaptive systems that can dynamically respond to user cognitive load [55, 138]. ET patterns underlie visual search strategies, informing how users interact with MR elements [135]. Such information is relevant for developing MR interfaces that efficiently adapt to user needs in real-time, optimizing the user experience in increasingly complex MR environments [99].

So far, several studies have investigated the impact of different factors on visual search performance in VR, such as the distracting effect of perceptual load [117] and the effects of field-of-view and eccentricity [65, 147]. In AR, most of the research has focused on visual guidance for visual search, i.e., interacting layered interfaces designed in accordance with cueing and search principles of attention theory [15] or employing an attention funnel technique as a 3D cursor to guide the user's attention [14]. However, Eyraud et al. [52] found that AR enhances the user's ability to allocate their attention when task-relevant correctly, but it distracts and degrades their attention allocation when the displayed content is task-irrelevant. Dixon et al. [43] demonstrated how distraction due to AR assistance in medical surgery caused an inattentional blindness effect. This implies that AR could be obtrusive and potentially introduce undesirable side effects, such as split attention and increased visual complexity. Finally, research investigating the implications of AV in visual search remains relatively limited [150].

In the following, we address the previously explained research gap by investigating visual search along the Reality-Virtuality continuum, providing insights into the effects of the manifestation. Therefore, we designed a within-subjects experiment with three different MANIFESTATIONS: *AR*, *AV*, and *VR*. Our research goal is to investigate how the processing of target information in visual search tasks varies across the MR continuum under different task difficulty conditions. We adopt a multimodal approach to achieve this, involving the collection of electroencephalographic (EEG) and eye tracking data, along with behavioral and subjective responses. Specifically, our objectives are threefold: firstly, we aim to discern how different MR manifestations, such as AR, AV, and VR, affect user performance and subjective workload. Secondly, we focus on understanding cognitive resource allocation in visual search tasks, indexed by FRP P3 [55]. Lastly, we assess the influence of the MR continuum on visual search efficiency using eye tracking metrics such as fixations, saccades, and pupil size. These metrics directly measure visual search efficiency and mental workload.

We found that FRP correlate of resource allocation (P3) [85, 87, 126] indexed an increased need for attentional resources in AR. In line with this, our ET results showed that the visual search was more scattered and cognitively demanding in AR than in VR and AV, as shown by fixation features and pupil sizes. In addition, users reported increased perceived workload and distraction in AR compared to VR and AV. Our findings carry implications for understanding the visual search behavior in MR environments. This work contributes to physiological computing for MR by identifying how specific physiological features show different patterns for search in MR experiences. We envision that the physiological correlates identified can be employed as features for future MR interfaces, ensuring that users can seamlessly navigate and effectively interact with digital information while minimizing cognitive strain.

2 Related Work

This section provides an overview of how users perform visual searches in general and search within concrete manifestations of the Reality-Virtuality continuum (e.g., AR and VR). Finally, we review electrophysiological and eye tracking correlates of visual search efficiency that motivated our dependent variables' choice.

2.1 Human Capabilities and Strategies in Visual Search

Visual search is a complex cognitive process where attention is methodically directed toward potential targets amidst various distractors. As highlighted by Wolfe and Horowitz [170], searching involves deliberately directing attention towards objects that might be the sought target [170]. Posner [128] describe shifting of visual attention as a spatial orientation towards a new target [128], which can be broken down into three distinct phases: disengagement from the current target, the transition of attention between stimuli, and finally, focusing on the new target [129]. To explain visual search, Treisman and Gelade's *Feature-Integration Theory (FIT)* posits that it consists of two stages: (1) a preattentive stage where basic features are processed simultaneously, and (2) a focused attention stage where intricate feature combinations are serially identified [155]. Later, Wolfe et al. [169] proposed a refinement with the *Guided Search* model, which also consists of two stages: (1) an initial parallel phase (similar to Treisman and Gelade [155]), and (2) a guided stage where attention is influenced by activation maps from the prior stage, making the search more efficient [169].

Today, we know five factors are strongly relevant for visual search: salience, feature guidance, scene structure, historical search patterns, and the relative value of targets and distractors [170]. Distractors have played a significant role in the study of visual search, as they are key elements that influence the efficiency and accuracy of the search process. Distractors can capture attention, leading to longer search times and increased error rates [23]. This phenomenon is particularly evident in situations where distractors share similar features with the target object, i.e., increasing perceptual load. However, salient distractors lead to shorter search times as it is less challenging to distinguish between relevant and irrelevant stimuli [95]. Therefore, in our study, we investigate low and high perceptual load by varying the number of distractors and using distractors similar to the target to increase the difficulty of the search.

2.2 Visual Search Across Realities

Visual search spans both physical and digital realities, where digital realities take forms such as Augmented Reality (AR), Augmented Virtuality (AV), and Virtual Reality (VR). Milgram and Kishino's taxonomy helps in understanding these various realities and how they interconnect [112]. Their concept, the reality-virtuality continuum, stretches from reality at one end to virtuality at the other. All the elements that fall between these two extremes are categorized as Mixed Reality (MR), including AR and AV. Although Milgram and Kishino originally described MR as a subset of VR [112], suggesting AR falls under VR as well, contemporary research views them as either distinct concepts or positions VR as a subset of MR [151]. In our research, we adopt the view that VR is part of the MR spectrum, recognizing that physical elements still play a role in shaping the virtual experience, following the classification by Skarbez et al. [148] and Auda et al. [4].

In the evolving field of visual search performance, various studies have emphasized specific aspects within their investigated physical and digital realities. In HCI, researchers oftentimes explore the effectiveness of different modalities, such as visual, auditory, or vibrotactile cues, in guiding the user's attention [17, 27]. Many of these studies focus on visual search within the human field-of-view of 180 degrees horizontally [164]. Hence, we follow the same focus. Another emphasis of previous work is to investigate specific manifestations. For example, researchers investigate AR in realistic wide-area outdoor environments and its effects on human behavior, such as the impact of varying lighting conditions and cognitive load on task performance [89]. These studies employ different numbers of search targets, including both physical and virtual objects, and varying methods for identifying and classifying them (Warden et al. [165] showed 48 objects to participants). In some cases, the technology is tailored to create robust and efficient search functions without needing detailed environmental models, utilizing optical markers and projections [22, 58]. Others innovate by employing novel techniques like swarm motion to maintain immersion in VR, while still guiding user attention [94].

Despite the diverse approaches, there is, to our knowledge, a significant lack of research that systematically compares different realities, such as AR, AV, and VR, for visual search. Hence, in our study, we will compare these realities as our primary conditions.

2.3 Fixation-Related Potentials in Visual Search

A shared goal in visual neuroscience and HCI is to understand how we perceive real-world and MR scenarios [48, 133]. From this perspective, EEG is a crucial method to investigate the brain's electrophysiological responses during perceptual and cognitive processing. EEG's high temporal resolution and sensitivity to changes in perceptual and cognitive demands make it particularly suitable for implicitly evaluating user interactions and UX [57, 102] in dynamic computing environments, such as those in MR [39, 115]. Here, Fixation-related potentials (FRPs) are electrophysiological potentials event-locked to the onset of eye fixations rather than the onset of stimuli, i.e., Event-Related

Potentials [175]. FRPs offer a fine-grained temporal resolution, allowing for the analysis of attentional load when we explore an environment and move our eyes to gather information in natural viewing conditions, allowing for the examination of complex, ecological visual stimuli. Earlier FRP components related to stimulus identification and perceptual features have been reported in fixed gaze tasks in the time window between 150-300 ms [81, 130, 153].

Recent investigations have shown how late FRP components in target fixations in visual search can be associated with well-established EEG cognitive resource allocation measures, such as the P3 component [90, 166]. Brouwer et al. [21] showed that FRPs event-locked to target fixations produced a large late cognitive potential similar to the P3 component. Kaunitz et al. [87] extended this finding to ecological settings in a free-viewing search paradigm. Finally, Wenzel et al. [166] showed how EEG and eye tracking data can provide implicit information about the relevance of items on the screen for potential use in online applications. While several studies have investigated the effects of low-level visual features and target-non-target characteristics on FRP components, it is still unclear if FRPs are affected by task demands. As different MR manifestations show inherent visual complexity, blending, and information richness, fixation-related electrophysiological activity might underlie different levels of task demands, affecting the amplitude of FRP components at later (P3) processing stages. Ries et al. [138] confirmed this result in a dual-task, controlled desktop setting. Here, we want to test whether findings on FRPs can likewise index and discriminate resource allocation across the MR continuum.

2.4 Eye Tracking Correlates of Visual Search

Eye tracking is a non-invasive method widely used to measure visual search performance, offering insights into cognitive processes and mental states [74, 105, 173]. Eye movements are a strong visual attention indicator, making its measurement a preferred approach for analyzing the visual search process and behavior across task complexities. During visual search, eye movements are guided by cognitive processes that help efficiently locate the target [49]. Hence, researchers have used eye tracking for measuring visual search behavior and efficiency, with a particular focus on the spatiotemporal dynamics of search behavior across multiple trials [74]. Eye tracking data are typically expressed as *fixations* (stable gaze points with brief pauses) and *saccades* (rapid eye movements in between fixations) [105]. During a visual search, individuals alternate between fixations and saccades, with the whole sequence termed 'scanpath'. The nature of the task may influence different outcomes [26]. For instance, shorter scan paths, with fewer fixations and shorter saccades, indicate a focused and efficient search strategy.

Our overarching goal in this work is to leverage eye tracking metrics to evaluate perceptual load during visual search across the MR continuum. We review metrics derived from fixation and saccadic features, as well as pupil size, that have been applied to investigate information processing, efficiency, and cognitive load related to visual search [91, 113, 124, 154]. In combination, these metrics reveal patterns in visual search behavior and play a crucial role in building a comprehensive understanding.

2.4.1 Fixation metrics. From fixations, the most commonly used eye tracking metrics for attention allocation and cognitive processing are *fixation duration* and *fixation count* [20, 47, 162]. They reflect the users' ability to extract and process information [83, 127]. In visual search, fixation duration provides insights into how long it takes to locate a target among distractors—effectively signaling the cognitive effort required for target identification [20, 127]. Fixation count quantifies the frequency of pauses at specific points of interest, revealing the distribution of attention across different areas. In high perceptual load tasks, longer fixation durations and counts are expected, as users struggle to efficiently locate the target, while the opposite is anticipated under low load conditions [2, 136].

2.4.2 Saccade metrics. We employ *saccade frequency* and *saccade amplitude* as metrics to understand the visual search patterns. Saccades are rapid eye movements that occur between two fixation events and involve a shift of visual attention towards the target. Efficient visual search often involves well-planned saccades that minimize the number of eye movements required to locate the target [127]. In low perceptual load tasks, we anticipate a lower frequency of saccades [60] and a smaller average saccade amplitude, which is the distance the eyes move during a single saccadic eye movement [113]. Environmental factors can introduce bias in saccadic patterns, as recent analyses demonstrated that distractor items with certain colors could strongly influence the guidance of saccadic eye movements [127].

2.4.3 Pupil Size. Pupil size serves as an indicator of visual search efficiency [25, 120, 149], with studies consistently demonstrating its association with task difficulty, a concept established by the pioneering work of Hess and Polt [70]. Consequently, pupil size has been proposed as a near-real-time metric for measuring cognitive load, effectively gauging mental effort during tasks [99, 101]. In high perceptual load tasks, the mean and standard deviation of the pupil size increases in response to the heightened visual processing [8, 20, 84, 116]. While more advanced pupil-based metrics have been proposed to measure workload [45], it is crucial to acknowledge that environmental factors, such as scene colors, brightness, and movement, can influence pupil size, complicating its interpretation [80, 91, 118, 149].

The correlates between fixation, saccades, and pupil size metrics aid in understanding the dynamics of visual search [173], informing how different MR manifestations impact visual search efficiency and workload under different levels of task difficulty.

3 User Study on Visual Search in Mixed Reality

Our study investigates how users identify relevant information and suppress distracting ones across different manifestations of the MR continuum under varying levels of task difficulty. To achieve this, we choose a visual search task to evaluate resource allocation and cognitive load in MR settings as it involves continuous use of attentional and perceptual resources, making them highly sensitive to changes in cognitive demands. This task requires users to locate and identify objects among distractors, a process that closely mimics everyday cognitive challenges encountered in diverse MR settings such as cognitive training [42], information retrieval [13], and daily tasks (such as finding items while cooking [38]). Despite the established research on visual search in controlled settings [48] with known physiological correlates [74, 85], there is a lack of systematic evaluation of how visual search behavior and its physiological indicators translate across MR manifestations. Thus, to bridge this gap, we designed a study where participants perform a visual search task (as modeled after Dey et al. [42]) across AR, AV, and VR scenarios using the validated MR toolkit VRception [63]. To ensure the external validity of our results, we are not only focusing on the continuum but also incorporating task difficulty in the form of perceptual load as a validation control variable.

Drawing from previous work from visual search, we formulate the following research questions:

RQ1: Do different MR manifestations impact performance and perceived workload differently?

RQ2: How do cognitive resource allocation in a visual search task vary across the continuum, as indexed by Fixation-Related P3?

RQ3: Does the MR continuum impact eye tracking correlates of visual search efficiency (fixations and saccades), and workload (pupil size)?

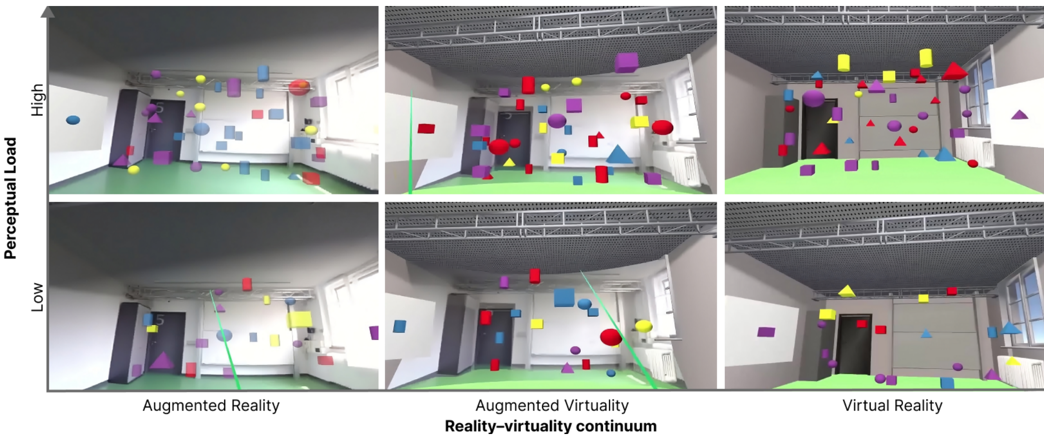


Fig. 2. *Experiment Conditions.* Participants performed a visual search task across the reality-virtuality continuum (AR, AV, and VR) with two possible levels of perceptual load (low and high). The first row shows the screen captures of the high PerceptualLoad conditions across the continuum, while the bottom row depicts low PerceptualLoad conditions across the continuum.

