

Reaction Time as a Proxy for Presence in Mixed Reality with Distraction

Yasra Chandio*
University of Massachusetts Amherst

Victoria Interrante†
University of Minnesota Twin Cities

Fatima M. Anwar ‡
University of Massachusetts Amherst

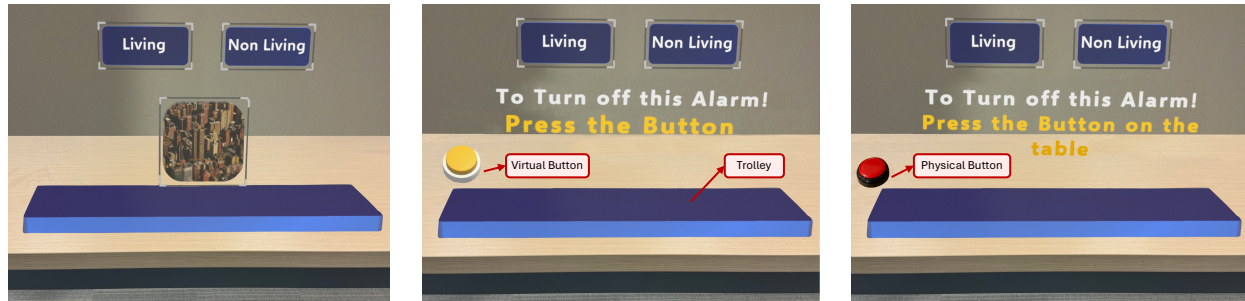


Fig. 1: View of the virtual scene across three MR conditions: In No Distraction (ND), participants sort images as living or non-living without distractions. In Congruent Distraction (CD), they sort images with virtual distractions. In Incongruent Distraction (ID), they sort images with real-world distractions. Red boxes and arrows label scene elements and are not part of the experiment.

Abstract—Distractions in mixed reality (MR) environments can significantly influence user experience, affecting key factors such as presence, reaction time, cognitive load, and Break in Presence (BIP). Presence measures immersion, reaction time captures user responsiveness, cognitive load reflects mental effort, and BIP represents moments when attention shifts from the virtual to the real world, breaking immersion. While prior work has established that distractions impact these factors individually, the relationship between these constructs remains underexplored, particularly in MR environments where users engage with both real and virtual stimuli. To address this gap, we have presented a theoretical model to understand how congruent and incongruent distractions affect all these constructs. We conducted a within-subject study ($N = 54$) where participants performed image-sorting tasks under different distraction conditions. Our findings show that incongruent distractions significantly increase cognitive load, slow reaction times, and elevate BIP frequency, with presence mediating these effects.

Index Terms—Mixed Reality, Presence, Distraction

1 INTRODUCTION

In Mixed Reality (MR) environments, *presence* refers to the immersive experience of being fully situated (“*being there*”) in a virtual environment (VE) [14, 51]. A high sense of presence is a critical factor for determining user immersion [22], engagement [55], task performance [27], and interaction quality [6]. Presence is largely constructed through cognitive and perceptual processes [56, 74]. Cognition shapes how users interpret and engage with the VE, while perception involves the sensory inputs that maintain the coherence of that experience. However, disruptions in either the real-world (noises, touch, etc.) or VE (system lags, glitches etc.) can affect these processes, often causing inconsistencies between the user’s mental model and the VE [35], leading to the *break in presence (BIP)* [59, 62], causing the user’s attention to shift from the VE back to real-world. Real-world elements inherently differ from virtual stimuli in their unpredictability and sensory modality, requiring users to reallocate cognitive attention across domains [35]. This perceptual discontinuity may increase the likelihood of BIP and reduce presence.

Traditionally, subjective methods like post-experience questionnaires

*e-mail: ychandio@umass.edu

†e-mail: interran@umn.edu

‡e-mail: fanwar@umass.edu

have been used to measure presence [49]. These methods, however, have significant limitations. They fail to capture the real-time, moment-to-moment fluctuations in presence [50] and are subject to biases in how participants recall or interpret their experiences [13]. This makes it difficult to objectively measure presence as it dynamically changes throughout the MR experience. Prior studies have examined the correlation between presence and reaction time by manipulating perceptual factors such as object realism and plausibility [10] and the link between the user’s cognitive state and performance to assess presence through reaction time [11].

While the correlation between presence and reaction time has been established, less is known about how this relationship is affected by distractions, particularly in MR environments where users remain connected to the physical world. These environments introduce additional complexity because real or virtual distractions can disrupt a user’s attention [20, 29] and potentially lead to a BIP [9]. In MR, distractions are not just an inevitable part of the experience; they also provide a key challenge for maintaining consistent presence [12, 24] and low reaction times [76]. Distractions in MR environments can be categorized as congruent (aligned with the task) or incongruent (deviating from the task) [46, 74]. *Congruent distractions (CD)* might be more easily absorbed into the experience, but they still increase cognitive load as they require mental effort to integrate them into the ongoing task [66]. However, *incongruent distractions (ID)* are more likely to cause significant disruptions because they conflict with the user’s task goals, forcing the brain to process unrelated information [77]. This mismatch between the distraction and the task increases cognitive load [43] and makes it more difficult to maintain focus, thereby increasing the likelihood of BIP [63]. These distractions can negatively impact users’ reaction

times [36], as their attention is forced to shift between real-world and virtual stimuli [75]. To address this gap in the literature, we conducted an exploratory within-subjects study (N = 54). We evaluated the relationship between presence, cognitive load, BIP, and reaction time across **CD**, **ID**, and no-distraction (**ND**) conditions caused by secondary tasks. We aim to explore how different types of distractions impact these key variables and how they interact in real-time MR environments.

To explore this relationship further, we conducted an exploratory within-subjects study (N = 54). We evaluated the relationship between presence, cognitive load, BIP, and reaction time across congruent (**CD**), incongruent (**ID**), and no-distraction (**ND**) conditions induced by secondary tasks. Assuming that these secondary tasks successfully create distinct distraction scenarios, our study aims to address the following key research questions (RQ):

RQ1: How do different types of distractions (congruent vs. incongruent) influence the likelihood of a BIP?

RQ2: What is the effect of distractions on cognitive load, and how does BIP interact with cognitive load?

RQ3: What is the relationship between presence and reaction time, and how does cognitive load mediate this relationship?

2 BACKGROUND AND RELATED WORK

2.1 Terminologies

In this paper, when we refer to **MR**, we align with the definition based on Milgram’s reality-virtuality continuum, positioning it near Augmented Reality (AR), where virtual components blend naturally with the physical world.

Presence refers to the psychological sense of “being there” in a VE, requiring visual realism and natural interactions [51].

Cognitive Load refers to the mental effort required to process information and perform tasks.

Reaction time represents the time it takes for a user to respond to a specific stimulus in the VE, distinct from the total time required to complete an entire task, often referred to as task completion time; these terms are not used interchangeably for the rest of the paper.

Immersion refers to the objective quality of the system’s ability to provide a coherent and rich sensory experience that fully engages the user [22]. Immersion represents the degree of sensory fidelity, which influences but does not equate to presence. Cognitive and perceptual processes are the main factors influencing presence and our main focus in this paper.

2.2 Measuring Presence

Traditionally, presence has been measured using subjective self-reports. Popular questionnaires include the Presence Questionnaire (PQ) [78], Igroup presence questionnaire (IPQ) [53, 54], and others [16, 23, 33, 34, 46, 47]. These tools evaluate the user’s subjective sense of presence after engaging with the VE, asking participants to reflect on how immersed they felt. These questionnaires have been instrumental in understanding users’ overall impressions of immersion, but they are limited by their retrospective nature, which can introduce bias and fail to capture real-time fluctuations in presence [50]. The second challenge with subjective questionnaires in MR is that many AR/MR-specific questionnaires [17, 21, 48, 64, 68] lack the validation and reliability testing as their VR counterparts [65]. These newer tools often have not undergone rigorous validation processes like Cronbach’s alpha, Kaiser-Meyer-Olkin tests, or factor analyses [19, 69]. Therefore, researchers frequently rely on well-established VR questionnaires, which, while not perfect, offer more reliable alternatives than unvalidated MR-specific tools [73]. Prior studies have shown that VR-developed questionnaires can still be effective in AR/MR contexts, especially when all users experience similar environments, even if they are not fully immersive [42, 70].

Alternatively, researchers have explored objective measures of presence that offer real-time feedback. Physiological measures such as heart rate and electrodermal activity [37] and neurological signals [15] have been proposed as potential indicators of presence. However,

these measures can be intrusive and may disrupt the immersive experience by adding physical discomfort. Another objective measure is reaction time, a behavioral metric reflecting cognitive and motor response to stimuli. Reaction time is less invasive than physiological measures and offers a way to track presence continuously throughout the MR experience [39, 76]. Faster reaction times are often associated with higher levels of presence, while slower reaction times may signal BIP [18]. Reaction time has been widely used in various fields to assess attention [29] and decision-making [28] and is gaining attention in MR research for its potential to serve as a proxy for real-time presence [3, 10, 59, 67].

