Investigating the Correlation Between Presence and Reaction Time in Mixed Reality

Yasra Chandio (iD), Noman Bashir (iD), Victoria Interrante (iD), and Fatima M. Anwar (iD)

Abstract—Measuring presence is critical to improving user involvement and performance in Mixed Reality (MR). *Presence*, a crucial aspect of MR, is traditionally gauged using subjective questionnaires, leading to a lack of time-varying responses and susceptibility to user bias. Inspired by the existing literature on the relationship between presence and human performance, the proposed methodology systematically measures a user's reaction time to a visual stimulus as they interact within a manipulated MR environment. We explore the user reaction time as a quantity that can be easily measured using the systemic tools available in modern MR devices. We conducted an exploratory study (N=40) with two experiments designed to alter the users' sense of presence by manipulating *place illusion* and *plausibility illusion*. We found a significant correlation between presence scores and reaction times with a correlation coefficient -0.65, suggesting that users with a higher sense of presence responded more swiftly to stimuli. We develop a model that estimates a user's presence level using the reaction time values with high accuracy of up to 80%. While our study suggests that reaction time can be used as a measure of presence, further investigation is needed to improve the accuracy of the model.

Index Terms—Mixed reality, Presence

---◆---

1 INTRODUCTION

Mixed Reality (MR) is gaining importance in science, education, training, and entertainment, offering new ways of interaction and engagement with the real and virtual worlds. The technological advancements in MR tools have facilitated an enhanced sense of *presence*, allowing users to behave within an MR environment as they would in the real world. *Presence* is typically described as the subjective experience of being in a simulated place or environment and the user's readiness to respond to virtually generated sensory data as if they were real [19, 21, 63]. This includes interacting naturally and appropriately with virtually generated sensory data. Just as in the real world, an individual should be able to bend down, grab an object on the floor in a virtual environment, feel its weight, and lift it if desired. This is achieved through the sense of one's body movement and position, which matches the sensory data presented in the virtual environment. High presence does not necessarily require high fidelity to physical reality but rather that individuals can behave as if the sensory data they are experiencing is real. This approach to measuring presence allows for observing and evaluating an individual's behavior in real and virtual environments. A high *presence* is desired in any simulated virtual environment, as it allows the user to engage in an immersive, realistic, and involved experience. Many studies investigate the notion of *presence* and describe the factors contributing to the sense of *presence*. According to Witmer & Singer [92], Slater & Steed [75] and others [19, 21, 61, 63], there are two main aspects of *presence*: *place illusion* and *plausibility illusion*. *Place illusion* refers to the *sense of being there*. In an MR environment, this corresponds to how the virtual content appears indistinguishable from the real world. *Plausibility illusion* refers to the sensation that the observed events in a virtual environment occur. Users will feel involved in the environment when both *place illusion* and *plausibility illusion* occur. This involvement leads to users responding realistically to the environment, resulting in a greater sense of engagement in an MR environment. Immersion and participation are necessary to experience *presence* [92].

A prerequisite to improving the *presence* of a user is the ability to measure and quantify *presence*. While conventional measures of presence have been defined for virtual environments that surround and isolate a user from the real world [84], we are measuring presence as the subjective experience that a particular object exists in a user's environment, even when that object does not [83]. Due to the subjective nature of *presence*, the most popular method to measure *presence* is the use of subjective questionnaires [28, 38, 64, 65, 92]. Questionnaires ask users to self-report their sense of *presence* by answering questions that attempt to assess *presence*, usually after the user has left the virtual environment. The subjective questionnaires can be quickly administered, graded, and interpreted without affecting the user experience. However, questionnaires cannot measure the time-varying qualities of presence and can produce unstable, inconsistent, and irreproducible responses due to the prior experience of the participants [22]. The well-known shortcomings of presence questionnaires have led researchers to explore alternative approaches to assessing presence, including behavioral responses such as postural response [24], hand and eye response [89], and startle response [91], which are produced automatically, without conscious thought, thus avoiding user bias. However, the assessment of behavioral responses is susceptible to experimenter bias and is highly sensitive to environmental factors and content [37, 76]. Various physiological responses can be measured to assess *presence*, such as a change in heart rate [27, 42], skin conductance [49], and body temperature [36]. However, physiological measures can also be noisy and unreliable, especially under non-stationary conditions, and may not capture differences in presence in situations of low emotional valence [49].