We used a within-participants experimental design. The independent variables were Manifestation (three levels: AR/AV/VR) and PerceptualLoad (two levels: Low / High). To avoid learning effects, we counterbalanced the order of conditions in a balanced Latin Williams square design with six levels [160]. Therefore, independent variables were manipulated using a 3×2 experimental design.

3.1 Mixed Reality Manifestation Conditions

Here, we describe the implementation of the tree levels of our Manifestation independent variable. For a graphical depiction, we refer the reader to Figure 2.

AR condition. In the *AR* conditions, participants execute a visual search task in a real-life 360-video scenario, like if they are asked to look for virtual objects on a shelf. Here, virtual objects have typical AR features. Thus, they are see-through, while the surrounding physical environment maintains the color of the pre-recorded video and transparency. To optimize the visibility of virtual objects against real-world backgrounds, we employed a transparency level of T0.5 (alpha = 0.5) for stimuli in the AR condition, following Hussain and Park [77]'s recommendations. We chose to employ a pre-recorded 360-video of the empty 'real' room in the background to control for background luminance as variations in external environment luminance can lead to changes in color perception in AR [73, 174]. Here, the background was maintained fully opaque mimicking a real environment (alpha = 1.0).

AV condition. In the *AV* condition, participants perform the visual search task in an AV environment, where there are shared features elements from both AR and VR scenarios, i.e., the floor is virtual, and parts of the closet are virtual and physical, see Figure 2. Virtual objects exhibit virtual reality (VR) features, exhibiting an opaque appearance. On the other hand, the physical objects within the environment retain color of the pre-recorded video blended with the VR scene. Our approach to AV aims to be primarily virtual; however, parts of the physical reality are integrated into the experience. We chose a ratio of real and virtual elements that mirrors the approach in AR, where this ratio is flipped (c.f., AR and AV on the reality-virtuality continuum [112]). Consistent with prior studies on AV [*e.g.* 110, 139], we incorporated aspects of the actual room into our virtual experience. We aimed to ensure a fair comparison, in which the ratio of real to virtual mirrors AR without focusing too much on specific objects, as depicted in Figure 2. Here, we used a transparency level of 0.5 for the background and 1.0 for the visual search stimuli.

VR condition. In the *VR* condition, participants carry out the visual search task in an immersive 3D rendering of the room used for the AR condition. The virtual objects and background exhibit VR features in this environment, creating a fully opaque appearance by applying a transparency level of 1.0.

3.1.1 Perceptual Load Conditions. Our PerceptualLoad manipulation consisted of 25 (Low) vs 40 (High) search display set, based on previous work that showed how set size yields inefficient performances between 30 and 60 items [71, 76, 119, 145, 156]. Among distractors, we presented a third of objects with the same color, a third with the same shape as the target, and a third that did not share either color or shape features, i.e., 8 in the low PerceptualLoad and 13 in the high PerceptualLoad conditions, respectively.

3.1.2 Dependent Variables. We collected a set of multimodal variables: (i) FRPs (P3b), (ii) eye tracking features, (iii) behavioral accuracy, (iv) reaction times (RTs), as a proxy for task efficiency„ and to evaluate workload and usability, (v) subjective workload (NASA TLX). We additionally collected three 7-point Likert-scale responses with a unipolar rating scheme [16, 98] (1 : Strongly Disagree, 7 : Strongly Agree) on MR usability, i.e., on how the background was *Distracting* ("The background distracted me strongly from the visual search task"), how the virtual objects were *Overwhelming* (“I perceived the number of virtual objects as very overwhelming": Overwhelming) and, finally how the background made the task more or less *Difficult* ("The background made it very difficult to distinguish the target object from other objects": Difficult).

3.2 Procedure

Upon the participants’ arrival, the experimenter provided them with details about the study’s process and obtained written informed consent. Next, the experimenter set up the water-based EEG recording. Then, participants wore the HTC Vive Pro Eye headset and completed a five-point eye tracking calibration. The study began with a training phase, where participants familiarized themselves with the Visual Search task in a Training block set in a neutral VR environment, i.e., Unity Default Skybox. Participants performed 20 Visual Search trials with a low PerceptualLoad search display. To start the experimental phase, they should have reached at least 80% accuracy. Otherwise, they repeated the training block. Finally, the experimental phase started with the first Visual Search block, following a Latin Williams square design [160]. Each Visual Search block followed the trial structure depicted in Figure 4 and encompassed 50 trials, as depicted in Figure 3. After each block, participants filled in the NASA TLX questionnaire[66] and ad-hoc UX survey. NASA-TLX and UX surveys were administered via the VR Questionnaires Toolkit [53]. On average, the experiment lasted about one hour.

3.3 Task

Participants performed a visual search task across three different MR manifestations. Participants were asked to select the target item with the VIVE controller trigger button among various distractors. Participants were instructed on the target features by presenting an image of the target object in the middle of the user’s view. The sides (left-right) where the target object image was presented were randomized across trials to avoid habituation effects in the eye tracking patterns. To select an object, participants were instructed to visually explore the MR environment and use the

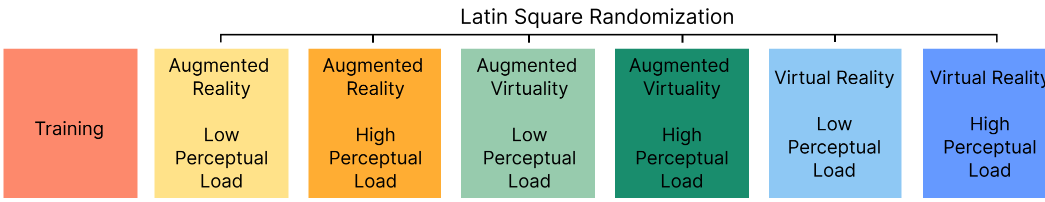


Fig. 3. *Experiment Procedure.* The experiment encompassed seven different blocks. In the first block, participants performed a training session until they reached an accuracy level of 80 % in the visual search task. Finally, the experimental blocks started, manipulating Manifestation and PerceptualLoad using a Latin Square randomization. Refer to Section 3.1 for a complete description of the experimental conditions.

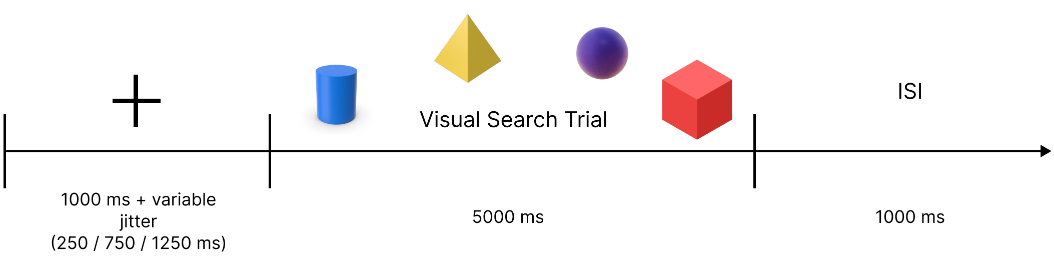


Fig. 4. *Visual Search Trial Structure.* The visual search trial was structured into three phases: Initially, participants were shown a fixation cross for a baseline duration of 1000 ms, followed by an additional, randomly assigned jitter duration of either 250 ms, 750 ms, or 1250 ms. This meant the total fixation cross-presentation varied between 1250 ms, 1750, and 2250 ms in each trial depending on variable jitter duration with no objects in the background. Subsequently, participants had 5000 ms to identify the target among distractors, followed by a 1000 ms interstimulus interval (ISI). Each condition involved 50 trials per participant.

controller's ray cast to aim at the target object. Once the target object was identified and aligned with the raycast, participants pressed the Vive controller's trigger button. Participants hold only one controller using their dominant hand. They were not instructed to return their arms to a neutral position between trials during the visual search tasks. This setup allowed for a more naturalistic interaction with the MR environments, as participants were free to move their arms without the constraint of returning them to a starting position after each trial. They were required to respond to the stimuli as fast and accurately as possible. Participants performed 50 trials per experimental block.

3.4 Trial Structure

Our trial structure was inspired by the work of Forschack et al. [56], following a real-world visual search task approach [29, 168]. The structure of the task, as depicted in Figure 4, was as follows: (1) participants were asked to fixate the fixation cross (+) with a pseudorandom duration (1250, 1500, or 1750 ms) at the center of the camera rig; (2) participants performed the visual search trial; (3) after selection, an inter-stimulus interval (ISI) of 1 second was presented with no cross or objects presented to allow for neural and attentional reset and counteract fatigue effects [3, 172]. Participants had 5000 ms after visual search display onset to select the target among distractors.

3.4.1 Stimuli. The stimulus set consisted of five types of objects: a fixation cross (red) and four possible target/distractor objects (cube, cylinder, pyramid, sphere). Colors were defined by the following RGB values: red = (191, 24, 24); blue = (255, 255, 255); purple = (29, 119, 47), yellow = (255, 255, 0), following recommendations from previous work to avoid the influence of complementary color and environment background on the search task [19, 40]. As luminance can significantly impact pupil size and act as a confounding variable [24], we controlled for luminance in the Unity VR Editor, setting brightness at 0.7. We displayed the stimuli using a video see-through technique. Here, we used a pre-recorded video to provide the same real-world experience for all participants in the AR and AV conditions displayed via the same device. Moreover, to validate our approach, we measured luminance across conditions. We measured the light levels inside the headset using a lux meter sensor (LT300, Extech, USA). Using 20 measurements per condition, we found an AR in the average lux was 44.4 (SD=.9), 43.3 (SD=1.0) lux in AV, and 44.1 (SD=.9) lux in VR. Those values align with luminance guidelines (below 200 nits) based on eye tracking best practices to avoid confounders for pupil size computation [24, 109].

Participants were placed at a 1.5-meter distance from the fixation cross, resulting $5.72 \times 5.72 \times 5.72$ degrees visual angle in size. Regardless of the perceptual load condition (High vs. Low), we displayed objects within the participant's frustum, maintaining eccentricity stable (8°), allowing for greater ERPs component sensitivity and statistical power [121]. By displaying objects within the frustum, they are positioned within the participant's field of view, making them easily accessible and visible without requiring extensive head or eye movements, minimizing EEG artifacts. Additionally, presenting objects within the frustum aligns with users' natural viewing behavior, enhancing the study's ecological validity and ensuring a more realistic representation of visual search behaviors in everyday MR environments.

3.5 Apparatus and Data Recording

The MR Visual Search task was implemented in Unity (Version 2020.3.8 LTS). We presented the MR manifestations using an HTC VIVE Eye Pro headset with a display resolution of 2880×1600 pixels combined (Field of View: 110 °). For environment tracking, we use two HTC Vive lighthouses 2.0. We acquired two physiological measurements: EEG signal using a LiveAmp amplifier connected via Bluetooth (BrainProducts GmbH, Germany, 500 Hz) and eye tracking data via HTC Vive Pro Eye headset (120 Hz). Physiological data were streamed within the Unity VR environment within the Lab Streaming Layer (LSL) framework[1] to the acquisition PC (Windows 10, Intel Core i7-11700K, 3.60 GHz, 16GB RAM). For recording the real-life 360-video used in the AR and AV conditions, we used the Insta360 Pro with a resolution of 3840x1920 pixels (4K) with 120fps.

3.5.1 EEG Recording & Preprocessing. We acquired EEG data (sampling rate = 500 Hz) via LiveAmp amplifier (Brain Products, Germany) 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 ≤20 kΩ. We set the reference at FCz during the recording, while FPz was used as ground. Electrodes were placed 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 [61]. We first automatically detected bad or outliers channels via random sample consensus (RANSAC) method [11] of spherical splines for estimating scalp potential based on algorithms proposed by Perrin et al. [123]. We then applied a notch filter (50 Hz) and band-passed the signal between (1-30 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 [96]. We automatized the labeling and rejection process of ICA components via the MNE plugin 'ICLabel' [78, 97, 125]. Epochs that showed blinks, eye movement, muscle, or single-channel artifacts in any of the electrodes were rejected. Only

[1] https://github.com/labstreaminglayer/

trials with a correct response and last fixation on the target were used; error trials or distractor fixations were excluded from analyses.

3.5.2 FRP Analysis. For FRP analyses, we segmented continuous signals between 250 ms before and 500 ms after the onset of each fixation. For the trials considered, no additional saccades were present during this period. We chose this epoching based on previous FRP analyses and eye tracking research showing that EEG activity can be measured before saccade onset [86] but also posterior to the fixation onset [21]. Since the artifact of the previous saccade was restricted from 100 ms to 0 ms before the fixation onset, the baseline for each channel was defined between 250 ms and 100 ms before the onset of the current fixation and subtracted. For FRP computation, we chose O1, O2, and Oz electrodes based on previous work highlighting recruitment of occipital areas in visual processing tasks involving depth and spatial perception [114].

3.5.3 Eye Tracking Recording & Preprocessing. We collected the head position and directional 3D vectors from Unity. We recorded eye tracking (ET) data using the built-in eye tracker in the HTC Vive Pro Eye headset (120 Hz) through the SRAnipal eye tracking SDK, giving us eye-directional 3D vectors relative to both the world and the head. To facilitate our analysis, we initially converted the head directional 3D vector, eye-in-world directional 3D vector, and eye-in-head directional 3D vector into 2D Fick angles using the Fick-gimbal method [68]. This conversion involved two rotations, one about the vertical axis and the other about the nested horizontal axis, effectively characterizing the position of each vector. These resulting 2D Fick angles for eye and head directions were the basis for our subsequent analyses. For analysis, we focused on the eye tracking data for each trial until the participant selected the target. Unlike EEG, eye tracking is no longer relevant after the target selection. Given the relatively short duration of our trials, averaging approximately 2.71 ± 1.01 seconds across all trials, we conducted our eye tracking analysis at the trial level rather than examining individual behaviors within each trial.