2.3 Break in Presence (BIP)

Slater et al. [59] defined BIP as moments when users shift focus from the virtual to the real-world, causing a noticeable drop in presence [9, 77]. BIP can be triggered by the following elements:

- (a) **External distractions.** are real-world interruptions such as ambient noises, physical disturbances, or system notifications. External distractions force the user to divide their attention between the real-world and the VE, increasing cognitive load and making it harder to maintain presence [62].
- (b) **Internal inconsistencies.** Glitches, lags, or unrealistic behaviors in the VE can create perceptual dissonance, forcing the user to question the validity of the VE and breaking their mental model [58], leading to BIP [59, 72].
- (c) **Cognitive overload.** occurs when users are overwhelmed by task complexity or by processing real and virtual stimuli simultaneously [26], or discontinuation for secondary task [46]. This leads to slower reaction times, reduced performance, and a higher likelihood of BIP [8, 20].

BIP has also been explored as an objective measure of presence [4], and capturing real-time fluctuations. Instead of relying solely on post-experience questionnaires, researchers can track presence dynamically by observing when BIPs occur. While BIP-specific validated tools are limited, drops in presence scores have been used as indirect indicators of BIP [60]. For instance, researchers like Slater et al. [59] asked participants to say “here” when experiencing a BIP, leading to the development of a presence counter, which uses a Markov chain to track BIP frequency [4]. However, reliance on self-reports introduces biases and delays, as participants had to be trained to recognize and report BIP events, potentially influencing their behavior. Subsequent studies linked physiological responses, such as heart rate and skin conductance, to correlate with BIP events [37, 57]. However, while physiological measures offer a more objective approach for continuously monitoring presence without requiring active user input, they can be intrusive and difficult to implement in naturalistic settings. Despite these limitations, BIP remains valuable for presence research, especially when combined with measures like reaction time.

2.4 Distractions Causing Disruptions in MR

Disruptions in MR arise from the dual interaction of the real and virtual worlds, which can complicate how users process sensory information and manage cognitive resources. Cognitive load and perceptual consistency are key factors in maintaining presence [32], and disruptions in either of these domains can lead to a BIP [9]. Cognitive load theory [66] divides cognitive load into intrinsic (task-related), extraneous (task-unrelated), and germane (learning-related) types. In MR environments, extraneous cognitive load is often increased by distractions from the real-world, such as ambient noise or physical movements, which divert attention away from the virtual task [30]. Incongruent stimuli, unrelated to the virtual task, demand additional mental processing and increase cognitive load, which can overwhelm the user’s ability to stay immersed in the VE [43].

Perception in MR relies on coherent sensory inputs from both the real and virtual worlds. Perceptual disruption, such as mismatched visual, auditory, or haptic feedback, can create a dissonance that undermines the user’s sense of presence. These disruptions require users to

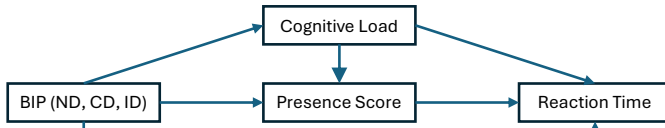


Fig. 2: The Cognitive Distraction Model for MR (CDM-MR).

reallocate cognitive resources to resolve the conflict between real and virtual stimuli, increasing the likelihood of a BIP [35]. Perceptual disruptions are particularly problematic in MR, where users are constantly required to navigate and reconcile inputs from both the real and virtual worlds [72].

2.5 BIP, Presence, Reaction time

As users experience BIP, their cognitive resources are redirected, which may result in slower reaction times [63]. Bracken et al. [7] found that reaction times increased during BIP events, showing that secondary task reaction time effectively measures disruptions in presence and attention. George et al. [20] found that interruptions in VR increase cognitive workload, disrupting presence and impairing reaction time, a key concern in MR where both real and virtual interruptions affect performance [28]. Schrader [52] expanded on this by examining how presence influences cognitive load and task performance, finding that increased cognitive load leads to slower reaction times and immersion. Breves and Stein [8] demonstrated how cognitive load can disrupt presence. Additionally, Nordahl et al. [41] proposed using controlled distractions as a measure of presence, showing that the ability of a system to keep users engaged despite distractions could be an effective indicator of immersion. Payne et al. [45] provided a framework for understanding the impact of secondary tasks on immersion in VE, which we extend by focusing on MR environments and scope down to use it as a tool to cause distractions.

2.6 Contribution Beyond Related Work

While previous studies have explored presence, BIP, cognitive load, and reaction time independently, none have examined how these factors intersect in MR environments. This study offers several key contributions: (1) We introduce the theoretical cognitive distraction model for MR, explaining how distractions in MR could impact users (§3). (2) We empirically demonstrate the distinct effects of distractions on cognitive load, reaction time, and presence (§4). (3) We highlight reaction time as a real-time proxy for presence, offering a non-intrusive method to assess immersion (§5). Unlike prior studies that linked reaction time to presence without considering the impact of distractions, our study adds depth by exploring how distractions modulate the relationship between reaction time and presence (§6), particularly in environments with both virtual and real-world stimuli.

3 APPROACH

3.1 Conceptual Model

In this section, we present our theoretical *Cognitive Distraction Model for MR (CDM-MR)* and integrate the concepts discussed in the background to explain how distractions influence the user experience in MR, as shown in Figure 2.

3.1.1 Cognitive Load and Attention

To understand how distractions impact presence and BIP, we first examine the role of cognitive load and attention. In MR environments, managing both virtual and real-world stimuli requires significant cognitive effort, particularly when distractions are present. We assume that real-world stimuli are inherently more disruptive due to their sensory modality and external origin; however, individual differences in attention switching may vary. Cognitive load theory [66] suggests that distractions increase extraneous cognitive load; because of that, when distraction occurs, users have fewer resources to maintain a presence in the VE [63]. This reflects the user’s need to allocate mental resources across multiple competing sources of information. This is supported by attention theory [40], which posits that human attentional capacity is

finite. When users are faced with increasing cognitive demands, particularly when it occurs abruptly due to unexpected distractions, their ability to maintain attention on a primary task may be reduced. This aligns with the concept of dual-task interference [44], which suggests that performing multiple tasks concurrently can lead to performance declines as attentional resources are divided between tasks, especially when the two competing tasks are conflicting, such as often incongruent nature of transitioning between real and virtual inputs, this task-switching incur cognitive costs due to attention reallocation [38].

3.1.2 Reaction Time and Presence

BIP itself introduces additional cognitive load. Once a BIP occurs, users must expend mental effort to reorient themselves back into the VE, thereby increasing the cognitive load beyond what is required for the primary task alone. We hypothesize that this creates a feedback loop: distractions increase cognitive load, leading to BIP, further exacerbating cognitive load as the user attempts to regain immersion in the virtual world. As the cognitive resources are redirected due to BIP, it may result in slower reaction times [7]. This concept is further supported by Baumeister’s resource depletion model [5], which suggests that cognitive resources are finite, and each competing task or distraction (in our case, BIP) depletes the available attentional capacity for the primary task. In MR environments, frequent BIP events could lead to decreased presence and degraded performance due to the increased mental effort required to re-engage with the virtual world [63].

3.1.3 Distractions Types

In §2.4, we discussed how distractions in MR affect users cognitively and perceptually; now we will discuss the nature of distractions by categorizing them into two distinct types:

Congruent distractions (CD), align with or are either contextually relevant to the task being performed in the VE or integrated within the VE to initiate parallel secondary tasks. For example, imagine a user in an MR environment navigating through a virtual space while following virtual cues. **CD** might involve a virtual pop-up or instruction that fits naturally within the virtual task, such as an additional navigation prompt or task-related feedback. This type of distraction is integrated within the VE and aligns with the user’s mental model of the task, making it easier to absorb without breaking the illusion of immersion. The assumption is that the likelihood of a BIP is lower because the distraction is part of the VE.

Incongruent distractions (ID), involve stimuli that break the illusion of immersion and repeatedly pull the user back to the real world. For example, an unrelated real-world notification, a sudden noise, or a secondary task that requires interactions in the real-world would force the user to shift their attention away from the virtual space. These distractions break immersion and increase the likelihood of BIP by diverting attention away from the virtual task and drawing focus back to the real world. The critical difference between **CD** and **ID** lies in the cognitive and attentional demands they impose. In MR, **CD** may add complexity but is easier to incorporate into the task structure, whereas **ID** disrupts cognitive flow by introducing unrelated real-world interactions and creating this back-and-forth between virtual and real words, making it harder for users to maintain a presence.

3.2 Hypotheses

Building on the *CDM-MR*, our experimental design will focus on testing the following hypotheses:

- **H1**: Incongruent distractions cause BIP, while congruent distractions do not cause BIP.
- **H2**: Distractions introduced through secondary tasks increase cognitive load.
 - **H2a**: Incongruent distractions result in a greater increase in cognitive load compared to congruent distractions.
 - **H2b**: BIP leads to an increase in cognitive load.
 - **H2c**: BIP increases the cognitive load associated with distraction tasks.
- **H3**: Presence and reaction time are correlated.