Given the state-of-the-art, there is a need for an approach to quantify *presence* that is objective, quantitative, not consciously affected by the participant and/or experimenter, and could be used at runtime without interfering with the virtual experience. It should take advantage of existing interactions (and underline quantities) in the virtual scene and measure *presence* through these interactions rather than making additional external interventions. One such underlying quantity is *reaction time*. *Reaction time* or response time refers to the time taken between when humans perceive something and when humans respond to it. *Reaction time* is dictated by the cognitive ability to detect, process, and respond to a stimulus [18, 90], and can be easily measured using the systemic tools available in modern MR devices such as Microsoft HoloLens 2 [4].

Our work investigates a fundamental question in MR: Would an individual experiencing more *presence* systematically show faster *reaction*

- *Yasra Chandio is with University of Massachusetts Amherst. E-mail: ychandio@umass.edu*
- *Noman Bashir is with University of Massachusetts Amherst. E-mail: nbashir@umass.edu.*
- *Victoria Interrante is with University of Minnesota Twin Cities. E-mail: interran@umn.edu.*
- *Fatima M. Anwar is with University of Massachusetts Amherst. E-mail: fanwar@umass.edu.*

times? If the answer is yes, we could use a systemic metric such as *reaction time* to quantify *presence* in a non-intrusive, objective, and unbiased manner. There is a large body of work investigating the relationship between *presence* and human performance [10, 16, 39, 47, 48, 53]. Natalia et al. showed a negative correlation between task completion time and *presence* when the sense of *presence* is altered by multisensory feedback [16]. Matteo et al. show a negative correlation between performance and *presence* when the sense of *presence* is changed by varying the perceptual load [48]. Maneuvrier et al. showed that presence promoted spatial cognition performance and that the presence-performance relationship was not mediated by other human factors [47]. Furthermore, human performance is often used as an argument for the good predictive validity of questionnaires [28].

To understand the relationship between *presence* and *reaction time*, we conducted a study in which we varied the *presence* of users by manipulating *place illusion* and *plausibility illusion* while they were interacting with an MR environment. We designed two sets of experiments. In one set, we only manipulated the appearance of the virtual object, and in the other set, we manipulated a non-task-relevant behavior of the virtual object. All other aspects of the experiments, such as the interaction mechanism, frequency of interactions, and physical environments, were kept the same. We systematically measured the *reaction time* of users in response to a visual stimulus. Our post-experience questionnaires show a significant change in presence in each experiment between the manipulation conditions. Similarly, we observed a significant change in user *reaction time* as the sense of *presence* changed. Our analysis shows a correlation between *presence* and *reaction time*.

In our attempt to understand the relationship between *presence* and *reaction time*, this work makes the following contributions.

Contribution 1: We propose the use of *reaction time* of a user as a measure of *presence*. We also develop a non-intrusive, systemic approach to measuring the user *reaction time* that relies on existing interactions in MR environments.

Contribution 2: We devise experiments that alter the sense of *presence* of a user by manipulating *place illusion* and *plausibility-illusion*. In designing experiments, we control for other factors that are known to impact user performance. While we use the experiments to demonstrate change in presence, we also demonstrate, as a byproduct, that presence questionnaires typically used in fully virtual environments can also be used in MR environments.

Contribution 3: We conduct an exploratory lab study ($N = 40$) that demonstrates a negative correlation between the sense of *presence* and the user *reaction time* when responding to a visual stimulus.

Contribution 4: We develop a model that estimates a user's presence level using the reaction time values as input. Our evaluations demonstrate that model has high accuracy (up to 80%), which can be further improved with data from a larger number of users.

2 BACKGROUND AND RELATED WORK

In this section, we define the relevant terminology, discuss the existing work on the concept of presence in MR, and present existing methods for quantifying presence.