3.5.4 Fixation-Saccade Analysis. We calculated fixation and saccades using pymovements [92], an open-sourced Python package for analyzing eye tracking data. We chose pymovements' implementation of the ID-T algorithm [141] with specific fixation thresholds set at a minimum fixation duration of 83ms and a maximum dispersion of 1.8 degrees [9, 158]. This allowed us to calculate the following fixation-based metrics: total fixation duration, average fixation duration, fixation count, and the time from the start of the visual search trial to the last fixation within the trial. For saccadic analysis, we employed pymovements' implementation of the microsaccade algorithm [51]. This enabled us to calculate two key saccade metrics: saccade amplitude and saccade frequency. Saccade amplitude was computed as the angular distance in degrees between the saccade onset and offset, while saccade frequency was determined by dividing the number of saccades within a trial by the trial duration.

3.5.5 Pupil Size Analysis. To compute the average pupil size for each trial [88, 118], we initiated the process by removing the pupil baseline for each eye. This baseline was established as the mean pupil size observed during the phase of the trial when the fixation cross was displayed. Subsequently, we computed the mean of the normalized left and right pupil sizes and then calculated the mean and standard deviation of the combined pupil size to facilitate our analysis at the trial level.

3.6 Participants

A sample of 24 participants voluntarily participated in the study recruited via institutional mailing lists, social networks, and convenience sampling. The number of participants recruited for our study aligns with recent investigations on the relationship between participants number and EEG data reliability for relatively long tasks [159] and in line with previous work in HCI [42], visual search [56] and physiological computing domains [35, 132]. Three participants were excluded due to inadequate EEG data quality, as identified by RANSAC, which revealed that over 50% of electrodes for computing, FRPs (01, 02, Oz) were classified as "bad," compromising data reliability due to a low SNR as described in Bigdely-Shamlo et al. [11]. This led to a total of 21 participants (M = 25.57, SD = 2.64; 8 female, 13 male, none diverse). We surveyed participants' familiarity with AR, AV, and VR devices as in previous work [30]. All participants reported prior experience with AR (M = 3.76, SD = 1.6), with AV (M = 2.45, SD = 1.07) and VR (M = 4.5, SD = 1.07) on a scale from 1 (not at all familiar) to 7 (extremely familiar). None of the participants reported a history of neurological, psychological, or psychiatric symptoms. The study met the criteria for fast-track conditions set by the local institutional ethics board, i.e., participants were not exposed to any risks such as deception, excessive stress, or recording of sensitive information.

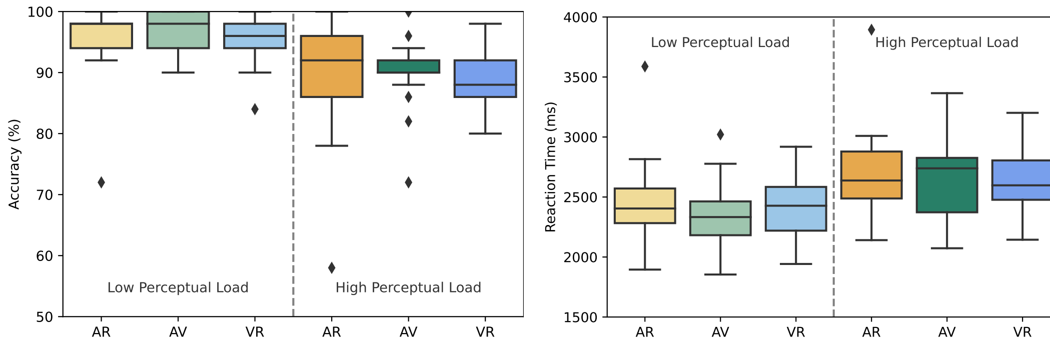
4 Results

In this section, we first present results on behavioral accuracy and reaction times, FRPs analysis on average peak amplitude for P3, eye tracking features, and subjective scores on perceived workload (NASA-TLX) and ad-hoc UX surveys. We employ a Generalized Linear Mixed Model (GLMM) to investigate differences in the reaction time, FRPs, and eye tracking feature distributions. We report the effect sizes as odds ratios for behavioral accuracy as a dichotomous variable, which reflect the magnitude of effect for each predictor in the model [82]. For continuous variables (FRPs, reaction times, eye tracking features), we employ the delta total (δ_t) as our effect size measure [69]. For subjective scores, upon the normality, using the Shapiro-Wilk test, we use two-way ANOVAs for parametric analysis and ART ANOVAs [167] for the non-parametric data. Furthermore, for post-hoc comparisons, we use either t-test or ART-C test [50] to report our results. For the ANOVA models, we computed effect sizes using Eta squared (η^2), and for ART ANOVAs, we first applied the ART procedure and then calculated Eta squared on the ranked data. Lastly, we investigated potential fatigue or learning effects due to the repeated measures structure of our experimental design. Statistical analysis focused on identifying significant linear or non-linear trends in the data. We specifically looked for increases in RT, visual search accuracy, and NASA-TLX scores, which could indicate rising fatigue, or a decrease in these metrics, which might suggest learning effects. We report these results in Section A.1 and make them openly available in Section 7.

4.1 Behavioral Data

4.1.1 Accuracy. Given its dichotomic nature, response accuracy at each trial was analyzed by conducting a Generalized Linear Mixed Model (GLMM) with logit link function using the glmer function from the lme4 library [7]. Behavioral responses and conditions were entered into the model as fixed effects while a random intercept varying among participants was entered into the model as random effects, leading to the formula: `accuracy ∼ Manifestation * PerceptualLoad + (1|participant)`.

A linear mixed-effects analysis was conducted to assess the influence of Manifestation and PerceptualLoad on the accuracy, including a random intercept for participants. Manifestation levels AV and VR did not show significant differences from AR, with b = .0092, $t(138)$ = 1.012, p = .313, and b = .005, $t(138)$ = .598, p = .551, respectively. Low Perceptual Load significantly differed from the difficult level (b = .0686, $t(138)$ = 9.276, p < .001). These findings suggest that while PerceptualLoad has a significant impact on accuracy, the modality of mixed reality (Manifestation) did not exert a significant effect in the given conditions. Specifically, accuracy was higher when the perceptual load was low than the high perceptual load condition. We observed the following odds ratio: in *AV*, the odds ratio was 1.01, 95% CI [0.99, 1.03], suggesting that a one-unit increase in *AV* is associated with a 1.1% increase in the odds of the outcome. Similarly, for *VR*, the



(a) Behavioral Accuracy

(b) Reaction Times

Fig. 5. Boxplot of accuracy and reaction times. a) The figure presents accuracy, showing no significant differences between AV and VR compared to AR, but a notable increase in low perceptual load conditions. b) The figure shows reaction times, with significant reductions in AV and VR compared to AR, and markedly faster in low perceptual load conditions.

odds ratio was approximately 1.00, 95% CI [.98, 1.02], indicating a marginal change in the odds with a one-unit increase. Additionally, for *Low Perceptual Load*, the odds ratio was 1.07, 95% CI [1.05, 1.09], showing a 7% increase in the odds of the outcome compared to the high perceptual load condition.

4.1.2 Reaction Times. Raw RTs were analyzed using a Generalized Linear Mixed Model (GLMM) using lme4 library [7]. Error trials were excluded from the analysis (formula: `RTs ~ Manifestation * PerceptualLoad + (1|participant)`). Overall, we computed 276.71 average amount of trials (SD = 12.84) per participant. We chose this approach because, besides considering variability across individuals, it allows us to control for many longitudinal effects during the task without transforming the data [100]. First, there are the effects of learning and fatigue [6]. Second, the response in a trial is usually heavily influenced by what happens in the preceding trial [93, 106, 168]. Sources of experimental noise are brought under statistical control using mixed-effects models.

A linear mixed-effects analysis was conducted to assess the influence of Manifestation and PerceptualLoad on reaction time (RT), including a random intercept for participants. For the Manifestation, both AV and VR levels exhibited significant reductions in reaction time compared to the AR level: AV ($b = -56.20$,, $t = -2.400$, $p = .017$) and VR ($b = -70.63$, $t = -3.01$, $p = .003$). Additionally, low PerceptualLoad led to a significant decrease in reaction time compared to the difficult level ($b = -259.70$, $t = -13.55$, $p < .001$). These findings indicate that PerceptualLoad and Manifestation influence reaction time. Overall, participants reacted faster in the low PerceptualLoad condition and responded more quickly in the AV and VR conditions than in AR. The model's δ_t indicated significant effects for AV ($b = -.07$, 95% CI [-.12, -.01]) and VR ($b = -.08$, 95% CI [-.14, -.03]). Moreover, PerceptualLoad in the 'Low' condition showed a substantial effect ($b = -.31$, 95% CI [-.35, -.26]), denoting its large role among the predictors.

4.2 EEG Data - Fixation Related P3

A linear mixed-effects analysis was utilized to test the impact of Manifestation and PerceptualLoad on the P3 amplitude post-fixation onset (see Figure 6), integrating a random intercept for individual participants (formula: `P3Amplitude ~ Manifestation * PerceptualLoad + (1|participant)`). The model held significant explanatory power with a conditional R^2 of .44.

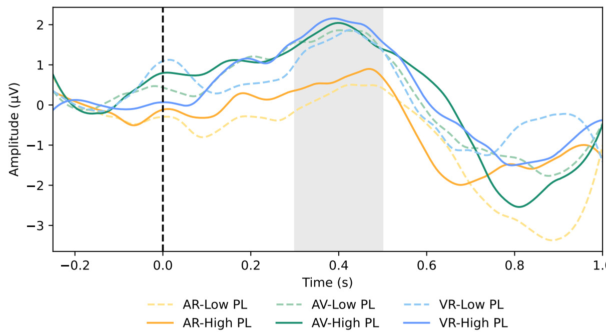


Fig. 6. Grand Average event-locked to fixation onset. Data reflect the results obtained from parieto-occipital ROI for each MR and PerceptualLoad condition. The plot suggests a pronounced MR Step influence on P3 amplitude, with marked variations between AV and VR compared to AR.

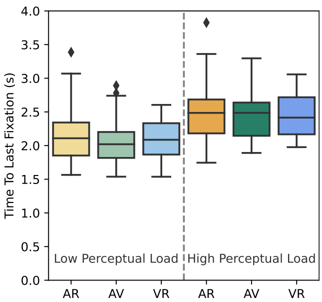
Within this model, the AV and VR levels of Manifestation significantly showed increased P3 amplitude, demonstrated by beta values of 2.88 (t = 2.59, p = .011) and 3.20 (t = 2.88, p = .005), respectively. In contrast, the low PerceptualLoad yields a non-significant and minor negative effect, with a beta of -.42 ($t = -.46$, p = .645). These findings highlight the main effect of the Manifestation on allocating attentional resources during visual search tasks, as reflected in the P3 amplitude. Specifically, the AV and VR levels were associated with heightened attentional engagement, potentially facilitating more efficient visual search processes compared to the AR level. In contrast, the PerceptualLoad appeared to have a negligible impact on the resource allocation in visual search. The model's δ_t indicated significant effects for *AV* (b = .43, 95% CI [.10, .76]) and *VR* (b = .48, 95% CI [.15, .81]). Moreover, *Perceptual Load* (Low) showed a smaller effect ($b = -.06$, 95% CI [-.33, .21]), though its role among the predictors was not as pronounced.

4.3 Eye Tracking Data

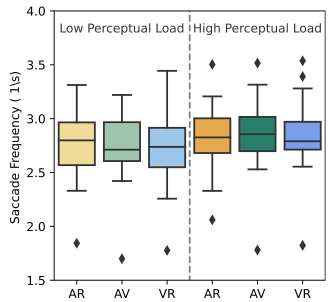
To compute the average pupil size for each trial [88, 118], we initiated the process by removing the pupil baseline for each eye. This baseline was established as the mean pupil size observed during the phase of the trial when the fixation cross was displayed. Subsequently, we computed the mean of the normalized left and right pupil sizes and then calculated the mean and standard deviation of the combined pupil size to facilitate our analysis at the trial level.

We used a general linear mixed-effects modeling to evaluate the effect of the independent variables, i.e., Manifestation and PerceptualLoad on the set of eye tracking features (formula: `ET feature ~ PerceptualLoad * Manifestation + (1|participant)`). The analysis was conducted using the REML method, and here, we highlight the significant findings on the fixed effects and interactions. Overall, no significant interactions were detected.

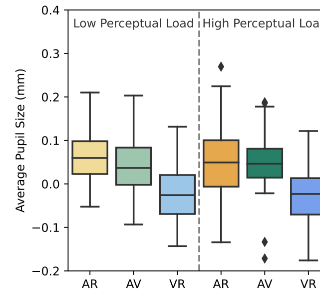
4.3.1 Fixation Duration. Analysis revealed that the effect of Manifestation [AV] on fixation duration was not statistically significant (beta = -.03, 95% CI [-.07, .01], $t(7191) = -1.35$, p = .179), with an average fixation duration of 1.78 seconds (SD = .684) for Low and 2.06 seconds (SD = .777) for High Perceptual load conditions. However, the effect of Manifestation [VR] was found to be statistically significant and negative (beta = -.04, 95% CI [-.08, -2.24e-03], $t = -2.07$, p = .038), where the average fixation durations were 1.77 seconds (SD = .66) for Low and 2.05 seconds (SD = .775) for High Perceptual conditions. This suggests that the VR condition led to shorter fixation durations than the AV condition. Additionally, the effect of low PerceptualLoad was statistically



(a) Time to Last Fixation



(b) Average Saccade Frequency



(c) Average Pupil Size

Fig. 7. Boxplot on selected eye tracking features. The left figure illustrates the Time to Last Fixation, showing faster times in low perceptual load and a slight increase in the AV condition. The center figure presents the Average Saccade Frequency, with minimal visual differences between conditions. The right figure depicts Average Pupil Size, revealing a notable decrease from AR to VR, particularly under varying perceptual loads. Error bars represent the standard deviation from the mean.

significant and negative (beta = -.28, 95% CI [-.31, -.25], $t = -17.51$, $p < .001$). This indicates that lower perceptual load is associated with shorter fixation durations, as evidenced by the average fixation durations of 1.77 seconds (SD = .660) for VR easy and 1.78 seconds (SD = .684) for AV easy conditions, compared to 2.05 seconds (SD = .775) for VR difficult and 2.06 seconds (SD = .777) for AV difficult conditions. In terms of standardized effect sizes, for the AR condition was .22, 95% CI [.06, .38], serving as the baseline for comparison. The δ_t for AV was -.04, 95% CI [-.09, .02], indicating a slightly shorter fixation duration compared to AR. For VR, the δ_t was more pronounced at -.05, 95% CI [-011, .00], suggesting a greater reduction in fixation duration relative to AR. The effect of low PerceptualLoad was the most substantial (δ_t = -.37, 95% CI [-.42, -.33]), indicating a significant decrease in fixation duration compared to the high perceptual load.