- **H3a**: Presence mediates the relationship between cognitive load and reaction time.

3.3 Design

For our study, we introduced distractions using secondary tasks [7, 45] rather than other common methods such as mimicking system glitches [46], whiteouts [18, 58], or interface interruptions [59]. Our decision stems from the need to create controlled and repeatable disruptions that allow us to evaluate their impact on our study variables. We assume that secondary tasks effectively simulate the types of interruptions users experience in real-world MR settings, though they may not capture spontaneous environmental changes. Unlike system glitches or interface errors, which are often unpredictable and difficult to standardize across participants, secondary tasks provide a consistent and clear way to assess how participants reacted between the primary and secondary stimuli. This approach allows us to maintain experimental control by ensuring that each participant encounters distractions of the same nature and frequency, making comparisons across conditions more reliable. Additionally, while glitches or sudden interruptions may cause frustration or cognitive overload, they could also introduce unintended variables such as frustration, confusion, or system usability issues that are difficult to disentangle from presence or BIP. In contrast, secondary tasks can be tailored to focus directly on cognitive load without introducing usability confounds [62].

3.3.1 Task Structure

Primary Task. Participants complete an image sorting task across three conditions: No Distraction (**ND**), **CD**, and **ID**. Each image is either a living or non-living object. The participant categorizes the image by dragging it to one of two labels: Living or Non-Living. This task remains consistent across all three conditions. To ensure the images used do not trigger any unintended emotional responses (fear, disgust, confusion, etc.), we select 90 images from the Open Affective Standardized Image Set (OASIS) [31]. This image set contains 900 images rated by 822 participants on pleasantness and excitement. We chose images with the highest “pleasantness” ratings to minimize potential triggers.

Secondary Task. **ND** serves as the baseline; participants perform the primary image sorting task without distractions. In the **CD** and **ID**, participants perform secondary tasks. In the **CD**, the primary task remains the same as in **ND** but with an added congruent distraction. A virtual buzzer appears randomly between trails (details in §4.3), prompting the participant to turn off the buzzer by pressing a virtual button. The distraction is congruent because the button is integrated into VE, reducing the degree to which it disrupts the immersive experience. Therefore, while the task becomes more complex due to the secondary task, we hypothesize that the cognitive transition between sorting images and turning off the buzzer remains smooth because both actions occur in the same virtual space.

In the **ID**, participants perform the primary task with a real-world distraction. A pop-up appears in the VE, but to turn off the buzzer, the participant presses a physical button in the real world outside the VE, located in the same physical position as the virtual button in the **CD**. This forces the participant to break immersion to perform the real-world task. Unlike the **CD**, **ID** requires participants to repeatedly shift between the VE and real-world tasks, introducing a cognitive discontinuity that we hypothesize potentially increases cognitive load and reduces presence. We assume that requiring participants to interact with a real-world object physically reliably triggers a BIP event; however, individual variation in spatial awareness and task familiarity may modulate this effect

Moreover, as reaction time is a primary measure, we designed both tasks to be simple enough for any adult to perform, avoiding requiring specialized skills. This minimizes cognitive overhead, allowing us to focus on the effects of distractions without adding unnecessary complexity.

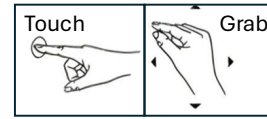


Fig. 3: Hand gesture for virtual button press and dragging images.

3.3.2 Task Interaction

We use near interactions on the Hololens2 [1], as several studies have confirmed the accuracy and reliability of its hand tracking [10, 61], allowing participants to interact with virtual objects using natural hand gestures. For grab interactions (to move/drag the image in the primary task), participants use their thumb and index finger together. They perform a pinch gesture, pinching their thumb and index finger near a virtual object (image) to grab it, as shown in Figure 3. Once the object is grabbed, participants can move their hand to move it to the desired location and release it by unpinching their fingers. For the virtual button press (distraction in **CD**), we use touch interaction; participants touch their fingers near a virtual button, and the Hololens’ hand tracking system detects this gesture, allowing the button to be pressed as it would in the real world, as shown in Figure 3. The participants extend their index finger and bring it close to a virtual button. As the finger approaches, a slight pressing motion simulates a button press (poke gesture), and the system detects the interaction.

To detect the real-world button press in the **ID**, we used audio processing in our Unity app to capture the sound emitted by the physical button. When the button is pressed, it produces a distinct beep sound at a frequency of 220 Hz with a sine waveform. We processed the incoming audio using the Fast Fourier Transform (FFT) to analyze the frequency spectrum. The FFT algorithm identifies the 220 Hz beep, signaling that the button has been pressed. We also implemented a 10-second timeout for the buzzer, ensuring it automatically turns off if the signal is not detected or the button is not pressed in the VE (**CD**) or the real world (**ID**).

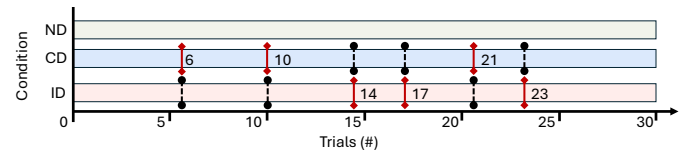


Fig. 4: The two patterns of distractions for participants: red with diamond ends and black with circular ends. Each participant experiences (6, 10, 21) in one condition and (14, 17, 23) in the other.

3.3.3 Order and Pattern of Distraction

To model the occurrence of distractions and mimic real-world unpredictability, we generated two distinct Poisson distributions. This maintains comparable difficulty and randomness across participants while preventing predictability or adaptation from using a single distribution. These distributions introduce distractions between the 6th and 25th trials, a boundary determined through the pilot study (§4.5) to avoid early distractions. Using a Poisson process with a rate parameter of $\lambda = 0.15$, we generate exactly three distractions, subsampling any excess events to maintain randomness while controlling the number. Distribution 1 (trials 6, 10, 21) is applied to the first distraction condition, and distribution 2 (trials 14, 17, 23) to the other. This setup (shown in Figure 4) focuses on how the nature of distractions impacts our variables rather than their timing.

4 USER STUDY

4.1 Participants

We conducted an a priori power analysis for our within-subject study comparing three conditions: **ND**, **CD**, and **ID** with a medium effect size (0.5), an error probability $\alpha = 0.05$, and a desired power of 0.95 the analysis indicated a minimum of 43 participants. To account for potential attrition of 25%, we recruited 54 participants from 18 to 35 years ($M=23$, $STD=4.2$), 26 were female, 1 non-binary, and 27 male.

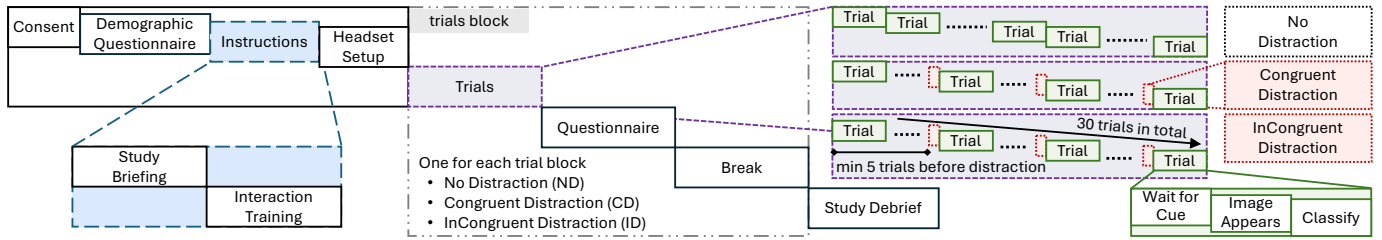


Fig. 5: User study procedure.

In terms of VR experience, 8 had never used it, 19 used it rarely, 20 occasionally, and 7 frequently. For the AR/MR experience, 21 had never used it, 16 rarely, 14 occasionally, and 3 frequently. Regarding gaming, 6 participants had no experience, 7 rarely played, 12 occasionally, and 29 frequently. All participants provided written informed consent, receiving \$15 for their participation. Each participant had normal or corrected-to-normal vision. The study was approved by the University of Massachusetts Amherst IRB, Human Research Protection Office.

4.2 Material

We used the HoloLens 2 [2] for its transparent display, spatial audio, integrated eye, spatial, and hand-tracking, and dual displays with a 1440×936 resolution and 110 FoV with Qualcomm Snapdragon 850 CPU. The VE is implemented in Unity with XR plugins and custom scripts for hand tracking and recording reaction time. Unity’s Microphone API and audio processing are also used to detect the real-world button beep in the **ID** to turn off the buzzer.

4.3 Experimental Task

This section summarizes the cue, interaction, and feedback across all conditions in primary and secondary tasks.

Cue. In all conditions, during each trial of the primary task, an image to be sorted appears in the VE on a slow-moving trolley, prompting participants to sort it as living or non-living, as shown in Figure 1. A virtual pop-up with a buzzer appears randomly in secondary task conditions (**ID**, **CD**). In **ID**, the buzzer lacks spatial audio to mimic a real-world source, while in **CD**, spatial audio signals that the sound originates from the VE.