2.1 Terminology

Different forms of *reality* depend on how much of the physical world is part of the user's experience and how the user interacts with the virtual objects in the scene, as shown in Figure 1. Defining MR remains challenging, with no universally agreed-upon comprehensive definition [80]. Virtual Reality (VR) immerses users in a wholly digital realm, while Augmented Reality (AR) superimposes digital elements into our real world. Augmented Virtuality (AV) is an immersive experience, complete or partial, with added elements of 'reality' such as video or texture mapping. MR is an umbrella term that encompasses both AR and AV. Our MR concept leans towards AR on Milgram's reality-virtuality spectrum, where users interact primarily with virtual objects while being able to see the real world around them. MR represents a spatial alignment between the real and virtual worlds, allowing users to
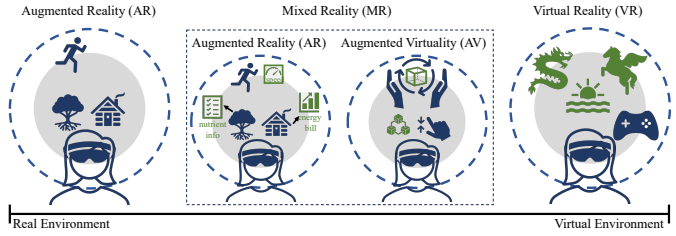

Fig. 1: *The Reality-Virtuality Continuum [52]. Virtual objects are colored green, and real-world objects are colored blue.*

interact with and manipulate both real and virtual environments. We will use MR as a blanket term throughout the paper [52].

2.2 Factors Affecting Presence

Presence is a phenomenon of awareness based on the interaction between sensory stimulation, environmental factors that encourage involvement and allow immersion, and internal tendencies to become involved and interact with virtual objects [92]. Sheridan [72] laid the foundation for determining the underlying presence factors, such as sensory information, sensor control, and motor control. Slater and Wilbur [73, 78] expanded on Sheridan's work to determine major factors affecting user presence. According to them, two main factors contribute the most to user presence. (1) *Place illusion* refers to the appearance of a virtual environment (or virtual object in the case of AR/MR). It can be affected by the realism of the virtual content, the consistency of the view between the headset and the direct view from the users' eyes (including displacement or latency issues), the lack of haptic feedback, and the awareness of the headset. (2) *Plausibility illusion* refers to the behavior of a virtual environment (or virtual object in the case of AR/MR). It can be affected by scene elements not obeying the laws of physics, cause-effect relationships not coupled as expected, actions that do not have the expected outcome, and events in the virtual environment not conforming to familiar expectations. It is argued that when both *place illusion* and *plausibility illusion* occur, users will feel involved and respond realistically to the environment, which will lead users to experience greater engagement in a virtual environment [92].

2.3 Presence Measures

Previous research highlights the importance of presence as an outcome of virtual environments (AR/VR/MR) [19, 63, 72, 92]. The most commonly accepted method of evaluating *presence* is the self-report questionnaire [87]. Typically, presence questionnaires are used after participants engage in an AR/VR/MR environment, making them post-experience questionnaires. Over time, previous research has developed, refined, and validated questionnaires for measuring presence. Bystorm et al. [14] proposed an integrated theoretical framework for studying presence that includes aspects related to task performance. In their work, they used two questionnaires to measure interaction fidelity.

The Witmer & Singer Presence Questionnaire (PQ) is one of the most widely used [92] to assess involvement/control, naturalness, and interface quality. Other popular post-experience questionnaires include the Igroup Presence Questionnaire (IPQ) [64, 65] and the Slater-Usoh-Steed Questionnaire (SUS) [75]. The PQ questions are based on the following factors: the ability to control the environment and "naturalness" of control over the environment, coherence and consistency of information from different senses, and distractions a participant may experience in a virtual environment. They also consider environment realism and meaningfulness, as well as the sense of disorientation when returning to the real world. On the other hand, the SUS and IPQ questions are based on three factors: the sense of physically being in the virtual environment, the extent to which the virtual environment feels real, and the extent to which the participant feels involved. There are additional questionnaires that measure various aspects of realism relating to the virtual scene, such as the ITC-Sense of Presence Inventory [44], the Kim and Biocca questionnaire [40], the Object presence questionnaire [84], the reality judgment and presence questionnaire [8], the Swedish viewer-user presence questionnaire (SVUP) [43], and others [10, 15, 56]. However, these questionnaires do not aim to measure

the time-varying qualities of presence. They can produce unstable, inconsistent, and unreproducible responses, and are susceptible to user bias. To address the issues of temporal information issue, Schwind et al. [66] used integrated questionnaires to measure presence. Integrating questionnaires directly into the virtual experience has been explored to assess the virtual experience as the user is going through it [62, 66, 68]. However, the problem of defining a clear baseline remains.