4.3.2 Time to Last Fixation (T2LF). Results showed that both Manifestation and PerceptualLoad significantly influence T2LF. Manifestation effect was significant and negative, with AV and VR conditions showing a shorter time to last fixation, reducing the response by -.057 (SE = .027), $t = -2.114$, $p = .035$, and -.072 units (SE = .027), $t = -2.643$, $p = .008$, respectively, compared to the AR condition. Regarding PerceptualLoad, the Low PerceptualLoad showed a negative significant effect, significantly decreased by -.385 units (SE = .022), $t = -17.394$, $p < .001$. The standardized effect sizes further elucidate the impact of each variable. Here, the δ_t for AV was -.05, 95% CI [-.12, .02], indicating a modest but not statistically significant reduction in T2LF compared to AR. For VR, the δ_t was -.08, 95% CI [-.16, -.01], suggesting a more pronounced and statistically significant reduction in T2LF compared to AR. The effect of Low PerceptualLoad was the most substantial, with a δ_t of -.38, 95% CI [-.46, -.31], highlighting its strong negative influence on T2LF. This suggests that conditions with lower perceptual load significantly shorten the time to last fixation.

4.3.3 Fixation Count. In the fixed effects analysis for Manifestation factor, both the AV and VR levels induced a non-significant negative effect on Fixation Count, decreasing it by -.17 units (SE = .123), $t = -1.41$, $p = .158$ and -.15 units (SE = .123), $t = -1.24$, $p = .214$, respectively. Furthermore, the low level of PerceptualLoad emerged with a significant negative impact on Fixation Count, reducing it by -1.87 units in comparison to the 'difficult' level (SE = .1), $t = -18.68$, $p < .001$. Regarding standardized effect sizes, the AR condition was at .23, 95% CI [.10, .36], serving as the baseline for comparison. The δ_t for AV was -.04, 95% CI [-.09, .01], and for VR it was -.03, 95% CI [-.09, .02], indicating slight but non-significant decreases in Fixation Count compared to AR. The δ_t for low PerceptualLoad was notably more substantial at -.41, 95% CI [-.45, -.37], demonstrating a significant reduction in Fixation Count compared to the high perceptual load.

4.3.4 Saccade Frequency. For Manifestation, both the AV and VR levels did not show any significant positive effects on Saccade Frequency, increasing it by .01 units (SE = .017), t = .79, p = .432 and .0044 units (SE = .017), t = .26, p = .798, respectively. In contrast, the PerceptualLoad factor demonstrated a significant and negative influence when the PerceptualLoad was low, decreasing Saccade Frequency by -.094 units compared to the high PerceptualLoad level (SE = .014), $t = -6.70$, $p < .001$. This significant negative effect indicates that participants tend to have fewer saccades when tasks are less perceptually demanding, potentially reflecting a reduced workload in these conditions. Regarding the standardized effect sizes: the AR condition, was at .06, 95% CI [-.13, .25]. The δ_t for AV was .02, 95% CI [-.03, .07], and for VR it was approximately .0066, 95% CI [-.04, .06], both showing non-significant changes in Saccade Frequency compared to AR. Conversely, the δ_t for low PerceptualLoad was -.14, 95% CI [-.18, -.10], marking a significant decrease in Saccade Frequency compared to the high perceptual load, thus aligning with the observation of fewer saccades in less demanding conditions.

4.3.5 Saccade Amplitude. When testing Manifestation factor fixed effects, the AV level presented a non-significant negative influence on saccade amplitude, reducing it by -.16 units (SE = .094), $t(7188) = -1.69$, p = .090. In the same direction, VR level was non-significant, $t = -.06$, p = .951. PerceptualLoad factor was significant and positive when the level was Low, increasing saccade amplitude by .82 units compared to the 'difficult' level (SE = .077), $t(7188) = 10.77$, $p < .001$. This substantial positive effect suggests that participants have larger saccade amplitudes when the tasks are easier, which might indicate a broader or more relaxed visual exploration strategy under lower PerceptualLoad conditions. When investigating the standardized effect sizes, AR condition was at -.10, 95% CI [-.25, .04]. The δ_t for AV was -.05, 95% CI [-.10, .01], and for VR it was approximately -.0017, 95% CI [-.05, .05], both showing minimal changes in saccade amplitude compared to AR. However, the δ_t for low PerceptualLoad was .24, 95% CI [.19, .28], demonstrating a significant increase in saccade amplitude under less demanding conditions.

4.3.6 Average Pupil Size. For Manifestation, both AV and VR levels significantly reduced mean pupil size. The AV level led to a decrease of -.01 units, $t(7191) = -2.57$, $p = .010$, while the more pronounced VR level decreased it by -.08 units, $t(7191) = -15.64$, $p < .001$. These findings suggest varying workloads or visual engagements across the MR continuum. In contrast, the Low PerceptualLoad did not significantly affect mean pupil size, with a change of .002 units, t = .44, p = .661. We computed the standardized effect sizes, and we found that for AR condition, it was .16, 95% CI [.02, .29], indicating the baseline mean pupil size. The δ_t for AV was -.07, 95% CI [-.12, -.02], showing a moderate reduction in pupil size compared to AR. However, the effect was more substantial for VR, with a δ_t of -.42, 95% CI [-.47, -.37], indicating a significant decrease in pupil size. Meanwhile, the effect of low PerceptualLoad on pupil size was negligible (δ_t = .0096, 95% CI [-.03, .05]), reinforcing the finding of its non-significant impact on mean pupil size.

4.4 Subjective Data

We analyzed perceived workload reports via the NASA-TLX questionnaire [67] via either two-way ANOVAs or ART ANOVAs [167] depending on normality testing. We additionally collected three 7-point Likert-scale responses on UX, related to how *Distracting* was the background, how the number of objects was *Overwhelming*, and lastly, how much the background impacted the task difficulty, i.e., *Difficult*.

4.4.1 Perceived Workload. The raw NASA TLX scores deviated from normality as per Shapiro-Wilk normality testing $W = .972$, $p < .001$. An ART ANOVA was conducted and showed a significant main effect of Manifestation on perceived workload, $F(2, 100) = 10.273$, $p < .0001$. This suggests that the Manifestation differentially impacts perceived workload. The main effect of PerceptualLoad was also significant, $F(1, 100) = 76.179$, $p < .0001$, indicating that workload perception varies significantly between high and low PerceptualLoad. Additionally, there was a significant interaction between Manifestation and PerceptualLoad, $F(2, 100) = 5.517$, $p < .001$, suggesting that the effect of interaction modality on perceived workload is not consistent across different levels of PerceptualLoad. Post hoc contrasts for Manifestation revealed a significant difference in perceived workload between AR and AV, $t(100) = 3.45$, $p < .01$, with an increased workload for AR condition. In the same direction, we found a significant difference between AR and VR, $t(100) = 4.278$, $p < .001$. The comparison between AV and VR did not show any significance$t(100) = .841$, $p > .05$. Regarding PerceptualLoad, there was a significant difference in perceived workload between the high and low PerceptualLoad conditions, with the first showing increased workload, $t(100) = 8.728$, $p < .0001$. We then compute the effect size as partial eta squared to elucidate the impact of each factor. For Manifestation, the effect size was substantial ($\eta^2_{\text{partial}} = .17$, indicating that this factor accounted for approximately 17% of the variance in perceived workload. The effect size for PerceptualLoad was even more pronounced ($\eta^2_{\text{partial}} = .43$, explaining about 43% of the variance and underscoring its significant impact on workload perception. Additionally, the interaction between Manifestation and PerceptualLoad had an effect size $\eta^2_{\text{partial}} = .10$, suggesting a meaningful but less pronounced impact compared to the main effects.

4.4.2 User Experience.

Distracting. Participants' ratings on the 7-point Likert scale item "*The background made it very difficult to distinguish the target object from other objects*" deviated from normality as indicated by the Shapiro-Wilk normality test, $W = .845$, $p < .001$. ART ANOVA detected a series of significant results. We found a significant main effect of Manifestation on participants' ratings, $F(2, 100) = 39.669$, $p < .001$. The main effect of PerceptualLoad was also significant, $F(1, 100) = 36.148$, $p < .001$. Finally, there was a significant interaction between Manifestation and PerceptualLoad, $F(2, 100) = 5.24$, $p < .001$. Post hoc contrasts for Manifestation demonstrated a significant difference between AR and AV $t(100) = 4.546$, $p < .001$. There was also a significant difference between AR and VR, with the same direction $t(100) = 8.857$, $p < .001$. Lastly, comparing AV and VR again yielded a significant difference $t(100) = 3.617$, $p = .0002$. Regarding PerceptualLoad, participants rated the difficulty in distinguishing the target object as significantly higher in the high PerceptualLoad condition than in the Low condition, $t(100) = 6.012$, $p < .001$. The effect sizes analysis revealed the substantial impact of each factor. The partial Eta squared for Manifestation was .44, suggesting that this factor accounted for approximately 44% of the variance in the ratings, indicating a very strong effect. The effect size for PerceptualLoad was .27, explaining about 27% of the variance, denoting a significant influence on how participants perceived distractions. Additionally, the interaction between Manifestation and PerceptualLoad had an effect size of .31, highlighting a considerable combined impact of these factors on participants' perception of distractions.

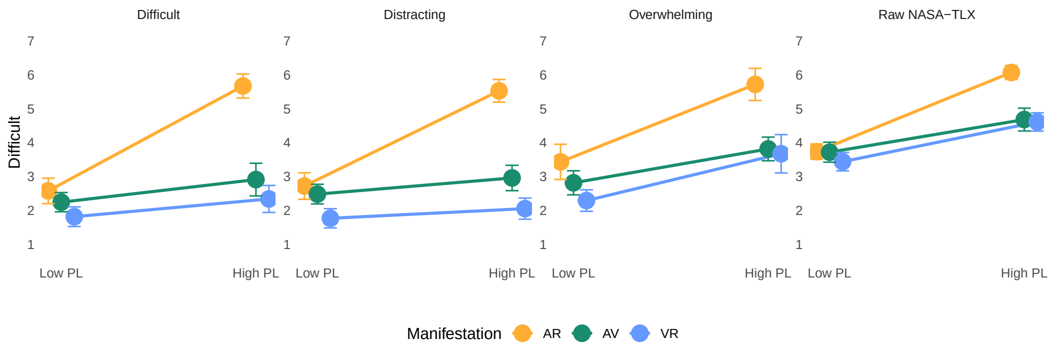


Fig. 8. *Subjective Data.* Line plots for Raw NASA-TLX, UX Difficult, UX Distracting, UX Overwhelming 7-point Likert scales (normalized to 7-points only for graphical illustration). Here, we found the main effects of Manifestation and PerceptualLoad across all subjective measures. Error bars are displayed as standard error from the mean.

Overwhelming. Participants' ratings on the number of items perceived as overwhelming deviated from normality according to Shapiro-Wilk $W = .853$, $p < .001$. An ART ANOVA was performed to analyze the aligned rank-transformed data. There was a significant main effect of Manifestation on the difficulty ratings, $F(2, 100) = 17.752$, $p < .001$. Post hoc contrasts revealed that participants in the AR condition rated it significantly harder to distinguish the target object than those in the AV condition $t(100) = 3.829$, $p < .001$. Similarly, the AR condition was rated as significantly more difficult than the VR condition $t(100) = 5.868$, $p < .001$. The AV and VR conditions did not differ, with participants reporting being subjectively overwhelmed similarly $t(100) = 2.038$, $p = .13$. This suggests that the type of Mixed Reality (MR) step influenced participants' perception of difficulty, with the AR environment being perceived as the most challenging. The effect of PerceptualLoad was also significant, $F(1, 100) = 52.14$, $p < .001$. Participants rated the task as significantly more challenging in the high PerceptualLoad condition than in the low PerceptualLoad condition, with an estimated difference of 33, $t(100) = 7.22$, $p < .001$. This indicates that irrespective of the MR step, tasks with a high PerceptualLoad were perceived as inherently more difficult. Furthermore, there was a significant interaction between Manifestation and PerceptualLoad, $F(2, 100) = 3.919$, $p < .05$. The partial Eta squared for Manifestation was .26, indicating that approximately 26% of the variance in difficulty ratings can be attributed to the Manifestation factor, suggesting a strong influence. The effect size for PerceptualLoad was even more significant at .34, accounting for about 34% of the variance and highlighting its considerable impact on perceived task challenge. The interaction effect between Manifestation and PerceptualLoad had an effect size of .07, denoting a smaller but still noteworthy combined influence of these factors on participants' perception of being overwhelmed.

Difficult. Scores on the perceived difficulty item due to the MR background deviated from normality as determined by the Shapiro-Wilk normality test, $W = .820$, $p < .001$. An ART ANOVA first detects a main effect of Manifestation on the difficulty ratings, $F(2, 100) = 30.621$, $p < .001$. Post hoc comparisons elucidated that participants perceived the AR condition as significantly more difficult than the AV condition, $t(100) = 5.547$, $p < .001$. Similarly, distinguishing the target object in the AR environment was deemed significantly more challenging than in the VR environment $t(100) = 7.554$, $p < .001$. The AV and VR conditions showed no perceptible differences

$t(100) = 2.006$, $p = .14$. This suggests that the choice of MR step distinctively influenced participants' perceptions, with the AR modality consistently emerging as the most taxing. Regarding PerceptualLoad, there was a clear distinction between the high and low PerceptualLoad conditions. Participants perceived the high PerceptualLoad tasks as considerably more challenging than their low PerceptualLoad counterparts $t(100) = 7.366$, $p < .001$. This shows the influence of task complexity on participants' perceptions, independent of the MR environment. Lastly, the interaction between Manifestation and PerceptualLoad was significant, $F(2, 100) = 19.248$, $p < .001$. This interaction implies that the effects of MR step on perceived difficulty were not uniform but contingent upon the task's inherent PerceptualLoad.

The partial Eta squared for Manifestation was .38, indicating that Manifestation accounted for approximately 38% of the variance in perceived difficulty, marking a substantial effect. The effect size for PerceptualLoad was .35, explaining about 35% of the variance and highlighting its significant role in shaping difficulty perceptions. Additionally, the interaction effect between Manifestation and PerceptualLoad had an effect size of .28, demonstrating a combined influence on participants' perceptions of difficulty.

5 Discussion

In this work, we evaluated the impact of different MR manifestations and task difficulty on behavioral and physiological correlates of visual search efficiency, resource allocation, and perceived workload.