Interaction. For the primary task, participants perform a pinch gesture to grab and drag the image to the appropriate label (living or non-living) in the VE. For the secondary task in **CD**, participants press a virtual button with a poking gesture to stop the buzzer. In **ID**, participants turn off the buzzer by pressing a physical button outside the VE on the physical table in front of them.

Feedback. In all conditions, once the image is sorted correctly, it disappears from the VE, signaling the completion of the primary task. For the secondary task, the buzzer stops when the virtual or physical button is pressed, resuming the primary image sorting task.

The equivalence between **CD** and **ID** lies in the interaction structure: both involve turning off a buzzer in the same location, requiring the same hand movement, with one in the VE and the other in the real world. We assume that participants interpret the physical and virtual button presses as functionally equivalent despite differences in sensory feedback, though additional factors such as tactile familiarity may influence their response.

4.4 Measurement

We measure presence with the PQ and IPQ after each condition. The PQ evaluates factors such as the possibility to act (ACT) and examine (EXAM), realism (REAL), self-evaluation (EVAL), and interface quality (IFQUAL). The IPQ measures factors such as spatial (SP) and general presence (GP), realism, and involvement (INV). The presence scores are derived from 33 items (14 IPQ and 19 PQ) on a 7-point scale. To better fit the MR with minimal text change as validated by [72, 74], only one item was adapted, changing the “computer-generated world” to “augmented reality” (as it is more common term than MR, more

details in §2.2). For each participant, presence was measured once at the end of each condition, resulting in 3 presence data points per participant (one for **CD**, **ID**, and **ND** conditions). Higher scores indicate a stronger sense of presence, and lower scores suggest reduced presence. Presence is used as a dependent variable to evaluate the effects of distraction type (**H1**, **H3**), as well as an independent variable (**H1**) and mediator in relation to cognitive load (**H2c**) and reaction time (**H3a**).

In the absence of a widely validated BIP measurement tool, we used the same presence questionnaires (PQ and IPQ) as a proxy for BIP. Prior research supports using presence scores to infer BIP events, as significant drops in presence often indicate moments when the user’s attention shifts from the virtual to the real world [59]. This approach assumes that presence and BIP are inversely related; a decrease in presence correlates with an increased likelihood of BIP. While this method captures aggregate trends, it does not detect the precise moment a BIP occurs, which remains a limitation.

Cognitive load is measured with the NASA Task Load Index (NASA-TLX) [25] after each condition. It is treated as a dependent variable (**H2**, **H2a**, **H2b**, **H2c**) to assess the impact of distraction types and BIP on participants’ cognitive workload. Cognitive load also functions as an independent variable (**H3a**) when exploring its influence on reaction time, with presence acting as a mediator. Our software recorded reaction time on the HoloLens2 in milliseconds (ms) continuously during the image sorting task across all conditions (30 data points each). We analyzed reaction time as a dependent variable to examine its relationship with presence (**H3**) and cognitive load (**H3a**) in response to different types of distractions.

4.5 Pilot Study

We conducted a pilot study with five participants to refine key parameters for the main experiment, including trial duration, frequency, timing and placement of distractions, and the selection of buzzer and beep (real-world) sounds and visual cues. We primarily relied on participant feedback to avoid experimenter bias, except for the trial duration, which was determined based on the reaction time data. The maximum reaction time recorded was 6.1 seconds, and the minimum was 2.2 seconds. To ensure consistency, we set the trial duration at 10 seconds, allowing ample time for all participants; however the extreme outliers may not fit this range. After testing trial counts between 20 and 40, we selected 30 trials to balance task engagement and minimize fatigue while providing sufficient opportunities for multiple distractions. We also tested various buzzer sounds and visual cues, choosing noticeable but not overly disruptive options. Based on feedback, the distraction timing was optimized to start from the 6th trial and end at the 25th, with 3 (we tested 2-6) distractions being the most optimal among all the participants. We also fine-tuned the scale of images and labels for clarity. We adopted a slow-moving trolley mechanism for the image to appear in the scene, which participants found intuitive. Finally, we tested real-world button press beeps with two waveforms (sine, square) and frequencies ranging from 100 to 440 Hz to ensure reliable detection; participants preferred the 220 Hz beep with a sine wave.

4.6 Study Procedure

The procedure for the study involved several steps, as shown in Figure 5. First, participants were given a consent form to read and sign before proceeding with the experiment. After consent was obtained,

Table 1: Summary of statistics for presence scores (PQ + IPQ), NASA TLX assessment, and reaction time across conditions.

	N	μ	σ	Std. Error	95% CI (mean)		Min	Max
					CI[5%]	CI[95%]		
Presence Score (PQ + IPQ, Scale: 0-7)								
ND	54.0	4.65	0.85	0.12	3.18	5.75	2.59	5.83
CD	54.0	4.34	0.97	0.13	2.59	5.73	2.21	6.18
ID	54.0	4.19	0.88	0.12	2.56	5.35	2.29	6.52
Cognitive Load (NASA TLX, Scale: 0-7)								
ND	54.0	1.62	1.13	0.15	1.0	4.67	1.0	4.80
CD	54.0	2.92	1.13	0.15	1.33	4.80	1.00	5.00
ID	54.0	4.03	1.12	0.15	1.26	4.80	1.00	6.00
Reaction Time (ms)								
ND	54.0	3.90	1.45	0.20	1.45	6.22	1.0	6.67
CD	54.0	5.50	2.21	0.30	2.26	8.93	1.0	8.52
ID	54.0	7.88	2.54	0.35	3.15	10.57	1.0	9.72

the participants were briefed on the experiments. During the briefing, participants were informed about the MR app and what to expect during the experiment. They were also instructed on the headset interactions and gestures needed to complete them, interactions required during the tasks, and how to perform them. The briefing included details about the visual stimuli used in the experiment, such as their color, shape, duration, cues, and feedback mechanism in each condition. Following the briefing, participants were asked to fill out a demographic questionnaire, which included questions about their gender, age, and familiarity with technology. The core experiment consisted of three conditions: **ND**, **CD**, and **ID**. These conditions were presented in a randomized and counterbalanced order using a Latin Square design to reduce any order effects across participants. In addition, a random distribution was assigned to the distraction tasks for each user, such that when Distribution 1 was used in the **CD**, Distribution 2 was assigned to the **ID**, and vice versa (§3.3.3). This ensured that no participant experienced the same distribution for both distraction conditions, preventing adaptation and predictability in the task structure. Participants were then put on the headset, and each condition was presented one after the other, with a break to complete the questionnaires assessing their cognitive load and presence for each condition.

In all three conditions, participants performed 30 trials (10 seconds apart) of the primary image sorting task, where they were shown a series of images and asked to categorize them as either “living” or “non-living” by performing a pinch-and-drag gesture. Participants completed two practice interactions before each condition, which were recorded but excluded from the analysis. In the **ND**, participants performed the sorting task without any distractions, which took ≈ 5 minutes to complete. In the **CD**, participants experienced a distraction in the form of a virtual pop-up accompanied by a buzzer. The pop-up appeared randomly during the trials, and participants were required to turn off the buzzer by performing a poking gesture to press a virtual button that appeared next to the pop-up, taking ≈ 6 minutes to complete. In the **ID**, participants faced a similar distraction, but this time, the buzzer sound came from the real world. Instead of pressing a virtual button, participants turned off the buzzer by pressing a physical button outside the VE on the table in front of them; in total, this condition took ≈ 6 minutes. After completing all three conditions, participants were debriefed, and the purpose of the study was explained. The total duration of the session was ≈ 45 minutes, including consent, briefing, training, headset calibration, the three experiments, three rounds of questionnaires, and debriefing.

5 RESULTS

5.1 Statistical Analysis

Statistical analysis was performed using Python’s Pingouin Library [71]. We first report detailed descriptive statistics for each condition (**ND**, **CD**, **ID**) for presence scores (using combined PQ and IPQ scores), cognitive

Table 2: Within subjects (**ND**, **CD**, **ID**) repeated measures ANOVA using presence scores as the dependent variable. **SS**: Sum of Squares, **df**: Degrees of Freedom, **MS**: Mean Squares, **F**: F-values, **p**: uncorrected p -value, **ng2**: generalized eta-square effect size, and **eps**: Greenhouse-Geisser epsilon factor.

Source	SS	df	MS	F	p	ng2	eps
Condition	5.88	2	2.94	12.51	0.00	0.04	0.93
Error	24.91	106	0.23				

load (using NASA TLX assessment), and reaction time in Table 1. We used Q-Q plots to confirm the normality and Levene’s test to verify the homoscedasticity for all conditions and variables. To examine the effect of condition on the presence score, we performed repeated measures ANOVA and reported the results in Table 2. To further check the ANOVA results and understand the difference between conditions, we performed a post-hoc t-test and reported the statistics in Table 3. Similarly, to understand the effect of condition on cognitive load, we performed repeated measures ANOVA and reported the results in Table 5. We also performed a post-hoc t-test for this case and reported the statistics in Table 6. We confirmed the sphericity of the data using the Mauchly test. Also, Cronbach’s alpha value of 0.874 with a 95%CI of [0.816, 0.917] demonstrated the high reliability and internal consistency of presence scores across conditions.