Other measures are continuous assessments [34, 61], psychophysical measures like cross-modality matching, free-modulus magnitude estimation, and paired comparisons. Subjective qualitative methods include autoconfrontation [60], focus group exploration [23], interaction analysis [79], free format self-reports [86], and repertory grid technique [81]. There are also several subjective corroborative measures to evaluate presence indirectly, such as Break In Presence (BIP) [13], duration estimations by users [33], attention awareness [17], and simulator sickness questionnaire [38]. Most of these methods ask participants to complete questionnaires that often contribute to the phenomenon of break-in presence [13] and are prone to participants' bias.

Objective measures are mostly captured by evaluating the behavioral and physiological responses of the users and are often used as corroborative measures. Popular physiological measures are cardiovascular measures [42], skin measures [49], ocular measures [41], and facial electromyography [32]. Presence is also known to be measured through neural correlates like Electroencephalogram (EEG) and Functional Magnetic Resonance Imaging (fMRI) [31]. Physiological measures can reduce user bias but may also be prone to inaccuracy and unreliability [49]. Some commonly used behavioral measures are based on assessing facial expression [76], postural [24], startle [91], reflex and social responses, and pointing conflicting cues [37].

2.4 Presence and Performance

Many studies have discussed the association of presence in the context of task performance in a virtual environment [10–12, 35, 45, 74, 75], we discuss only a few of them here. Slater et al. [74] conducted experiments to assess the influence of presence on performance while the participants learned to play three-dimensional chess. They noted that presence refers to the behavioral and psychological responses of people. Similarly, Barfield et al. [10] suggested that task performance measures can be used as objective corroborative indicators of presence. A few such methods are task completion time and error rate [11], the number of actions [74], and secondary task performance [35]. Though it is generally assumed that higher levels of presence are associated with better task performance [12], the exact causal link between presence and task performance is unclear.

Slater et al. [74] explored the relationship between presence and performance. While keeping other factors such as relevant background knowledge and users' ability the same, results suggested that increasing presence by increasing the richness of the virtual environment improved task performance. The study also found that reported presence was higher for egocentric than exocentric immersion, but a causal relationship between presence and task performance was not established [78]. It is also noted that motor behavior is strongly influenced by perceptual uncertainty and the expected consequences of actions [26] that can affect user's characteristics, such as ability and motivation, that will influence task performance [30]. IJsselsteijn et al. [35] noted that it is reasonable to assume that several characteristics of a virtual environment will similarly influence presence and task performance. They further expanded that performance on a secondary task can serve as a measure of the amount of effort and attention allocated to the primary task. The more effort is dedicated to the primary task, the more performance on the secondary task will decrease. A similar argument can be made in the case of presence: if more attention is allocated to the mediated virtual environment, performance on a secondary task will decrease. Szczurowski and Smith made it reasonable to assume that the nature of presence is subconscious [85]. They argue that if the presence exists outside the subjective feeling domain, it's unlikely to be a conscious process. No one must remind themselves about staying present in the real world. It's also unlikely that it is possible to force yourself into feeling present in a virtual environment. IJsselsteijn, Szczurowski, Smith, and others [9, 55, 85, 90] concluded that reaction

time or error rate could be used as task performance measures for presence evaluation.

Therefore, we make a case for measuring subjective feelings of presence with an objective, reproducible approach and producing a stable response without interference in the virtual scene. Despite being aware of problems associated with a subjective measurement method, we are deciding to use questionnaires as the baseline for our current study to test our hypotheses about relations between different constituents of *presence* and users' *reaction time*.

2.5 Presence in Less-immersive AR/MR Environments

Presence governs aspects of a user's autonomic responses and behavior in a virtual environment, whereas immersion refers to a quantifiable description of a technology [77]. Wilbur and Slater argue that the degree of immersion can be objectively assessed as the characteristics of a technology [78], and has dimensions such as the extent to which a display system can deliver an inclusive, extensive, surrounding and vivid illusion of a virtual environment to a user [82]. AR/MR elicits a different sense of presence: "It is here" presence [46]. Although AR/VR/MR is very different from a technical perspective, a common feature they share is that virtual objects exist in a curated environment: real (in the case of AR and MR) or virtual (in the case of VR). Therefore, the common approach to measuring presence in various virtual environments, from the least immersive environment (AR) to the fully immersive environment (VR), questions whether one has a sense of being in or interacting with the virtual environment. Although there have been attempts to develop presence methods [25, 57, 84] and measurement tools [59, 83] exclusively for AR/MR, these tools only measure factors that may influence presence, rather than directly measuring the subjective sense of presence itself [88].