5.1 Summary of Results

In our study investigating the effects of different MR environments on visual search tasks, we observed notable differences in cognitive load and visual search efficiency across AR, AV and VR. The behavioral data analysis revealed that perceptual load significantly impacts search accuracy and reaction times, with lower perceptual loads leading to higher accuracy and faster responses. Notably, participants in the AV and VR environments demonstrated quicker reaction times than those in the AR setting, suggesting that these environments may facilitate faster cognitive processing. FRP analysis highlighted increased P3 amplitudes in AV and VR compared to AR. This suggests a greater demand for cognitive resources in AR when contrasted with AV and VR, potentially contributing to the more efficient processing observed in these environments. Eye tracking metrics supported these findings, showing shorter fixation durations and quicker times to last fixation in VR, which suggest a more efficient visual search process. Furthermore, the average pupil size was smaller in VR, indicating lower cognitive load due to decreased visual complexity than AR. Further, participants reported finding the AR environment more demanding and distracting, consistent with the objective measures of increased cognitive load. These results show how AV and VR environments might require less cognitive resources in visual search tasks, compared to AR.

5.2 Impact of Manifestation on User Performance

With our first research question (**RQ1**), we wanted to investigate whether our independent variable Manifestation affects users' performance in a visual search task. In particular, we were interested in objective performance measures, such as accuracy and reaction times, as well as subjective measures, such as perceived workload and individual task perception. While we did not find any differences in accuracy, we found that participants responded faster in the AV and VR conditions than in AR. Given the short duration of our search task (5s), drops in accuracy can mostly be attributed to misses rather than false positives. Thus, it is likely that we introduced a ceiling effect, and longer task durations would show differences in accuracy as well. Nonetheless, our findings of significantly different reaction times are consistently supported by our subjective measures (perceived workload and individual task perception), which depicted how challenging and demanding the AR condition was compared to the others. Participants expressed that they found the background more distracting and the overall task relatively more overwhelming in the AR environment. Compared to previous studies that explored concrete manifestations individually (e.g., in VR [42], in AR [111], or even outside of MR [137]), our work can address the gap of comparing user performance between manifestations.

Given the decreased performance of the AR condition, applying search-related AR applications may be problematic, particularly in safety-critical domains such as healthcare, aviation, or emergency response [72, 79], where a slightly delayed response can have serious consequences. Validating our experimental design, a low perceptual load resulted in higher accuracy rates and shorter reaction times than did a high perceptual load. Considering this, designers may need to strike a delicate balance in MR applications so that augmentation enhances the user's performance without overwhelming their cognitive capacities [44]. However, we do not advocate for the exclusion of AR in general. AR has specific use cases where physical aspects of the environment cannot be occluded, making it indispensable [12]. For scenarios where blending more virtuality in AR is not feasible, the trade-off should prioritize user safety while maintaining the necessary visibility and awareness of the real world. In remote or less critical scenarios, enhancing the virtual components to shift towards a more AV environment might be beneficial [110]. The decision to use AR or AV depends significantly on the context and requirements of the task, ensuring that while we aim to minimize potential distractions and cognitive overload, we do not remove the essential elements that AR brings to user interactions in the real world.

5.3 Cognitive Resource Allocation per Manifestation

To reiterate **RQ2**, we were interested in resource allocations across the different manifestations. In FRPs, we found increased P3 amplitude, indicating enhanced attentional and more efficient visual search processes in both the AV and VR conditions compared to AR. First, we confirm that FRPs are elicited across MR manifestations. FRPs are grounded in natural scanning behaviors, specifically the act of fixating on relevant information, a phenomenon central to visual search tasks and routinely manifested in everyday settings. Second, we show that the environmental visual feature of MR manifestations influences task demands and can affect electrophysiological correlates of resource allocation in visual tasks. Previous ERP studies have shown late (P3) visual ERPs during high auditory working memory load compared to low auditory working memory load [131] or when users are asked to ignore distracting auditory stimuli [59].

This aligns with insights from Seijdel et al. [142], which highlight the critical role of recurrent processing in navigating complex visual environments. Their work demonstrated that increased background complexity leads to more recurrent processing, which is crucial for object segregation and recognition. This is particularly relevant for the AR environment in our study, characterized by its richer textures and variable visual noise compared to the more controlled VR settings. Their findings indicate that as background complexity increases, so does the reliance on deeper, more recurrent neural processing. ERPs peaks can be delayed and diminished when participants are engaged in tasks within more complex backgrounds [62, 134]. These ERPs alterations suggest that the brain's capacity to process information efficiently is taxed more in complex settings, leading to potential delays in cognitive processing and increased cognitive load. This could explain the distinct cognitive demands and performance differentials we observed in the AR environment, reinforcing the need to consider visual complexity in designing MR systems that aim to support optimal allocation of attentional resources [33, 34, 101].

Moreover, our results replicate previous work from Ries et al. [138], demonstrating how fixation-related P3 can discriminate task demands, extending them to MR, particularly in its AR manifestation. This reinforces the notion that perceptual load and manifestation types significantly influence cognitive resource allocation in MR environments, as evidenced by the elicited FRPs. Moreover, this outcome is more relevant as it presents a novel contribution by being the first to elucidate FRPs across different MR environments. This result is particularly significant as FRPs are ecological ERPs, offering practical insights into real-world cognitive processing. This finding has significant implications for the design of MR systems, suggesting that optimizing attentional resources should be a key consideration across the MR continuum. For example, in environments where FRPs indicate a higher need for cognitive resources, designers might seek to reduce complexity or introduce features to aid information processing. These features could include directional cues, shifting the user's attention [64], or contextual highlighting and filtering, limiting the visual complexity [31].

5.4 Eye Tracking Correlates of Visual Search Efficiency

To approach **RQ3**, we considered the evidence of the impact of the manifestation on our eye tracking correlates of visual search efficiency. Drawing on foundational studies in visual search, such as those by Duncan and Humphreys [46], we know that stimulus similarity can influence search efficiency. Their work suggests that as target and distractor similarity increases, the search becomes more difficult. Thus, one would assume that searching for virtual objects in AV is easier than searching for them in VR because of the increased perceptual consistency, which could slow the search process [28, 140, 146]. However, VR fostered more efficient visual processing than AV, mainly evidenced by shorter fixation durations. Such differences might suggest that transparency and see-through objects in AR and AV elevate the attentional load, possibly due to their decreased visual salience [170]. Further integrating the Guided Search theory, it's plausible that the VR and AV modes streamlined the preattentive processing stages, guiding attention more effectively than in the AR mode. Regarding pupil size, in contrast to the abundant literature that often associates higher perceptual load with increased pupil dilation [20, 116], we observed no significant changes across the MR continuum. This could suggest a ceiling effect, wherein the perceptual load across conditions was high enough to maintain a consistent pupil size. However, even if pupil size did not differ between PerceptualLoad load levels, the Low condition was associated with faster visual processing. This was evidenced by shorter fixation durations and shorter times to the last fixation, potentially reflecting reduced workload and enhanced search efficiency. This trend of reduced cognitive demand with lower perceptual loads was further manifested in the fixation count and saccade frequency metrics, where participants exhibited fewer fixations and saccades, possibly because they adopted a more relaxed visual exploration strategy.

5.5 Implications for Adaptive Systems

Taking it one step further, EEG data not only holds the potential to study the electrophysiological correlates of attention but can also be used for adaptive systems [101, 115**?**]. FRPs have recently been employed as a feature to differentiate between relevant and non-relevant information in single-trial classification [144, 166]. Most current BCIs rely on stimulus-evoked signals for target identification, typically within the confines of fixed-location stimulus displays [115, 163]. In contrast, Finke et al. [55] demonstrated the potential of BCIs based on FRPs, which offer a more flexible approach. FRP-based BCIs are not constrained by stimulus-dependent tasks, allowing them to operate effectively in more complex and dynamic environments. Drawing from our results, the variation in FRP amplitude across MR manifestations indicates that users' resources and processing efficiency differ between AR, AV, and VR environments. This outcome could guide the development of adaptive interfaces by providing cues to the system on which objects in a scene are relevant to the user, allowing for modulating content or background virtuality according to the detected resource capacity.

Building on this, the overarching goal is to develop transitional interfaces capable of seamlessly switching between different points on the MR continuum [4, 161]. Previous transitional interfaces found applications for visual analytics [75], collaborative learning, and manipulation tasks [139], yet their application in MR visual search represents a novel direction. Here, we envision interfaces that could dynamically switch across AR, AV, and VR by utilizing FRP amplitude variations adapting demands, task features, and environment. Here, we argue that these aspects equally influence UX in MR interactions.

However, it is important to acknowledge that BCIs in MR systems can face specific challenges. A key consideration is that users might not wear MR headsets for prolonged sessions due to their still-to-be-improved usability and ergonomics [41, 110]. Moreover, the ongoing endeavor to embed EEG headsets in VR and AR largely depends on dry setups [10], with unclear outcomes on signal quality and its impact on high-accuracy classification [10], even though high-density dry setups appear promising for application and performance [54].

5.6 Limitations and Future Work

In our research, we mainly investigated the effects of the general MR environment. While our study is insightful into cognitive load and visual search within MR environments, the specificity of visual search tasks—which primarily involve identifying and locating visual targets—may not directly translate to other cognitive functions that involve different types of processing. For instance, perceptual decision-making [108] involves complex evaluations and choices among multiple options, which might be affected differently by the cognitive load induced by MR technologies. Similarly, divided attention tasks [1, 143], which require managing multiple stimuli simultaneously, could interact with MR in ways not observed in single-focus visual search tasks. Furthermore, multiple object tracking [122], involving continuous monitoring and memory engagement for several moving objects, presents dynamic challenges that MR might influence distinctively compared to the static or less complex demands of visual search. These cognitive tasks each involve differential selection processes that could respond to MR environments in ways not fully captured by our current findings, underscoring the need for future research to explore these varied interactions comprehensively. MR encompasses a continuum of environments with varying levels of virtuality, ranging from single virtual objects to fully immersive backgrounds. Future work should explore these different manifestations, investigating how varying degrees of virtuality within each manifestation influence user perception and interaction.

Moving forward and examining MR interactions at a finer level, we propose investigating the dual nature of targets and distractors encompassing virtual and physical objects. Previous work has shown that real-world objects evoke more sustained, stronger ERP responses than 2D images, but it is unclear if this also applies to virtual objects or FRPs [107, 114]. Moreover, our study's representation of the MR manifestations world was relatively simplistic, predominantly involving virtual stimuli. Even if our stimuli allowed for a more controlled setting, future work should aim for a more realistic simulation by incorporating visual clutter, whether it be virtual or physical [36, 110]. Previous work has shown that real-world objects evoke more sustained, stronger ERP responses than 2D images, but it is unclear if this also applies to virtual objects or FRPs [107, 114].

To ensure reliable and robust eye tracking results, we controlled that presented stimuli had consistent luminances and transparency. However, to design an adaptive system, we propose investigating varying luminance levels, specifically in AR, to evaluate the effect on pupil size and if such results can be applied in more ecological contexts. To broaden our understanding of attentional processes within MR environments, future studies should incorporate the analysis of the N2 component, particularly by designing experiments that include both target-present and target-absent conditions [103, 104]. This approach would enable us to explore the N2 component's role in attentional engagement and disengagement and understand if users differentially locate virtual and physical target information in MR. Such studies could employ varied background complexities and types of stimuli (physical/virtual) to delineate how the presence or absence of targets affects cognitive load and attention allocation.

Regarding transparency, we chose a video-see-through approach to avoid potential confounders associated with room luminance that are more prevalent in optical see-through MR systems. Instead, in optical see-through AR, the light from virtual surfaces is a blend of display-emitted light and environmental light, leading to semitransparent appearances distinct from real-world objects. This difference can influence the saliency of the virtual content [37] and be associated with increased temporal uncertainty in FRPs which may impact the accuracy of attention decoding in MR [157]. Moving forward, future research in optical see-through MR could explore how changes in ambient lighting affect the perception of transparency in virtual objects as compared to video see-through. This might involve studying how varying levels of environmental illumination influence the saliency of virtual content and how users adapt their visual attention in response to these changes. Such investigations could inform the development of adaptive optical see-through MR systems that adjust the rendering of virtual objects to optimize their visibility and coherence with the real-world environment. We found a difference between the main effect on accuracy and the effects of MR Manifestation and PerceptualLoad, which we interpret as a possible change in the visual search strategy. To verify, we propose investigating the performance trade-off via cognitive modeling using the Drift Diffusion Model (DDM) [152]. DDM enables simultaneous analysis of accuracy and reaction times to extract parameters that underlie performance trade-offs. Thus, it already applies to visual search [171] and HCI [32].

6 Conclusion

We presented a systematic investigation of how users detect target information across the reality-virtuality continuum in a visual search task. We applied a multimodal evaluation using behavioral, subjective, and physiological (EEG and eye tracking) measures underlying workload, resource allocation, and visual search efficiency. Our results confirm previous findings and extend them toward the MR continuum, showing how, regardless of the perceptual load of the scene, AR environments posed increased attentional demands on users, as shown by diminished P3 amplitude and eye tracking features. This claim is corroborated by perceived workload, perceived distractibility, and difficulty in the visual search task. Identifying distinct physiological patterns across the reality–virtuality continuum opens the field to designing MR systems grounded in physiological computing. Our results inform the design of future adaptive interfaces that employ users' physiological signals as input for effective interaction in MR systems.

7 Open Science

We encourage readers to reproduce and extend our results. Our collected dataset, MR scenarios, 3D models, and analysis scripts are open-sourced and available at this link: https://osf.io/fncj4/.

Acknowledgments

Francesco Chiossi was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with Project ID 251654672 TRR 161. This work has been funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 952026 (https://www.humane-ai.eu/).