5.1.1 Presence Scores

The analysis of presence scores (a combination of PQ and IPQ) across the three conditions (**ND**, **CD**, **ID**) reveals notable differences. Descriptive statistics (Table 1) show the highest presence scores in the **ND** condition ($M = 4.65, SD = 0.85$), followed by **CD** ($M = 4.34, SD = 0.97$), and the lowest in **ID** ($M = 4.19, SD = 0.88$). The 95% confidence intervals for the means suggest differences across conditions, particularly between **ND** and **ID**, where there is no overlap between the intervals. A repeated measures ANOVA (Table 2) was conducted to assess these differences statistically. The results indicated a significant effect of condition on presence scores ($F(2, 106) = 12.51, p < 0.001, \eta^2 = 0.04$). This finding suggests that the type of distraction, whether congruent or incongruent, significantly affects a user’s sense of presence in the MR. To further explore these effects, post-hoc paired t-tests were conducted to compare presence scores between conditions (Table 3). The comparison between **ND** and **CD** showed a significant reduction in presence in the **CD** condition ($t(53) = 3.14, p = 0.003, d = 0.34$), indicating that even **CD** can negatively impact the sense of presence. The **ND** vs. **ID** comparison revealed an even more significant reduction in presence with **ID** ($t(53) = 4.54, p < 0.001, d = 0.53$), demonstrating that incongruent distractions have a stronger detrimental effect on presence compared to congruent ones. These results confirm that both types of distractions reduce presence, with **ID** having a more pronounced effect.

In Table 4, we show the values for each condition by summarizing the distribution of presence scores, where each participant contributes one score per condition. For both the PQ and IPQ subscales, participants in the **ND** consistently reported higher presence scores compared to the **CD** and **ID**. In terms of the overall PQ score, participants in the **ND** had the highest average presence score ($\mu = 5.46, \sigma = 0.89$), followed by **CD** ($\mu = 4.88, \sigma = 1.02$) and **ID** ($\mu = 4.21, \sigma = 0.90$). The same trend holds across the PQ subscales, where **ND** consistently scores higher, particularly in the ACT subscale, with **ND** having a mean of 5.71, while **CD** and **ID** drop to 4.69 and 4.24, respectively. Similarly, the IPQ scores reflect a similar pattern. The **ND** condition reports a mean presence score of 4.37 ($\sigma = 0.88$), with the **CD** and **ID** scoring 3.78 ($\sigma = 1.08$) and 3.34 ($\sigma = 0.85$), respectively. INV, IPQ-REAL, and SP also show higher scores for the **ND**, which suggests that distractions substantially reduce participants’ perceived presence.

5.1.2 Cognitive Load

The cognitive load was measured using the NASA TLX assessment across three conditions (**ND**, **CD**, **ID**). Descriptive statistics (Table 1) show that cognitive load was lowest in the **ND** ($M = 1.62, SD = 1.13$), increased in the **CD** ($M = 2.92, SD = 1.13$), and was highest in the **ID**

Table 3: Paired samples t-test statistics for (ND, CD) and (ND, ID) using presence scores. **T**: T-value, **df**: Degrees of Freedom, **p**: p-value, **CI95%**: confidence interval of the differences in means, **cohen-d**: Cohen d effect size, and **power**: Achieved power of the test.

Pair	T	df	p	CI95%	cohen-d	power
ND, CD	3.14	53	0.003	[0.11, 0.50]	0.34	0.68
ND, ID	4.54	53	0.000	[0.26, 0.66]	0.53	0.97

Table 4: Summary of descriptive results PQ and IPQ questionnaires subscales scores across conditions: No distraction (ND), congruent distraction (CD), and incongruent distraction (ID). We provide mean (μ) and standard deviation (σ) values for each experiment. The subscales for PQ and IPQ questionnaires are realism (PQ-REAL, IPQ-REAL), possibility to act (ACT), interface quality (IFQUAL), possibility to examine (EXAM), self-evaluation of performance (EVAL), involvement (INV), general presence (GP), and spatial presence (SP).

Quest.	ND		CD		ID	
	μ	σ	μ	σ	μ	σ
PQ-REAL	5.08	1.07	4.65	1.16	4.17	1.00
ACT	5.71	1.02	4.69	1.26	4.24	1.02
QUAL	5.62	1.06	5.03	1.01	4.30	1.34
EXAM	5.36	1.06	5.04	1.07	4.23	1.18
EVAL	5.85	1.16	5.58	1.45	4.12	1.54
PQ	5.46	0.89	4.88	1.02	4.21	0.90
INV	3.90	1.12	3.26	1.22	2.91	0.98
IPQ-REAL	4.26	0.94	3.39	1.09	3.11	1.01
SP	4.72	1.13	4.27	1.42	3.78	1.19
GP	4.94	1.43	4.89	1.59	3.85	1.32
IPQ	4.37	0.88	3.78	1.08	3.34	0.85

($M = 4.03, SD = 1.12$). The 95% confidence intervals for the mean cognitive load across conditions indicate non-overlapping intervals, especially between ND and ID, suggesting significant differences between these conditions. To statistically assess the impact of the distractions on cognitive load, a repeated measures ANOVA was performed (Table 5). The results revealed a significant main effect of condition on cognitive load ($F(2, 106) = 15.60, p < 0.001, \eta^2 = 0.08$), indicating that the type of distraction significantly influences cognitive load in the MR. To further investigate these differences, post-hoc paired t-tests were conducted to compare cognitive load between conditions (Table 6). The comparison between ND and CD showed a significant increase in cognitive load in the CD ($t(53) = 3.94, p < 0.001, d = 0.60$), indicating that CD imposes additional cognitive load compared to ND. Similarly, the comparison between ND and ID showed an even larger increase in cognitive load for ID ($t(53) = 3.32, p < 0.001, d = 0.60$), highlighting the greater cognitive load caused by ID. These results demonstrate that both types of distractions significantly increase cognitive load, with ID having a more pronounced impact.

Although the accuracy of the task is not part of our hypotheses, Table 7 reports the accuracy of participants in sorting living and non-living images across three conditions. The highest accuracy is observed in the ND (86.80% mean accuracy), indicating that participants performed the sorting task most effectively when no distractions were present. In the CD condition, accuracy drops to 74.92%, suggesting that even CD may reduce task performance, although participants still maintain relatively high accuracy. The lowest accuracy is seen in the ID (52.42%), where incongruence appears to have caused significant difficulty in performing the task. The increasing standard deviation from ND (5.26%) to CD (14.85%) and ID (20.98%) reflects greater variability in participant performance as distractions are introduced.

5.1.3 Reaction Time

The reaction time results indicate clear differences across all conditions. In the ND, participants had the fastest average reaction time ($\mu = 3.90,$

Table 5: Within subjects (ND, CD, ID) repeated measures ANOVA using cognitive load as the dependent variable. **SS**: Sum of Squares, **df**: Degrees of Freedom, **MS**: Mean Squares, **F**: F-values, **p**: uncorrected p-value, **ng2**: generalized eta-square effect size, and **eps**: Greenhouse-Geisser epsilon factor.

Source	SS	df	MS	F	p	ng2	eps
Condition	16.72	2	8.36	15.60	0.00	0.08	1.00
Error	56.80	106	0.54				

Table 6: Paired samples t-test statistics for (ND, CD) and (ND, ID) using NASA TLX. **T**: T-value, **df**: Degrees of Freedom, **p**: p-value, **CI95%**: confidence interval of the differences in means, **cohen-d**: Cohen d effect size, and **power**: Achieved power of the test.

Pair	T	df	p	CI95%	cohen-d	power
ND, CD	3.94	53	0.000	[0.34, 1.03]	0.60	0.99
ND, ID	3.32	53	0.000	[0.27, 1.11]	0.60	0.99

$\sigma = 1.45$), suggesting minimal interference from distractions. In the CD condition, the average reaction time increased significantly ($\mu = 5.50, \sigma = 2.21$), with some participants reaching up to 9.52 seconds, as depicted in the cubic spline model in Figure 8 (a,b). Finally, in the ID, participants took the longest to react ($\mu = 7.88, \sigma = 2.54$), with recovery times remaining high across trials and individual times reaching 10.72 seconds in the distributions shown in Figure 8 (c,d). These results demonstrate that ID substantially impacted reaction times, with recovery taking much longer compared to CD and ND. The longer reaction times here most likely reflect BIP-related cognitive load rather than external influences such as confusion or fatigue, although this distinction may not always be clear.