While conventionally, presence has been defined for virtual environments that surround and isolate a user from the real world [84], Slater et al. [87] conducted a study on VR questionnaires (PQ and SUS) and concluded that the questionnaires that are developed for VR can still be useful when all users experience the same type of environment even if the environment is not fully immersive (AR/MR). They also concluded that the utility of questionnaires might be doubtful for comparing experiences across environments - such as immersive virtual compared to real, or desktop compared to immersive virtual, or a real environment with virtual objects to a fully virtual environment. Presence questionnaires are often utilized in research to explore the subjective experience of presence rather than the link between perceived presence and aspects of technology; therefore, they can be employed anywhere on the virtuality continuum in technological or real-world contexts. To this end, in this study, we refer to *presence* as the subjective experience that a particular object exists in a user's environment, even when that object does not [83]. This definition is more appropriate for assessing non-immersive displays such as AR/MR headsets [57]. We use this definition of *presence* for the rest of the paper.

2.6 Taxonomy of a virtual scene

As described by Wilbur and Slater [78], the place illusion (appearance realism) and the plausibility illusion (behavioral realism) are the main aspects of any virtual experience in any alternative reality medium.

Virtual scene represents the semantics of the virtual environment in three dimensions (3D) placed within the real environment.

Event An event happens in a computer system. For instance, adding or removing a virtual object from a virtual scene is considered an event.

Task refers to an observable activity with a start and an endpoint. In MR, tasks will be aligned with the start or end of an event, depending on the semantics of the virtual scene.

Interaction is defined as performing a physical action to perform the task in a virtual scene.

Cue is defined as the signal (visual or auditory) that is sent to the participant to initiate a task.

Feedback is the visual or auditory confirmation sent to the participant that the task is completed.

To create a virtual scene with context, immersion, and interaction, we need to craft our experiments so that participants feel engaged in the virtual world. However, the scene should neither be too complex

that a participant's cognitive load is consumed in understanding the scene nor should the scene be too simple that the participant feels disengaged [35, 77]. For example, we do not want a scene with Warcraft heavy-load games or a simple box with no semantic value to the participant. We need to create a balance between a semantically too complex and a simple virtual scene. The same applies to events, tasks, cues, interactions, and feedback from the virtual scene.

3 APPROACH

To validate *reaction time* as a measure of *presence*, we investigate the *correlation between presence and reaction time* in MR. We reference Insko et al.'s [61] criteria for a useful measure: *sensitivity* (to detect different levels of 'presence'), *reliability* (providing repeatable results), *validity* (correlating with existing 'presence' measures), and *objectivity* (free from participant's and experimenter's bias). Accordingly, we navigate the following design challenges (DCs):

DC1: How can we induce different feelings of *presence*? (sensitivity)
DC2: How do we *minimize confounding variables* while varying feelings of *presence*? (reliability)
DC3: How do we *establish a baseline* measure of presence and what should that baseline be? (validity)
DC4: What user *interaction mechanism* should we use to assess reaction times? (objectivity)

3.1 DC1 - Varying presence

We need an empirical setup that can measure presence in real-time while depicting a practical scenario for measuring varying feelings of presence. But first, we need to understand the main aspects of presence and what dimensions could be measured in those distinct but overlapping aspects. We describe two main aspects (place illusion and plausibility illusion) of presence in §2. Since we are using Mixed Reality (MR) as our experiment medium, place illusion and plausibility illusion need to be refined. In MR, measuring the place illusion could mean to what extent the virtual object appears indistinguishable from reality. The sense of plausibility can be described when users select a dominant space as the reference frame. Then virtual objects in real space or real objects in a virtual space would be perceived as plausible if the object behaves coherently to the dominantly perceived space as noted in place illusion. For example, plausibility would be lessened if gravitational forces were applied horizontally rather than vertically.

In summary, place illusion refers to the elements related to the *appearance* of the environment. In contrast, plausibility illusion refers to the elements related to the *behavior* of the objects in the environment. Therefore, we suggest that to vary the feelings of presence, we could manipulate *appearance* and manipulate the *behavior* of the object in the scene. As Slater describes, presence is affected by both realism and plausibility. To establish the relationship between response time and presence, irrespective of why the presence changes, we altered realism and plausibility to induce various levels of presence. Other factors can also impact presence, and we plan to explore their effect in the future.