References

[1] Yomna Abdelrahman, Anam Ahmad Khan, Joshua Newn, Eduardo Velloso, Sherine Ashraf Safwat, James Bailey, Andreas Bulling, Frank Vetere, and Albrecht Schmidt. 2019. Classifying Attention Types with Thermal Imaging and Eye Tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 69 (sep 2019), 27 pages. https://doi.org/10.1145/3351227

[2] Magdalena Andrzejewska and Anna Stolińska. 2016. Comparing the difficulty of tasks using eye tracking combined with subjective and behavioural criteria. *Journal of Eye Movement Research* 9, 3 (2016), 16. https://doi.org/10.16910/jemr.9.3.3

[3] Janette Atkinson and Oliver J Braddick. 1989. 'Where'and 'what'in visual search. *Perception* 18, 2 (1989), 181–189. https://doi.org/10.1068/p180181

[4] Jonas Auda, Uwe Gruenefeld, Sarah Faltaous, Sven Mayer, and Stefan Schneegass. 2023. A Scoping Survey on Cross-Reality Systems. *ACM Comput. Surv.* 56, 4, Article 83 (oct 2023), 38 pages. https://doi.org/10.1145/3616536

[5] Ronald T Azuma. 1997. A survey of augmented reality. *Presence: teleoperators & virtual environments* 6, 4 (1997), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

[6] R Harald Baayen, Douglas J Davidson, and Douglas M Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language* 59, 4 (2008), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

[7] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823* (2014). https://doi.org/10.48550/arXiv.1406.5823

[8] Jackson Beatty. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin* 91, 2 (1982), 276. https://doi.org/10.1037/0033-2909.91.2.276

[9] Jaap A Beintema, Editha M van Loon, and Albert V van den Berg. 2005. Manipulating saccadic decision-rate distributions in visual search. *Journal of vision* 5, 3 (2005), 1–1. https://doi.org/10.1167/5.3.1

[10] Guillermo Bernal, Nelson Hidalgo, Conor Russomanno, and Pattie Maes. 2022. Galea: A physiological sensing system for behavioral research in virtual environments. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, New York, NY, USA, 66–76. https://doi.org/10.1109/VR51125.2022.00024

[11] Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A Robbins. 2015. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in neuroinformatics* 9 (2015), 16. https://doi.org/10.3389/fninf.2015.00016

[12] Oliver Bimber and Ramesh Raskar. 2006. Modern approaches to augmented reality. In *ACM SIGGRAPH 2006 Courses* (Boston, Massachusetts) *(SIGGRAPH '06)*. Association for Computing Machinery, New York, NY, USA, 1–es. https://doi.org/10.1145/1185657.1185796

[13] Frank Biocca, Charles Owen, Arthur Tang, and Corey Bohil. 2007. Attention issues in spatial information systems: Directing mobile users' visual attention using augmented reality. *Journal of Management Information Systems* 23, 4 (2007), 163–184. https://doi.org/10.2753/MIS0742-1222230408

[14] Frank Biocca, Arthur Tang, Charles Owen, and Fan Xiao. 2006. Attention Funnel: Omnidirectional 3D Cursor for Mobile Augmented Reality Platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 1115–1122. https://doi.org/10.1145/1124772.1124939

[15] Leonardo Bonanni, Chia-Hsun Lee, and Ted Selker. 2005. Attention-based design of augmented reality interfaces. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1228–1231. https://doi.org/10.1145/1056808.1056883

[16] Jürgen Bortz and Nicola Döring. 2007. *Forschungsmethoden und Evaluation für Human-und Sozialwissenschaftler: Limitierte Sonderausgabe*. Springer-Verlag. https://doi.org/10.1007/978-3-540-33306-7

[17] Riccardo Bovo, Daniele Giunchi, Ludwig Sidenmark, Joshua Newn, Hans Gellersen, Enrico Costanza, and Thomas Heinis. 2023. Speech-Augmented Cone-of-Vision for Exploratory Data Analysis. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 162, 18 pages. https://doi.org/10.1145/3544548.3581283

[18] Doug A Bowman, Sabine Coquillart, Bernd Froehlich, Michitaka Hirose, Yoshifumi Kitamura, Kiyoshi Kiyokawa, and Wolfgang Stuerzlinger. 2008. 3d user interfaces: New directions and perspectives. *IEEE computer graphics and applications* 28, 6 (2008), 20–36. https://doi.org/10.1109/MCG.2008.109

[19] Cynthia A Brewer. 1996. Prediction of simultaneous contrast between map colors with Hunt's model of color appearance. *Color Research & Application* 21, 3 (1996), 221–235. https://doi.org/10.1002/(SICI)1520-6378(199606)21:3<221::AID-COL4>3.0.CO;2-U

[20] Anne-Marie Brouwer, Maarten A Hogervorst, Bob Oudejans, Anthony J Ries, and Jonathan Touryan. 2017. EEG and eye tracking signatures of target encoding during structured visual search. *Frontiers in human neuroscience* 11 (2017), 264. https://doi.org/10.3389/fnhum.2017.00264

[21] Anne-Marie Brouwer, Boris Reuderink, Joris Vincent, Marcel AJ van Gerven, and Jan BF van Erp. 2013. Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *Journal of vision* 13, 3 (2013), 17–17. https://doi.org/10.1167/13.3.17
[22] Andreas Butz, Michael Schneider, and Mira Spassova. 2004. SearchLight – A Lightweight Search Function for Pervasive Environments. In *Pervasive Computing*, Alois Ferscha and Friedemann Mattern (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 351–356. https://doi.org/10.1007/978-3-540-24646-6_26
[23] Marisa Carrasco, Denise L Evert, Irene Chang, and Svetlana M Katz. 1995. The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & psychophysics* 57 (1995), 1241–1261.
[24] Benjamin T Carter and Steven G Luke. 2020. Best practices in eye tracking research. *International Journal of Psychophysiology* 155 (2020), 49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010
[25] Rebecca L. Charles and Jim Nixon. 2019. Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics* 74 (2019), 221–232. https://doi.org/10.1016/j.apergo.2018.08.028
[26] Siyuan Chen, Julien Epps, Natalie Ruiz, and Fang Chen. 2011. Eye Activity as a Measure of Human Mental Effort in HCI. In *Proceedings of the 16th International Conference on Intelligent User Interfaces* (Palo Alto, CA, USA) *(IUI '11)*. Association for Computing Machinery, New York, NY, USA, 315–318. https://doi.org/10.1145/1943403.1943454
[27] Taizhou Chen, Yi-Shiun Wu, and Kening Zhu. 2018. Investigating Different Modalities of Directional Cues for Multi-Task Visual-Searching Scenario in Virtual Reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 41, 5 pages. https://doi.org/10.1145/3281505.3281516
[28] Xiaojiao Chen, Xiaoteng Tang, Ying Zhao, Tengyu Huang, Ran Qian, Jiayi Zhang, Wei Chen, and Xiaosong Wang. 2023. Evaluating Visual Consistency of Icon Usage in Across-Devices. *International Journal of Human–Computer Interaction* (2023), 1–17. https://doi.org/10.1080/10447318.2022.2162275
[29] Xin Chen and Gregory J. Zelinsky. 2006. Real-world visual search is dominated by top-down guidance. *Vision Research* 46, 24 (2006), 4118–4133. https://doi.org/10.1016/j.visres.2006.08.008
[30] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. 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*. 282–297. https://doi.org/10.1145/3472749.3474750
[31] Yi Fei Cheng, Hang Yin, Yukang Yan, Jan Gugenheimer, and David Lindlbauer. 2022. Towards Understanding Diminished Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (, New Orleans, LA, USA,) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 549, 16 pages. https://doi.org/10.1145/3491102.3517452
[32] Francesco Chiossi, Luke Haliburton, Changkun Ou, Andreas Martin Butz, and Albrecht Schmidt. 2023. Short-Form Videos Degrade Our Capacity to Retain Intentions: Effect of Context Switching On Prospective Memory. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 30, 15 pages. https://doi.org/10.1145/3544548.3580778
[33] Francesco Chiossi, Changkun Ou, and Sven Mayer. 2023. Exploring Physiological Correlates of Visual Complexity Adaptation: Insights from EDA, ECG, and EEG Data for Adaptation Evaluation in VR Adaptive Systems. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 118, 7 pages. https://doi.org/10.1145/3544549.3585624
[34] Francesco Chiossi, Changkun Ou, and Sven Mayer. 2024. 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* (Arenzano, Genoa, Italy) *(AVI '24)*. Association for Computing Machinery, New York, NY, USA, Article 25, 9 pages. https://doi.org/10.1145/3656650.3656657
[35] Francesco Chiossi, Yagiz Turgut, Robin Welsch, and Sven Mayer. 2023. Adapting Visual Complexity Based on Electrodermal Activity Improves Working Memory Performance in Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 196 (sep 2023), 26 pages. https://doi.org/10.1145/3604243
[36] Francesco Chiossi, Robin Welsch, Steeven Villa, Lewis Chuang, and Sven Mayer. 2022. Virtual reality adaptation using electrodermal activity to support the user experience. *Big Data and Cognitive Computing* 6, 2 (2022), 55. https://doi.org/10.3390/bdcc6020055
[37] Emily A Cooper. 2023. The Perceptual Science of Augmented Reality. *Annual Review of Vision Science* 9 (2023). https://doi.org/10.1146/annurev-vision-111022-123758
[38] Brendan David-John, Candace Peacock, Ting Zhang, T Scott Murdison, Hrvoje Benko, and Tanya R Jonker. 2021. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*. 1–7.
[39] Daniele De Massari, Daniel Pacheco, Rahim Malekshahi, Alberto Betella, Paul FMJ Verschure, Niels Birbaumer, and Andrea Caria. 2014. Fast mental states decoding in mixed reality. *Frontiers in behavioral neuroscience* 8 (2014), 415.

https://doi.org/10.3389/fnbeh.2014.00415

[40] Rasha Deeb, Kristien Ooms, Alžběta Brychtová, Veerle Van Eetvelde, and Philippe De Maeyer. 2015. Background and foreground interaction: Influence of complementary colors on the search task. *Color Research & Application* 40, 5 (2015), 437–445. https://doi.org/10.1002/col.21920

[41] Jessyca L Derby and Barbara S Chaparro. 2022. Evaluating the Usability of Augmented Reality (AR) Applications: The Impact of The Device Used. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 66. SAGE Publications Sage CA: Los Angeles, CA, 1355–1359. https://doi.org/10.1177/1071181322661431

[42] Arindam Dey, Alex Chatburn, and Mark Billinghurst. 2019. Exploration of an EEG-based cognitively adaptive training system in virtual reality. In *2019 ieee conference on virtual reality and 3d user interfaces (vr)*. IEEE, New York, NY, USA, 220–226. https://doi.org/10.1109/VR.2019.8797840

[43] Benjamin J Dixon, Michael J Daly, Harley Chan, Allan D Vescan, Ian J Witterick, and Jonathan C Irish. 2013. Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical endoscopy* 27 (2013), 454–461. https://doi.org/10.1007/s00464-012-2457-3

[44] Jayfus T Doswell and Anna Skinner. 2014. Augmenting human cognition with adaptive augmented reality. In *Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making through Adaptive Systems: 8th International Conference, AC 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Proceedings 8*. Springer, 104–113. https://doi.org/10.1007/978-3-319-07527-3_10

[45] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. The Index of Pupillary Activity: Measuring Cognitive Load Vis-à-Vis Task Difficulty with Pupil Oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173856

[46] John Duncan and Glyn W Humphreys. 1989. Visual search and stimulus similarity. *Psychological review* 96, 3 (1989), 433. https://doi.org/10.1037/0033-295X.96.3.433

[47] Christopher M. Durugbo. 2021. Eye tracking for work-related visual search: a cognitive task analysis. *Ergonomics* 64, 2 (2021), 225–240. https://doi.org/10.1080/00140139.2020.1822547

[48] Miguel P Eckstein. 2011. Visual search: A retrospective. *Journal of vision* 11, 5 (2011), 14–14. https://doi.org/10.1167/11.5.14

[49] Miguel P. Eckstein. 2011. Visual search: A retrospective. *Journal of Vision* 11, 5 (12 2011), 14–14. https://doi.org/10.1167/11.5.14

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

[51] Ralf Engbert and Reinhold Kliegl. 2003. Microsaccades uncover the orientation of covert attention. *Vision research* 43, 9 (2003), 1035–1045. https://doi.org/10.1016/S0042-6989(03)00084-1

[52] Robin Eyraud, Elisabetta Zibetti, and Thierry Baccino. 2015. Allocation of visual attention while driving with simulated augmented reality. *Transportation research part F: traffic psychology and behaviour* 32 (2015), 46–55. https://doi.org/10.1016/j.trf.2015.04.011

[53] Martin Feick, Niko Kleer, Anthony Tang, and Antonio Krüger. 2020. The Virtual Reality Questionnaire Toolkit *(UIST '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3379350.3416188

[54] Patrique Fiedler, Carlos Fonseca, Eko Supriyanto, Frank Zanow, and Jens Haueisen. 2022. A high-density 256-channel cap for dry electroencephalography. *Human brain mapping* 43, 4 (2022), 1295–1308. https://doi.org/10.1002/hbm.25721

[55] Andrea Finke, Kai Essig, Giuseppe Marchioro, and Helge Ritter. 2016. Toward FRP-based brain-machine interfaces—single-trial classification of fixation-related potentials. *PloS one* 11, 1 (2016), e0146848. https://doi.org/10.1371/journal.pone.0146848

[56] Norman Forschack, Christopher Gundlach, Steven Hillyard, and Matthias M Müller. 2022. Dynamics of attentional allocation to targets and distractors during visual search. *NeuroImage* 264 (2022), 119759. https://doi.org/10.1016/j.neuroimage.2022.119759

[57] Jérémy Frey, Maxime Daniel, Julien Castet, Martin Hachet, and Fabien Lotte. 2016. Framework for electroencephalography-based evaluation of user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2283–2294. https://doi.org/10.1145/2858036.2858525

[58] Hind Gacem, Gilles Bailly, James Eagan, and Eric Lecolinet. 2015. Finding Objects Faster in Dense Environments Using a Projection Augmented Robotic Arm. In *Human-Computer Interaction – INTERACT 2015*, Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler (Eds.). Springer International Publishing, Cham, 221–238. https://doi.org/10.1007/978-3-319-22698-9_15