5.2 Correlation and Mediation Analysis

Next, we quantified the strength and direction of the relationship between cognitive load, presence, and reaction time. First, we computed Pearson’s correlation coefficient and applied linear regression to assess the relationship between cognitive load and presence; with regression coefficients of -0.43 and 5.64 . The $R^2 = 0.35$ and a p-value of 0.000 indicate a strong negative relationship between cognitive load and presence. The Pearson’s correlation coefficients for the ND, CD, and ID are $-0.79, -0.73,$ and -0.69 , respectively, with an overall correlation of -0.59 . These results suggest that presence decreases across all conditions as cognitive load increases. Similarly, we analyzed the relationship between presence scores and reaction time using Pearson’s correlation and linear regression as shown in Figure 7, the regression coefficients were -1.78 and 13.87 , with an $R^2 = 0.42$ and a p-value of 0.000 . This indicates a significant negative relationship between presence and reaction time, where higher presence scores are associated with faster reaction times. The Pearson’s correlation coefficients for ND, CD, and ID are $-0.68, -0.84,$ and -0.80 , respectively, with an overall correlation of -0.64 .

We conducted a mediation analysis to understand how presence mediates the relationship between cognitive load and reaction time, shown in Figure 9. We computed both direct and indirect effects using this bootstrapping approach with 1000 resamples, and 95% CI revealed an indirect effect of presence, with a CI of $[0.20, 0.42]$. The total effect of cognitive load on reaction time was -1.78 with a standard error of 0.17 and a $p < 0.001$, indicating a significant relationship between cognitive load and reaction time. As expected, this suggests that a higher cognitive load leads to slower reaction times. The direct effect of cognitive load on reaction time, after accounting for the mediator (presence), was 1.68 with a standard error of 0.09 and a $p < 0.001$. This significant direct effect indicates that cognitive load still impacts reaction time independently of presence. The mediation analysis revealed an indirect effect of 0.30 , with a standard error of 0.07 . The indirect pathway was significant ($p < 0.001$), indicating that part of the effect of cognitive load on reaction time is mediated by presence. The coefficient of -0.44 with a standard error of 0.05 reflects that as cognitive load increases, presence decreases.

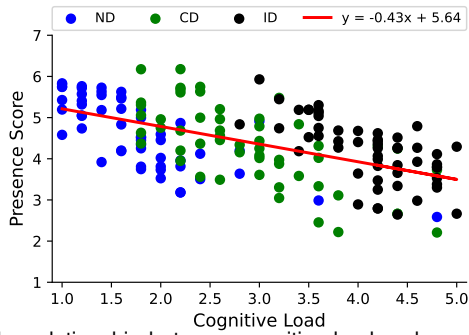


Fig. 6: The relationship between cognitive load and presence. The regression coefficients are -0.43 and 5.64. The value of R^2 is 0.35 with a p -value of 0.000. The Pearson’s correlation coefficients for **ND**, **CD**, **ID**, and overall are -0.79, -0.73, -0.69, and -0.59, respectively.

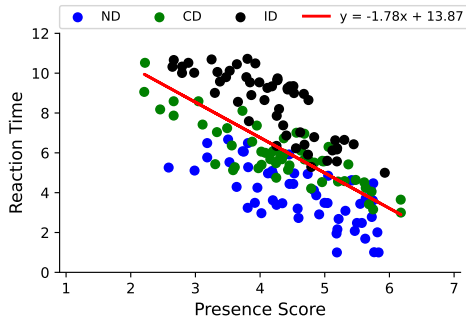


Fig. 7: The relationship between presence score and reaction time. The regression coefficients are -1.78 and 13.87 with R^2 of 0.42 and p -value of 0.000. The Pearson’s correlation coefficient is -0.68 for **ND**, -0.84 for **CD**, -0.80 for **ID**, and -0.64 overall.

Lastly, the overall mediation analysis in Figure 10 shows that cognitive load and presence scores significantly mediate the relationship between BIP and reaction time. The total effect of BIP on reaction time is significant, as demonstrated by a coefficient of 2.17(0.20), $p < 0.001$, suggesting that different BIP conditions impact how long participants take to respond.

Table 7: The accuracy of sorting living and non-living objects for study participants in **ND**, **CD**, and **ID** conditions.

Stat	ND	CD	ID
Mean (%)	86.80	74.92	52.42
Std. Dev. (%)	5.26	14.85	20.98

6 DISCUSSION

The study reveals how different distractions affect presence, cognitive load, and reaction time in MR. With medium effect sizes in cognitive load and presence score with high reliability (Cronbach’s alpha = 0.874), we assume that our manipulation to create distractions influences presence and cognitive load to a certain extent. We conducted a thematic analysis to identify comments related to real-world distraction. 34 participants explicitly mentioned pressing the button on the table, indicating that our manipulations were noticed.

6.1 Interpreting Results

Our result partially support **H1**. Both distraction types led to significant reductions in presence, but **ID** had a more pronounced effect compared to **CD**, as shown by the larger decrease in presence scores. Although we cannot directly measure BIP due to the lack of a validated tool, the reduction in presence scores suggests that BIP is more likely to occur

in the presence of **ID**, thereby partially supporting **H1**. For **H2**, the findings fully support this hypothesis. The results show that both **CD** and **ID** significantly increased cognitive load compared to the **ND**. The post-hoc t-tests indicate that **ID** leads to a greater increase in cognitive load compared to **CD**, supporting **H2a**. The evidence from the correlation analysis supports **H2b**. The significant negative correlation between presence and cognitive load across all conditions, particularly in **ID**, indicates that as presence decreases (a likely indication of BIP), cognitive load increases. Thus, **H2b** is accepted based on these indirect indicators of BIP leading to increased cognitive load. The mediation analysis suggests that presence mediates the relationship between cognitive load and reaction time, supporting the indirect link between cognitive load and presence in **H2c**. This finding shows that as cognitive load increases, it reduces presence, leading to slower reaction times. **H3** is supported by our data. There was a significant negative correlation between presence and reaction time across all conditions, indicating that as presence decreases, reaction time increases. **H3a** is partially supported; while the correlation between presence and reaction time is strong, the mediation analysis shows that presence only partially mediates the relationship between cognitive load and reaction time. This means presence influences reaction time, but cognitive load also directly affects it beyond the mediation by presence; thus, we fully accept **H3** and partially accept **H3a**.

6.2 Implications

Our study makes several contributions to the field of MR research. First, we advance the understanding of how distractions impact core experiential factors such as presence, BIP, cognitive load, and reaction time. Our work extends existing theories of presence and cognitive load by showing how these constructs interact under varying distraction conditions, leading to different performance outcomes. This theoretical insight could inform the design of more immersive and user-friendly MR systems by identifying key variables that need to be managed to maintain a high presence and low cognitive load. We introduce a conceptual model (CDM-MR) that positions presence as a mediating factor between cognitive load and reaction time. The integration of presence, cognitive load, and reaction time in a cohesive framework represents a novel contribution, offering a theoretical foundation for future studies aimed at optimizing MR environments. Lastly, by providing empirical evidence of these relationships, we contribute to the growing body of literature on user experience in MR environments, highlighting the nuances of different types of distractions (congruent and incongruent).

7 LIMITATIONS AND FUTURE WORK

In this section, we briefly list our limitations, study scope elements that may appear as limitations and future directions for research. This includes balancing experimental control, the scope of the research for a single paper, focusing on specific hypotheses and addressing the practical constraints.

7.1 Limitations

First, we used secondary tasks as the primary method of inducing BIP events. While this approach effectively controlled distractions in a repeatable manner, it does not represent the full range of distractions present in real-world MR environments, such as spontaneous environmental noise or system glitches. We did not include these other types of distractions because we aimed to create a controlled experimental setup that allowed us to systematically measure the effects of task-related distractions without introducing too many variables that could affect the consistency of the results.

Secondly, we used presence scores to infer BIP, which presents a limitation in accurately capturing BIP. This was necessary because no widely validated questionnaire or tool specifically designed to measure BIP exists. As a result, we had to rely on presence scores as a proxy, which may not fully capture the nuanced nature of BIP occurrences. However, for our study, the overall experience and the impact of BIP on cognitive load and reaction time were more critical than pinpointing the exact moments when BIP events occurred. Thus, while a more precise

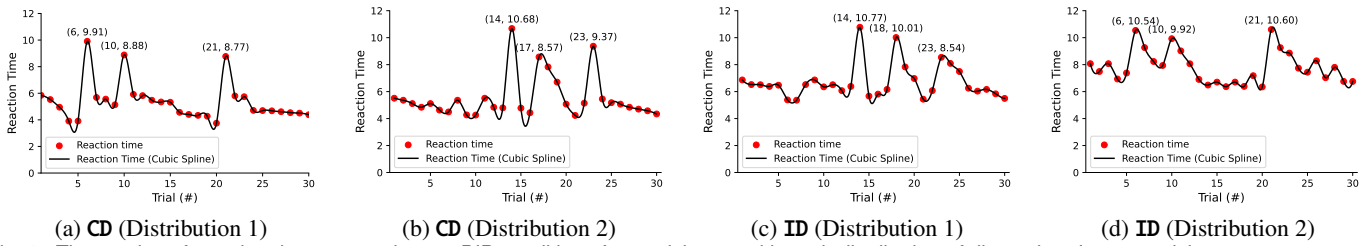


Fig. 8: Time series of reaction time across the two BIP conditions for participants with each distribution of distraction. In CD, participants can recover easily from the low presence, whereas in ID, they take much longer to recover from a break-in presence.