[59] Elena Gherri and Martin Eimer. 2011. Active listening impairs visual perception and selectivity: An ERP study of auditory dual-task costs on visual attention. *Journal of cognitive neuroscience* 23, 4 (2011), 832–844. https://doi.org/10.1162/jocn.2010.21468
[60] Joseph H. Goldberg, Mark J. Stimson, Marion Lewenstein, Neil Scott, and Anna M. Wichansky. 2002. Eye Tracking in Web Search Tasks: Design Implications. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications* (New Orleans, Louisiana) *(ETRA '02)*. Association for Computing Machinery, New York, NY, USA, 51–58. https://doi.org/10.1145/507072.507082
[61] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, et al. 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in neuroscience* (2013), 267. https://doi.org/10.3389/fnins.2013.00267
[62] Iris IA Groen, Sennay Ghebreab, Hielke Prins, Victor AF Lamme, and H Steven Scholte. 2013. From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. *Journal of Neuroscience* 33, 48 (2013), 18814–18824.
[63] Uwe Gruenefeld, Jonas Auda, Florian Mathis, Stefan Schneegass, Mohamed Khamis, Jan Gugenheimer, and Sven Mayer. 2022. VRception: Rapid Prototyping of Cross-Reality Systems in Virtual Reality. In *CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3491102.3501821
[64] Uwe Gruenefeld, Tim Claudius Stratmann, Abdallah El Ali, Susanne Boll, and Wilko Heuten. 2018. RadialLight: Exploring Radial Peripheral LEDs for Directional Cues in Head-Mounted Displays. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 39, 6 pages. https://doi.org/10.1145/3229434.3229437
[65] Jacob Hadnett-Hunter, Eamonn O'Neill, and Michael J Proulx. 2022. Contributed session ii: Visual search in virtual reality (vsvr): A visual search toolbox for virtual reality. *Journal of Vision* 22, 3 (2022), 19–19. https://doi.org/10.1167/jov.22.3.19
[66] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908. https://doi.org/10.1177/154193120605000909
[67] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9
[68] Thomas Haslwanter. 1995. Mathematics of three-dimensional eye rotations. *Vision Research* 35, 12 (1995), 1727–1739. https://doi.org/10.1016/0042-6989(94)00257-M
[69] Larry V Hedges. 2007. Effect sizes in cluster-randomized designs. (2007). https://doi.org/10.3102/1076998606298
[70] Eckhard H Hess and James M Polt. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 3611 (1964), 1190–1192. https://doi.org/10.1126/science.143.3611.119
[71] Anne P Hillstrom. 2000. Repetition effects in visual search. *Perception & psychophysics* 62 (2000), 800–817. https://doi.org/10.3758/BF03206924
[72] Mauricio Hincapié, Andrea Caponio, Horacio Rios, and Eduardo González Mendívil. 2011. An introduction to Augmented Reality with applications in aeronautical maintenance. In *2011 13th international conference on transparent optical networks*. IEEE, New York, NY, USA, 1–4. https://doi.org/10.1109/ICTON.2011.5970856
[73] Daichi Hirobe, Yuki Uranishi, Jason Orlosky, Shizuka Shirai, Photchara Ratsamee, and Haruo Takemura. 2022. Objective Measurements of Background Color Shifts Caused by Optical See-Through Head-Mounted Displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, New York, NY, USA, 389–390. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00084
[74] Andrew Hollingworth and Brett Bahle. 2020. *Eye Tracking in Visual Search Experiments*. Springer US, New York, NY, 23–35. https://doi.org/10.1007/7657_2019_30
[75] Sebastian Hubenschmid, Jonathan Wieland, Daniel Immanuel Fink, Andrea Batch, Johannes Zagermann, Niklas Elmqvist, and Harald Reiterer. 2022. ReLive: Bridging In-Situ and Ex-Situ Visual Analytics for Analyzing Mixed Reality User Studies. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 24, 20 pages. https://doi.org/10.1145/3491102.3517550
[76] Johan Hulleman. 2020. Quantitative and qualitative differences in the top-down guiding attributes of visual search. *Journal of Experimental Psychology: Human Perception and Performance* 46, 9 (2020), 942. https://doi.org/10.1037/xhp0000764
[77] Muhammad Hussain and Jaehyun Park. 2023. Effect of Transparency Levels and Real-World Backgrounds on the User Interface in Augmented Reality Environments. *International Journal of Human–Computer Interaction* (2023), 1–10. https://doi.org/10.1080/10447318.2023.2212218
[78] Mainak Jas, Denis A Engemann, Yousra Bekhti, Federico Raimondo, and Alexandre Gramfort. 2017. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage* 159 (2017), 417–429. https://doi.org/j.neuroimage.

2017.06.030

[79] Gouri Jha, Lavanya shm Sharma, and Shailja Gupta. 2021. Future of augmented reality in healthcare department. In *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020*. Springer, 667–678. https://doi.org/10.1007/978-981-16-0733-2_47

[80] Brendan John, Pallavi Raiturkar, Arunava Banerjee, and Eakta Jain. 2018. An Evaluation of Pupillary Light Response Models for 2D Screens and VR HMDs. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 19, 11 pages. https://doi.org/10.1145/3281505.3281538

[81] Jeffrey S Johnson and Bruno A Olshausen. 2003. Timecourse of neural signatures of object recognition. *Journal of vision* 3, 7 (2003), 4–4. https://doi.org/10.1167/3.7.4

[82] David W Jr Hosmer, Stanley Lemeshow, and RX Sturdivant. 2004. Applied logistic regression. *Hoboken: John Wiley & Sons* (2004). https://doi.org/10.1002/9781118548387

[83] Marcel Adam Just and Patricia A Carpenter. 1976. Eye fixations and cognitive processes. *Cognitive Psychology* 8, 4 (1976), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

[84] Daniel Kahneman and Jackson Beatty. 1966. Pupil Diameter and Load on Memory. *Science* 154, 3756 (1966), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

[85] Juan E. Kamienkowski, Matias J. Ison, Rodrigo Quian Quiroga, and Mariano Sigman. 2012. Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of Vision* 12, 7 (07 2012), 4–4. https://doi.org/10.1167/12.7.4

[86] Juan E Kamienkowski, Matias J Ison, Rodrigo Quian Quiroga, and Mariano Sigman. 2012. Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of vision* 12, 7 (2012), 4–4. https://doi.org/10.1167/12.7.4

[87] Lisandro N Kaunitz, Juan E Kamienkowski, Alexander Varatharajah, Mariano Sigman, Rodrigo Quian Quiroga, and Matias J Ison. 2014. Looking for a face in the crowd: Fixation-related potentials in an eye-movement visual search task. *NeuroImage* 89 (2014), 297–305. https://doi.org/10.1167/12.7.4

[88] Peter Kiefer, Ioannis Giannopoulos, Andrew Duchowski, and Martin Raubal. 2016. Measuring Cognitive Load for Map Tasks Through Pupil Diameter. In *Geographic Information Science*, Jennifer A. Miller, David O'Sullivan, and Nancy Wiegand (Eds.). Springer International Publishing, Cham, 323–337. https://doi.org/10.1007/978-3-319-45738-3_21

[89] You-Jin Kim, Radha Kumaran, Ehsan Sayyad, Anne Milner, Tom Bullock, Barry Giesbrecht, and Tobias Höllerer. 2022. Investigating Search Among Physical and Virtual Objects Under Different Lighting Conditions. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3788–3798. https://doi.org/10.1109/TVCG.2022.3203093

[90] Albert Kok. 2001. On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 3 (2001), 557–577. https://doi.org/10.1017/S0048577201990559

[91] Thomas Kosch, Mariam Hassib, Daniel Buschek, and Albrecht Schmidt. 2018. Look into My Eyes: Using Pupil Dilation to Estimate Mental Workload for Task Complexity Adaptation. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188643

[92] Daniel G. Krakowczyk, David R. Reich, Jakob Chwastek, Deborah N. Jakobi, Paul Prasse, Assunta Süss, Oleksii Turuta, Paweł Kasprowski, and Lena A. Jäger. 2023. Pymovements: A Python Package for Eye Movement Data Processing. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications* (Tubingen, Germany) *(ETRA '23)*. Association for Computing Machinery, New York, NY, USA, Article 53, 2 pages. https://doi.org/10.1145/3588015.3590134

[93] Arni Kristjánsson. 2006. Simultaneous priming along multiple feature dimensions in a visual search task. *Vision research* 46, 16 (2006), 2554–2570. https://doi.org/10.1016/j.visres.2006.01.015

[94] Daniel Lange, Tim Claudius Stratmann, Uwe Gruenefeld, and Susanne Boll. 2020. HiveFive: Immersion Preserving Attention Guidance in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376803

[95] Rebecca K Lawrence and Jay Pratt. 2022. Salience matters: Distractors may, or may not, speed target-absent searches. *Attention, Perception, & Psychophysics* (2022), 1–12.

[96] Te-Won Lee, Mark Girolami, and Terrence J Sejnowski. 1999. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation* 11, 2 (1999), 417–441.

[97] Adam Li, Jacob Feitelberg, Anand Prakash Saini, Richard Höchenberger, and Mathieu Scheltienne. 2022. MNE-ICALabel: Automatically annotating ICA components with ICLabel in Python. *Journal of Open Source Software* 7, 76 (2022), 4484. https://doi.org/10.21105/joss.04484

[98] Petra Lietz. 2010. Research into questionnaire design: A summary of the literature. *International journal of market research* 52, 2 (2010), 249–272. https://doi.org/10.2501/S147078530920120

[99] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*. 147–160. https://doi.org/10.1145/3332165.3347945

[100] Steson Lo and Sally Andrews. 2015. To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in psychology* 6 (2015), 1171. https://doi.org/10.3389/fpsyg.2015.01171

[101] Xingyu Long, Sven Mayer, and Francesco Chiossi. 2024. Multimodal Detection of External and Internal Attention in Virtual Reality using EEG and Eye Tracking Features. In *Proceedings of Mensch und Computer 2024 (MuC '24)*. ACM, 15. https://doi.org/10.1145/3670653.3670657

[102] Romy Lorenz, Javier Pascual, Benjamin Blankertz, and Carmen Vidaurre. 2014. Towards a holistic assessment of the user experience with hybrid BCIs. *Journal of neural engineering* 11, 3 (2014), 035007. https://doi.org/10.1088/1741-2560/11/3/035007

[103] Steven J Luck and Steven A Hillyard. 1994. Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 3 (1994), 291–308.

[104] Steven J Luck and Steven A Hillyard. 1994. Spatial filtering during visual search: evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance* 20, 5 (1994), 1000.

[105] Bhanuka Mahanama, Yasith Jayawardana, Sundararaman Rengarajan, Gavindya Jayawardena, Leanne Chukoskie, Joseph Snider, and Sampath Jayarathna. 2022. Eye Movement and Pupil Measures: A Review. *Frontiers in Computer Science* 3 (2022). https://doi.org/10.3389/fcomp.2021.733531

[106] Vera Maljkovic and Ken Nakayama. 1994. Priming of pop-out: I. Role of features. *Memory & cognition* 22, 6 (1994), 657–672. https://doi.org/10.3758/BF03209251

[107] Francesco Marini, Katherine A Breeding, and Jacqueline C Snow. 2019. Distinct visuo-motor brain dynamics for real-world objects versus planar images. *Neuroimage* 195 (2019), 232–242. https://doi.org/10.1016/j.neuroimage.2019.02.026

[108] Nuno Cid Martins, Bernardo Marques, João Alves, Tiago Araújo, Paulo Dias, and Beatriz Sousa Santos. 2022. Augmented reality situated visualization in decision-making. *Multimedia Tools and Applications* 81, 11 (2022), 14749–14772.

[109] Nathan Matsuda, Alex Chapiro, Yang Zhao, Clinton Smith, Romain Bachy, and Douglas Lanman. 2022. Realistic luminance in VR. In *SIGGRAPH Asia 2022 Conference Papers*. 1–8. https://doi.org/10.1145/3550469.3555427

[110] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A dose of reality: Overcoming usability challenges in vr head-mounted displays. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2143–2152. https://doi.org/10.1145/2702123.2702382

[111] Tilo Mentler and Daniel Wessel. 2019. A Toolkit for Studying Attention and Reaction Times to Smartglass Messages in Conditions of Different Perceptual Load. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume V: Human Simulation and Virtual Environments, Work With Computing Systems (WWCS), Process Control 20*. Springer, 595–604. https://doi.org/10.1007/978-3-319-96077-7_64

[112] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329. https://doi.org/10.1.1.102.4646

[113] Nadine M. Moacdieh and Nadine Sarter. 2017. Using Eye Tracking to Detect the Effects of Clutter on Visual Search in Real Time. *IEEE Transactions on Human-Machine Systems* 47, 6 (2017), 896–902. https://doi.org/10.1109/THMS.2017.2706666

[114] Mehrdad Naderi, Tatjana Pladere, Reinis Alksnis, and Gunta Krumina. 2023. Brain activity underlying visual search in depth when viewing volumetric multiplanar images. *Scientific Reports* 13, 1 (2023), 7672. https://doi.org/10.1038/s41598-023-34758-9

[115] Chukwuemeka Nwagu, Alaa AlSlaity, and Rita Orji. 2023. EEG-Based Brain-Computer Interactions in Immersive Virtual and Augmented Reality: A Systematic Review. *Proc. ACM Hum.-Comput. Interact.* 7, EICS, Article 174 (jun 2023), 33 pages. https://doi.org/10.1145/3593226

[116] Manuel Oliva. 2019. Pupil size and search performance in low and high perceptual load. *Cognitive, Affective, & Behavioral Neuroscience* 19, 2 (2019), 366–376. https://doi.org/10.3758/s13415-018-00677-w

[117] Bettina Olk, Alina Dinu, David J Zielinski, and Regis Kopper. 2018. Measuring visual search and distraction in immersive virtual reality. *Royal Society open science* 5, 5 (2018), 172331. https://doi.org/10.1098/rsos.172331

[118] Oskar Palinko, Andrew L. Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 141–144. https://doi.org/10.1145/1743666.1743701

[119] Evan M Palmer, Todd S Horowitz, Antonio Torralba, and Jeremy M Wolfe. 2011. What are the shapes of response time distributions in visual search? *Journal of experimental psychology: human perception and performance* 37, 1 (2011), 58. https://doi.org/10.1037/a0020747

[120] Liping Pang, Yurong Fan, Ye Deng, Xin Wang, and Tianbo Wang. 2020. Mental Workload Classification By Eye Movements In Visual Search Tasks. In *2020 13th International Congress on Image and Signal Processing, BioMedical*

Engineering and Informatics (CISP-BMEI). 29–33. https://doi.org/10.1109/CISP-BMEI51763.2020.9263668

[121] Orestis Papaioannou and Steven J Luck. 2020. Effects of eccentricity on the attention-related N2pc component of the event-related potential waveform. *Psychophysiology* 57, 5 (2020), e13532. https://doi.org/10.1111/psyp.13532

[122] Youngmin Park, Vincent Lepetit, and Woontack Woo. 2008. Multiple 3d object tracking for augmented reality. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, 117–120. https://doi.org/10.1109/ISMAR.2008.4637336

[123] François Perrin, Jacques Pernier, Olivier Bertrand, and Jean Francois Echallier. 1989. Spherical splines for scalp potential and current density mapping. *Electroencephalography and clinical neurophysiology* 72, 2 (1989), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

[124] Ken Pfeuffer, Yasmeen Abdrabou, Augusto Esteves, Radiah Rivu, Yomna Abdelrahman, Stefanie Meitner, Amr Saadi, and Florian Alt. 2021. ARtention: A design space for gaze-adaptive user interfaces in augmented reality. *Computers & Graphics* 95 (2021), 1–12. https://doi.org/10.1016/j.cag.2021.01.001

[125] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. 2019. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage* 198 (2019), 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

[126] John Polich. 2007. Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology* 118, 10 (2007), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

[127] Marc Pomplun, Tyler W. Garaas, and Marisa Carrasco. 2013. The effects of task difficulty on visual search strategy in virtual 3D displays. *Journal of Vision* 13, 3 (08 2013), 24–24. https://doi.org/10.1167/13.3.24

[128] Michael I Posner. 1980. Orienting of attention. *Quarterly journal of experimental psychology* 32, 1 (1980), 3–25. https://doi.org/10.1080/00335558008248231

[129] Michael I. Posner and Steven E. Petersen. 1990. The Attention System of the Human Brain. *Annual Review of Neuroscience* 13, 1 (1990), 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

[130] Geoffrey F Potts, Mario Liotti, Don M Tucker, and Michael I Posner. 1996. Frontal and inferior temporal cortical activity in visual target detection: Evidence from high spatially sampled event-related potentials. *Brain topography* 9 (1996), 3–14. https://doi.org/10.1007/BF01191637

[131] Nikki Pratt, Adrian Willoughby, and Diane Swick. 2011. Effects of working memory load on visual selective attention: behavioral and electrophysiological evidence. *Frontiers in human neuroscience* 5 (2011), 57. https://doi.org/10.3389/fnhum.2011.00057

[132] Felix Putze, Susanne Putze, Merle Sagehorn, Christopher Micek, and Erin T. Solovey. 2022. Understanding HCI Practices and Challenges of Experiment Reporting with Brain Signals: Towards Reproducibility and Reuse. *ACM Trans. Comput.-Hum. Interact.* 29, 4, Article 31 (mar 2022), 43 pages. https://doi.org/10.1145/3490554

[133] Felix Putze, Athanasios Vourvopoulos, Anatole Lécuyer, Dean Krusienski, Sergi Bermúdez i Badia, Timothy Mullen, and Christian Herff. 2020. Brain-computer interfaces and augmented/virtual reality. *Frontiers in human neuroscience* 14 (2020), 144. https://doi.org/10.3389/fnhum.2020.00144

[134] Karim Rajaei, Yalda Mohsenzadeh, Reza Ebrahimpour, and Seyed-Mahdi Khaligh-Razavi. 2019. Beyond core object recognition: Recurrent processes account for object recognition under occlusion. *PLoS computational biology* 15, 5 (2019), e1007001.

[135] Natasha Anne Rappa, Susan Ledger, Timothy Teo, Kok Wai Wong, Brad Power, and Bruce Hilliard. 2022. 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), 1338–1350. https://doi.org/10.1080/10494820.2019.1702560

[136] Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3 (1998), 372. https://doi.org/10.1037/0033-2909.124.3.372

[137] Geraint Rees, Chris Frith, and Nilli Lavie. 2001. Processing of irrelevant visual motion during performance of an auditory attention task. *Neuropsychologia* 39, 9 (2001), 937–949. https://doi.org/10.1016/S0028-3932(01)00016-1

[138] Anthony J Ries, Jon Touryan, Barry Ahrens, and Patrick Connolly. 2016. The impact of task demands on fixation-related brain potentials during guided search. *PloS one* 11, 6 (2016), e0157260. https://doi.org/10.1371/journal.pone.0157260

[139] Joan Sol Roo and Martin Hachet. 2017. One Reality: Augmenting How the Physical World is Experienced by Combining Multiple Mixed Reality Modalities. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 787–795. https://doi.org/10.1145/3126594.3126638

[140] Inês Salselas, Rui Penha, and Gilberto Bernardes. 2021. Sound design inducing attention in the context of audiovisual immersive environments. *Personal and Ubiquitous Computing* 25 (2021), 737–748. https://doi.org/10.1007/s00779-020-01386-3

[141] Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. 71–78. https://doi.org/10.1145/355017.355028

[142] Noor Seijdel, Jessica Loke, Ron Van de Klundert, Matthew Van der Meer, Eva Quispel, Simon Van Gaal, Edward HF De Haan, and H Steven Scholte. 2021. On the necessity of recurrent processing during object recognition: it depends on the need for scene segmentation. *Journal of Neuroscience* 41, 29 (2021), 6281–6289.

[143] Hitesh Nidhi Sharma, Phoebe O. Toups Dugas, Ajit Jain, and Andruid Kerne. 2015. Designing to Split Attention in a Mixed Reality Game. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (London, United Kingdom) *(CHI PLAY '15)*. Association for Computing Machinery, New York, NY, USA, 691–696. https://doi.org/10.1145/2793107.2810289

[144] Mansi Sharma, Maurice Rekrut, Jan Alexandersson, and Antonio Krüger. 2022. Towards Improving EEG-Based Intent Recognition in Visual Search Tasks. In *International Conference on Neural Information Processing*. Springer, 604–615. https://doi.org/10.1007/978-3-031-30111-7_51

[145] Tomonari Shimomura and Takatsune Kumada. 2011. Spatial working memory load affects counting but not subitizing in enumeration. *Attention, Perception, & Psychophysics* 73 (2011), 1694–1709. https://doi.org/10.3758/s13414-011-0135-5

[146] Motofumi Shishikura, Hiroshi Tamura, and Ko Sakai. 2023. Correlation between neural responses and human perception in figure-ground segregation. *Frontiers in systems neuroscience* 16 (2023), 999575. https://doi.org/10.3389/fnsys.2022.999575

[147] Ludwig Sidenmark, Franziska Prummer, Joshua Newn, and Hans Gellersen. 2023. Comparing Gaze, Head and Controller Selection of Dynamically Revealed Targets in Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4740–4750. https://doi.org/10.1109/TVCG.2023.3320235

[148] Richard Skarbez, Missie Smith, and Mary C Whitton. 2021. Revisiting Milgram and Kishino's reality-virtuality continuum. *Frontiers in Virtual Reality* 2 (2021), 647997. https://doi.org/10.3389/frvir.2021.647997

[149] Alexis D. Souchet, Stéphanie Philippe, Domitile Lourdeaux, and Laure Leroy. 2022. Measuring Visual Fatigue and Cognitive Load via Eye Tracking while Learning with Virtual Reality Head-Mounted Displays: A Review. *International Journal of Human–Computer Interaction* 38, 9 (2022), 801–824. https://doi.org/10.1080/10447318.2021.1976509

[150] Rhaíra Helena Caetano e Souza and Eduardo Lázaro Martins Naves. 2021. Attention detection in virtual environments using EEG signals: a scoping review. *frontiers in physiology* 12 (2021), 727840. https://doi.org/10.3389/fphys.2021.727840

[151] Maximilian Speicher, Brian D. Hall, and Michael Nebeling. 2019. What is Mixed Reality?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300767

[152] Gabriela Tavares, Pietro Perona, and Antonio Rangel. 2017. The attentional drift diffusion model of simple perceptual decision-making. *Frontiers in neuroscience* 11 (2017), 468. https://doi.org/10.3389/fnins.2017.00468

[153] Simon Thorpe, Denis Fize, and Catherine Marlot. 1996. Speed of processing in the human visual system. *nature* 381, 6582 (1996), 520–522. https://doi.org/10.1038/381520a0

[154] Shurong Tong and Yafei Nie. 2023. Measuring Designers' Cognitive Load for Timely Knowledge Push via Eye Tracking. *International Journal of Human–Computer Interaction* 39, 6 (2023), 1230–1243. https://doi.org/10.1080/10447318.2022.2057898

[155] Anne M. Treisman and Garry Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12, 1 (1980), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

[156] Lana M Trick and Zenon W Pylyshyn. 1994. Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological review* 101, 1 (1994), 80. https://doi.org/10.1037/0033-295X.101.1.80

[157] Marija Ušćumlić and Benjamin Blankertz. 2016. Active visual search in non-stationary scenes: coping with temporal variability and uncertainty. *Journal of Neural Engineering* 13, 1 (2016), 016015. https://doi.org/10.1088/1741-2560/13/1/016015

[158] AV Van den Berg and EM Van Loon. 2005. An invariant for timing of saccades during visual search. *Vision Research* 45, 12 (2005), 1543–1555. https://doi.org/10.1016/j.visres.2004.12.018

[159] Alessia Vozzi, Vincenzo Ronca, Pietro Aricò, Gianluca Borghini, Nicolina Sciaraffa, Patrizia Cherubino, Arianna Trettel, Fabio Babiloni, and Gianluca Di Flumeri. 2021. The sample size matters: to what extent the participant reduction affects the outcomes of a neuroscientific research. A case-study in neuromarketing field. *Sensors* 21, 18 (2021), 6088. https://doi.org/10.3390/s21186088

[160] Bing-Shun Wang, Xiao-Jin Wang, and Li-Kun Gong. 2009. The construction of a Williams design and randomization in cross-over clinical trials using SAS. *Journal of statistical software* 29 (2009), 1–10. https://doi.org/10.18637/jss.v029.c01

[161] Nanjia Wang and Frank Maurer. 2022. A Design Space for Single-User Cross-Reality Applications. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces* (Frascati, Rome, Italy) *(AVI 2022)*. Association for Computing Machinery, New York, NY, USA, Article 31, 5 pages. https://doi.org/10.1145/3531073.3531116

[162] Qiuzhen Wang, Sa Yang, Manlu Liu, Zike Cao, and Qingguo Ma. 2014. An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems* 62 (2014), 1–10. https://doi.org/10.1016/j.dss.2014.02.007

[163] Yijun Wang, Xiaorong Gao, Bo Hong, Chuan Jia, and Shangkai Gao. 2008. Brain-computer interfaces based on visual evoked potentials. *IEEE Engineering in medicine and biology magazine* 27, 5 (2008), 64–71.

[164] Amelia C. Warden, Christopher D. Wickens, Domenick Mifsud, Shannon Ourada, Benjamin A. Clegg, and Francisco R. Ortega. 2022. Visual Search in Augmented Reality: Effect of Target Cue Type and Location. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 66, 1 (2022), 373–377. https://doi.org/10.1177/1071181322661260

[165] Amelia C. Warden, Christopher D. Wickens, Daniel Rehberg, Francisco R. Ortega, and Benjamin A. Clegg. 2023. Fast, Accurate, but Sometimes Too-Compelling Support: The Impact of Imperfectly Automated Cues in an Augmented Reality Head-Mounted Display on Visual Search Performance. arXiv:2303.14300

[166] Markus A Wenzel, Jan-Eike Golenia, and Benjamin Blankertz. 2016. Classification of eye fixation related potentials for variable stimulus saliency. *Frontiers in neuroscience* 10 (2016), 23. https://doi.org/10.3389/fnins.2016.00023

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

[168] Jeremy M Wolfe. 2020. Visual search: How do we find what we are looking for? *Annual review of vision science* 6 (2020), 539–562. https://doi.org/10.1146/annurev-vision-091718-015048

[169] Jeremy M Wolfe, Kyle R Cave, and Susan L Franzel. 1989. Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance* 15, 3 (1989), 419. https://doi.org/10.1037/0096-1523.15.3.419

[170] Jeremy M Wolfe and Todd S Horowitz. 2017. Five factors that guide attention in visual search. *Nature Human Behaviour* 1, 3 (2017), 0058. https://doi.org/10.1038/s41562-017-0058

[171] Jeremy M Wolfe and Michael J Van Wert. 2010. Varying target prevalence reveals two dissociable decision criteria in visual search. *Current biology* 20, 2 (2010), 121–124. https://doi.org/10.1016/j.cub.2009.11.066

[172] Geoffrey F Woodman and Steven J Luck. 1999. Electrophysiological measurement of rapid shifts of attention during visual search. *Nature* 400, 6747 (1999), 867–869. https://doi.org/10.1038/23698

[173] Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2018. Studying Eye Movements as a Basis for Measuring Cognitive Load. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188628

[174] Lili Zhang and Michael J Murdoch. 2018. Color matching criteria in augmented reality. *Journal of Perceptual Imaging* 1, 1 (2018), 10506–1. https://doi.org/10.2352/J.Percept.lmaging.2018.1.1.010506

[175] Ling Zhu and Jiufang Lv. 2023. Review of Studies on User Research Based on EEG and Eye Tracking. *Applied Sciences* 13, 11 (2023), 6502. https://doi.org/10.3390/app13116502

A Appendix

A.1 Order Effect Results

In examining potential fatigue or learning effects, we tested changes in RTs, accuracy, and NASA-TLX scores over the course of the experiment, which included a substantial 300 trials across six experimental conditions. Given the prolonged engagement required by our experimental design, we verified the potential effects of fatigue and possible learning effects over time. Before analysis, we tested the normality of RT, accuracy, and NASA-TLX scores for each block using the Shapiro-Wilk test to conduct repeated measures ANOVA (RM ANOVA) or Friedman test, for non-parametric data. Changes in subjective workload were evaluated using NASA-TLX scores collected at the end of each block of trials. Depending on their distribution, either RM ANOVA or Friedman tests were used to detect significant trends in these scores.

A.1.1 Reaction Times. First, the normality of the reaction time data was assessed using the Shapiro-Wilk test. The results indicated that the data were not normally distributed, $W = .954, p < .001$. Thus, a Friedman test was conducted to evaluate order effects on differences on reaction times. The test did not reveal significant differences in reaction times across the orders, $\chi^2(5) = 2.31, p = .805$. This result suggests that there were no statistically significant changes in reaction times across the different conditions of the experiment, indicating an absence of both fatigue and learning effects over the duration of the trials.

A.1.2 Accuracy. The Shapiro-Wilk test was conducted to assess the normality of the accuracy data. The results indicated that the accuracy data were not normally distributed, $W = .876, p < .001$. We then conducted a Friedman test to investigate differences in accuracy due to order effects. The test revealed no significant differences in accuracy across the orders, $\chi^2(5) = 2.78, p = .733$. These findings suggest that there were no statistically significant changes in accuracy that would indicate systematic effects such as learning or fatigue over the course of the experiment.

A.1.3 NASA-TLX scores. The normality of NASA-TLX scores were not normally distributed, $W = .972, p = .011$. Given the non-normal distribution of the NASA TLX scores, a Friedman test was conducted to investigate differences in perceived workload across the different orders of the experiment. The Friedman test did not reveal significant differences in NASA TLX scores across the orders, $\chi^2(5) = 8.78, p = .118$. This result suggests no statistically significant changes in perceived workload over time, indicating that the order of conditions did not differentially affect participants' perceived workload.

Received February 2024; revised May 2024; accepted June 2024