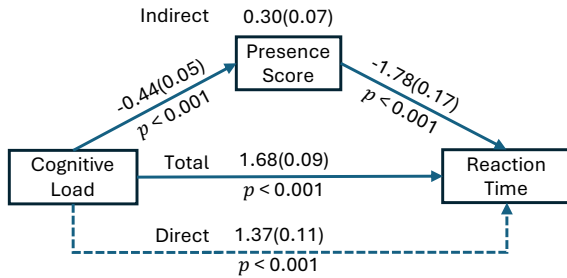


Fig. 9: Bootstrap analysis (95% CI 1000 resamples) for the mediation between cognitive load, presence, and reaction time, with a CI of [0.20, 0.42] for the indirect effect.

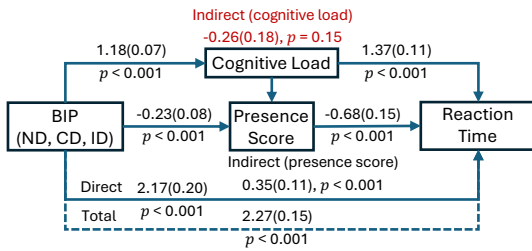


Fig. 10: Overall mediation analysis with cognitive load and presence scores as the mediators for the relationship between different conditions and the reaction time.

BIP measurement tool would have been ideal, the focus of our research on the broader effects of BIP justified the use of presence scores.

Additionally, we employed a limited number of questionnaires to measure cognitive load and presence. This choice was made to keep the study manageable in scope and to minimize participant fatigue, as extensive questionnaires could have led to longer sessions and lower data quality due to participant disengagement. Lastly, our sample of younger adults from a college town may limit the generalizability of our findings due to its limited age diversity. This participant pool reflects the practical limitation of being in a college town, where diversity in age and experience is constrained. However, since age was not the focus of any hypotheses, we did not prioritize age diversity in the study.

7.2 Study Scope

We chose not to explore the potential reversible relationship between presence and reaction time. While reduced presence correlates with slower reaction times, investigating whether faster reaction times directly enhance presence was not relevant to our study’s objectives. Since reaction time is already a well-established and objective metric, we focused on its direct relationship with presence rather than reversing the analysis. Exploring this reversibility would not have provided significant additional insights. Another scope limitation is the inability to establish causality, as it requires more complex experimental designs such as longitudinal studies, randomized controlled trials, or interventions. These approaches were not feasible due to time, scope, and resource constraints. Additionally, the tools and methods needed to

prove causality in the context of cognitive load, presence, and reaction time in MR environments are still evolving and may not be readily available or practical within the scope of this study.

7.3 Future research

In future research, we aim to address these limitations and expand the study’s scope. Incorporating physiological measures along with the reaction time, like EEG, could provide real-time data on brain activity, offering more insights into cognitive states during BIP events and fluctuations in presence. This would allow researchers to understand better the neural correlates of attention shifts between virtual and real-world stimuli. In terms of distractions, future work should explore more diverse methods, such as environmental stimuli (e.g., fluctuating light levels, ambient noise), system-based glitches (e.g., lags, graphical errors), or sensory overload (e.g., haptic disturbances). Investigating these distractions will provide a more comprehensive understanding of how various real-world disruptions impact cognitive load, reaction time, and presence.

Future work should also explore the relationship between BIP and task-technology fit. Understanding how well MR systems align with users’ tasks will offer valuable insights into presence, cognitive load, and reaction time. A better fit between task demands and technological capabilities could reduce cognitive strain and BIP events, thereby improving overall user performance and maintaining immersion. Studying this interaction will be essential for optimizing MR environments and ensuring that the technology effectively supports the user’s goals.

8 CONCLUSION

This study examined the impact of distractions on reaction time, presence, and cognitive load in MR environments, focusing on BIP events. The results show that ID increases cognitive load, slows reaction times, and reduces presence compared to CD. These findings highlight the importance of studying distractions to better manage their effects on performance and presence. Reaction time is strongly correlated with presence, offering real-time insights into immersion, but further research is needed to develop advanced measurement methods to capture these complex constructs fully.

ACKNOWLEDGMENTS

We thank the anonymous IEEE VR 2025 reviewers for their insightful comments and feedback. The research reported in this paper was sponsored by the National Science Foundation (NSF) under award 2237485.

REFERENCES

- [1] Direct hand manipulation. <https://learn.microsoft.com/en-us/windows/mixed-reality/design/direct-manipulation>. Accessed Sep-2024. 4
- [2] Hololens 2. <https://www.microsoft.com/en-us/hololens/hardware>. Accessed Sep-2024. 5
- [3] S. M. U. Arif, M. Brizzi, M. Carli, and F. Battisti. Human reaction time in a mixed reality environment. *Frontiers in Neuroscience*, 16:1410, 2022. 2
- [4] N. Averbukh and B. Velichkovskiy. Measurement of presence by a presence counter based on breaks in presence. In *IMS*, pp. 423–438, 2021. 2

- [5] R. F. Baumeister, M. Muraven, and D. M. Tice. Ego depletion: A resource model of volition, self-regulation, and controlled processing. *Social Cognition*, 18(2), 2000. 3
- [6] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. An introduction to 3-d user interface design. *Presence*, 10(1):96–108, 2001. 1
- [7] C. C. Bracken, G. Pettey, and M. Wu. Revisiting the use of secondary task reaction time measures in telepresence research: exploring the role of immersion and attention. *AI & society*, 29:533–538, 2014. 3, 4
- [8] P. Breves and J.-P. Stein. Cognitive load in immersive media settings: the role of spatial presence and cybersickness. *Virtual Reality*, 27(2):1077–1089, 2023. 2, 3
- [9] L. Brübach, F. Westermeier, C. Wienrich, and M. E. Latoschik. Breaking plausibility without breaking presence - evidence for the multi-layer nature of plausibility. *IEEE Transactions on Visualization and Computer Graphics*, 28, 2022. 1, 2
- [10] Y. Chandio, N. Bashir, V. Interrante, and F. M. Anwar. Investigating the correlation between presence and reaction time in mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 30(9), 2024. 1, 2, 4
- [11] Y. Chandio, V. Interrante, and F. M. Anwar. Human factors at play: Understanding the impact of conditioning on presence and reaction time in mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 30(5), 2024. 1
- [12] K.-W. Chen, Y.-J. Chang, and L. Chan. Predicting opportune moments to deliver notifications in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2022. 1
- [13] D. Cho, J. Park, G. J. Kim, S. Hong, S. Han, and S. Lee. The dichotomy of presence elements: The where and what. In *Proceedings IEEE Virtual Reality*, 2003. 1
- [14] I. Deary, D. Liewald, and J. Nissan. A free, easy-to-use, computer-based simple and four-choice reaction time programme: the deary-liewald reaction time task. *Behaviour Research*, 43, 2011. 1
- [15] A. Dey, J. Phoon, S. Saha, C. Dobbins, and M. Billinghurst. A neuro-physiological approach for measuring presence in immersive virtual environments. In *IEEE International Symposium on Mixed and Augmented Reality*, 2020. 2
- [16] J. Freeman, S. Avons, D. Pearson, and W. Ijsselsteijn. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence*, 8, 1999. 2
- [17] M. Gandy, R. Catrambone, B. MacIntyre, C. Alvarez, E. Eiriksdottir, M. Hilimire, B. Davidson, and A. C. McLaughlin. Experiences with an ar evaluation test bed: Presence, performance, and physiological measurement. In *IEEE International Symposium on Mixed and Augmented Reality*, 2010. 2
- [18] M. Garau, H. R. Widenfeld, A. Antley, D. Friedman, A. Brogni, and M. Slater. Temporal and spatial variations in presence: A qualitative analysis. In *Proceedings of the International Workshop on Presence*, 2004. 2, 4
- [19] A. Garcia-Santillan, E. Moreno-Garcia, J. Carlos-Castro, J. H. Zamudio-Abdala, and J. Garduno-Trejo. Cognitive, affective and behavioral components that explain attitude toward statistics. *Journal of mathematics research*, 4(5):8, 2012. 2
- [20] C. George, M. Demmler, and H. Hussmann. Intelligent interruptions for ivr: Investigating the interplay between presence, workload and attention. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018. 1, 2, 3
- [21] Y. Georgiou and E. A. Kyza. The development and validation of the ari questionnaire: An instrument for measuring immersion in location-based augmented reality settings. *International journal of human-computer studies*, 98, 2017. 2
- [22] M. Gonzalez-Franco, R. Pizarro, J. Cermeron, K. Li, J. Thorn, W. Hutabarat, A. Tiwari, and P. Bermell-Garcia. Immersive mixed reality for manufacturing training. *Frontiers in Robotics and AI*, 4:3, 2017. 1, 2
- [23] S. Grassini and K. Laumann. Questionnaire measures and physiological correlates of presence: A systematic review. *Frontiers in Psychology*, 11, 2020. 2
- [24] C. S. Green and D. Bavelier. Action video game modifies visual selective attention. *Nature*, 423(6939), 2003. 1
- [25] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research, 1988. *Advances in Human Psychology: Human Mental Workload*. Elsevier Science, 1988. 5
- [26] M. Hofer, T. Hartmann, A. Eden, R. Ratan, and L. Hahn. The role of plausibility in the experience of spatial presence in virtual environments. *Frontiers in Virtual Reality*, 1:2, 2020. 2
- [27] R. Kalawsky. The validity of presence as a reliable human performance metric in immersive environments. *3rd International Workshop on Presence*, 2000. 1
- [28] N. Khenak, J. Vezien, and P. Bourdot. Spatial presence, performance, and behavior between real, remote, and virtual immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, 26, 2020. 2, 3
- [29] S. J. Kim, T. H. Laine, and H. J. Suk. Presence effects in virtual reality based on user characteristics: Attention, enjoyment, and memory. *Electronics*, 10, 2021. 1, 2
- [30] M. Kocur, P. Schaubhuber, V. Schwind, C. Wolff, and N. Henze. The effects of self-and external perception of avatars on cognitive task performance in virtual reality. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2020. 2
- [31] B. Kurdi, S. Lozano, and M. R. Banaji. Introducing the open affective standardized image set (oasis). *Behavior research methods*, 49:457–470, 2017. 4
- [32] M. E. Latoschik and C. Wienrich. Congruence and plausibility, not presence: Pivotal conditions for XR experiences and effects, a novel approach. *Frontiers in Virtual Reality*, 3, 2022. 2
- [33] M. R. Leary and R. M. Kowalski. The interaction anxiousness scale: Construct and criterion-related validity. *Journal of personality assessment*, 61(1):136–146, 1993. 2
- [34] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff. A cross-media presence questionnaire: The itc-sense of presence inventory. *Presence: Teleoperators & Virtual Environments*, 10, 2001. 2
- [35] B. Liebold, M. Brill, D. Pietschmann, F. Schwab, and P. Ohler. Continuous measurement of breaks in presence: psychophysiology and orienting responses. *Media Psychology*, 20(3):477–501, 2017. 1, 3
- [36] B. Lok, S. Naik, M. Whitton, and F. P. Brooks. Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence*, 12(6):615–628, 2003. 2
- [37] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks. Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics*, 21, 2002. 2
- [38] S. Monsell. Task switching. *Trends in cognitive sciences*, 7(3):134–140, 2003. 3
- [39] K. Nesbitt, S. Davis, K. Blackmore, and E. Nalivaiko. Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays*, 48, 2017. 2
- [40] A. Ninio and D. Kahneman. Reaction time in focused and in divided attention. *Journal of Experimental Psychology*, 103(3):394, 1974. 3
- [41] R. Nordahl and D. Korsgaard. Distraction as a measure of presence: using visual and tactile adjustable distraction as a measure to determine immersive presence of content in mediated environments. *Virtual Reality*, 14:27–42, 2010. 3
- [42] D. Nunez. How is presence in non-immersive, non-realistic virtual environments possible? In *Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, 2004. 2
- [43] I. A. Odermatt, K. A. Buetler, N. Wenk, Ö. Özen, J. Penalver-Andres, T. Nef, F. W. Mast, and L. Marchal-Crespo. Congruency of information rather than body ownership enhances motor performance in highly embodied virtual reality. *Frontiers in neuroscience*, 15:678909, 2021. 1, 2
- [44] H. Pashler. Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, 116(2):220, 1994. 3
- [45] E. Payne, K. Battel, N. Baker, and N. Ward. Considerations when trying to quantify immersion with secondary tasks. *Frontiers in Virtual Reality*, 5:1212216, 2024. 3, 4
- [46] S. Putze, D. Alexandrovsky, F. Putze, S. Höffner, J. D. Smeddinck, and R. Malaka. Breaking the experience: Effects of questionnaires in vr user studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020. 1, 2, 4
- [47] B. Rammstedt and O. P. John. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41(1):203–212, 2007. 2
- [48] H. Regenbrecht and T. W. Schubert. Measuring presence in augmented reality environments: Design and a first test of a questionnaire. *ArXiv*, abs/2103.02831, 2021. 2
- [49] G. Riva, F. Davide, and W. Ijsselsteijn. Measuring presence: Subjective, behavioral and physiological methods. *Being there: Concepts, effects and*

- measurement of user presence in synthetic environments, 5, 2003. 1
- [50] S. Safikhani, M. Holly, A. Kainz, and J. Pirker. The influence of in-vr questionnaire design on the user experience. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, 2021. 1, 2
- [51] M. Sanchez-Vives and M. Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6, 2005. 1, 2
- [52] C. Schrader and T. J. Bastiaens. The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior*, 28(2):648–658, 2012. 3
- [53] T. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realism. *Zeitschrift für Medienpsychologie*, 15, 2003. 2
- [54] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10, 2001. 2
- [55] R. Skarbez, F. P. Brooks, and M. C. Whitton. Immersion and coherence: Research agenda and early results. *IEEE transactions on visualization and computer graphics*, 27(10):3839–3850, 2020. 1
- [56] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 2009. 1
- [57] M. Slater, A. Brogni, and A. Steed. Physiological responses to breaks in presence: A pilot study. In *Presence 2003: The 6th annual international workshop on presence*, vol. 157. Citeseer, 2003. 2
- [58] M. Slater, C. Guger, G. Edlinger, R. Leeb, G. Pfurtscheller, A. Antley, M. Garau, A. Brogni, and D. Friedman. Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence*, 15(5), 2006. 2, 4
- [59] M. Slater and A. Steed. A virtual presence counter. *Presence*, 9(5):413–434, 2000. 1, 2, 4, 5
- [60] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3, 1994. 2
- [61] I. Soares, R. B. Sousa, M. Petry, and A. P. Moreira. Accuracy and repeatability tests on hololens 2 and htc vive. *Multimodal Technologies and Interaction*, 5(8):47, 2021. 4
- [62] A. Steed, A. Brogni, and V. Vinayagamoorthy. Breaks in presence as usability criteria. In *Proceedings of HCI international*, 2005. 1, 2, 4
- [63] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *IEEE Virtual Reality*, 2016. 1, 3
- [64] B. Stevens and J. Jerrams-Smith. The sense of object-presence with projection-augmented models. In *International Workshop on Haptic Human-Computer Interaction*, 2000. 2
- [65] B. Stevens, J. Jerrams-Smith, D. Heathcote, and D. Callear. Putting the virtual into reality: Assessing object-presence with projection-augmented models. *Presence: Teleoperators & Virtual Environments*, 11, 2002. 2
- [66] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988. 1, 2, 3
- [67] K. Szczurowski and M. Smith. Measuring presence: hypothetical quantitative framework. In *23rd International Conference on Virtual System & Multimedia*, 2017. 2
- [68] A. Toet, T. Mioch, S. Gunkel, O. Niamut, and J. B. van Erp. Assessment of presence in augmented and mixed reality. 2021. 2
- [69] T. Q. Tran, T. Langlotz, and H. Regenbrecht. A survey on measuring presence in mixed reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–38, 2024. 2
- [70] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence*, 9, 2000. 2
- [71] R. Vallat. Pingouin: Statistics in Python. *Journal of Open Source Software*, 3(31), 2018. 6
- [72] F. Vona, M. Stern, N. Ashrafi, T. Kojić, S. Hinzmann, D. Grieshammer, and J.-N. Voigt-Antons. Investigating the impact of virtual element misalignment in collaborative augmented reality experiences. In *16th International Conference on Quality of Multimedia Experience*, 2024. 2, 3, 5
- [73] I. Wagner, W. Broll, G. Jacucci, K. Kuutii, R. McCall, A. Morrison, D. Schmalstieg, and J.-J. Terrin. On the role of presence in mixed reality. *Presence*, 18, 2009. 2
- [74] F. Westermeyer, L. Brübach, M. E. Latoschik, and C. Wienrich. Exploring plausibility and presence in mixed reality experiences. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2680–2689, 2023. 1, 5
- [75] C. Wienrich, P. Komma, S. Vogt, and M. E. Latoschik. Spatial presence in mixed realities—considerations about the concept, measures, design, and experiments. *Frontiers in Virtual Reality*, 2:694315, 2021. 2
- [76] M. Wiesing, G. Fink, and R. Weidner. Accuracy and precision of stimulus timing and reaction times with unreal engine and steamvr. *PLoS ONE*, 15, 2020. 1, 2
- [77] W. Wirth, T. Hartmann, S. Böcking, P. Vorderer, C. Klimmt, H. Schramm, T. Saari, J. Laarni, N. Ravaja, F. R. Gouveia, et al. A process model of the formation of spatial presence experiences. *Media psychology*, 9(3):493–525, 2007. 1, 2
- [78] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators & Virtual Environments*, 7, 1998. 2