3.2 DC2 - Controlling confounding variables

Various options exist for manipulating objects' appearance and behavior in the virtual environment. We must pick scenes and manipulations carefully to satisfy the reliability criteria. We have identified two constraints that will help limit the introduction of external variables and maintain symmetry across experiments and users.

- **Constraint 1:** Manipulations should not affect (increase or decrease) the overall complexity of the scene, including tasks, events, and interactions. (**simplicity** in scene)
- **Constraint 2:** Manipulations should also be free from additional confounding variables. Confounding variables are the extra variables that affect the actual relationship between the variables under study [54]. These output variables are *presence* and *reaction time*. (**symmetry** in the scene and users)

To satisfy the simplicity constraint, we will avoid unnecessary details in the scene (features mentioned above), and we are enforcing consistency in the non-manipulated conditions of our experiment. The selection of simple scenes is a conscious design choice to control the

effect of confounding variables. In a complex scene, it can be hard to isolate the effect of various factors on presence. These experimental conditions are the placement and duration of the visual stimulus, time and appearance of the cue, time and appearance of the feedback, number of tasks, duration of the experiment, and type of interaction. We also suggest that to keep the scene simple, we should pick a scene familiar to most college students (intended participants pool) and have only one virtual object in the scene at a time. However, we want to keep the participant engaged throughout the experiment, and we need a scene with some familiar semantics. Similarly, the manipulations that will vary the presence should be subtle and change only one scene to isolate the effect on our output variables. In [67], authors have proposed manipulating the scene's appearance by manipulating the visual fidelity. We vary presence by making the scene appear *realistic* in one scenario and *abstract* in another. Similarly, as shown in [13], plausibility illusion can be affected by challenging the physical laws in a scene. In behavior manipulation, we make the behavior of a scene naturally plausible in one scenario and implausible in another. We will maintain a controlled physical environment to minimize interference and satisfy symmetry constraints. We keep the same room, lighting, seating arrangement, study session duration, breaks, and order of the experiment across users and experiments.

To this end, we divide our experiment into two sets. In the first set, we vary the sense of presence by manipulating the *appearance* of the virtual object (place illusion). In the second set, we vary the sense of presence by manipulating the *behavior* of the virtual object (plausibility illusion). Each experiment will contain two blocks of trials: a *control* block and a *manipulated* block. In the *control* block, the scene contains virtual objects with a natural appearance and plausible behavior. However, in the *manipulated* block, the scene contains virtual objects whose appearance is manipulated to be visually unnatural (unrealistic) or in which the virtual objects exhibit implausible behavior.

3.3 DC3 - Establishing a baseline

Since presence itself has a subjective nature, it is logical that we also establish our baseline with a subjective measure. To understand the level of presence, we use the questionnaire as our baseline, which researchers commonly accept as the standard measure. This helps us set a starting point for our study [13, 58, 67]. Subjective questionnaires have been the standard measure of presence for many years. They are sensitive enough to find differences in presence when used to examine the difference between two visually similar fidelities [61]. The post-experience questionnaire provides scores that reflect the level of perceived presence in the scene. While subjective questionnaires have limitations, as discussed in §2.3, they are currently the only widely accepted method of quantifying presence. We measure various feelings of presence with the three most-cited and widely used questionnaires that measure presence. We use the Igroup Presence Questionnaire (IPQ) [64, 65], Slater-Usoh-Steed (SUS) [75], and Witmer & Singer (PQ) [92].

3.4 DC4 - Selecting an interaction mechanism

We need a task that engages the participant and helps us measure our output variable, *reaction time*. Through interactions, the participant can physically be involved in the scene. In designing the interaction, we want to avoid imposing differential barriers to task completion, such as placing a button at a height that is easier to reach for some participants than others. Similarly, we want to avoid designing unnecessarily complex interactions that could cause task performance to vary unpredictably, independent of the targeted manipulations. Additionally, we ensure that external interventions do not affect the reaction time. Therefore, we need to leverage existing elements in the taxonomy of the scene (cue, interactions, and feedback).

To solve these challenges, we employ HoloLens's "air tap" gesture [1] as our interaction mechanism. It is precise, and its ease of use is independent of a typical participant's height, reaching range, or any other physical attribute. The air tap gesture requires specific movements performed in a particular order. According to the instructions for air tap in the HoloLens 2 manual [50], the user needs to "hold your hand straight out in front of you in a loose fist, point your index finger
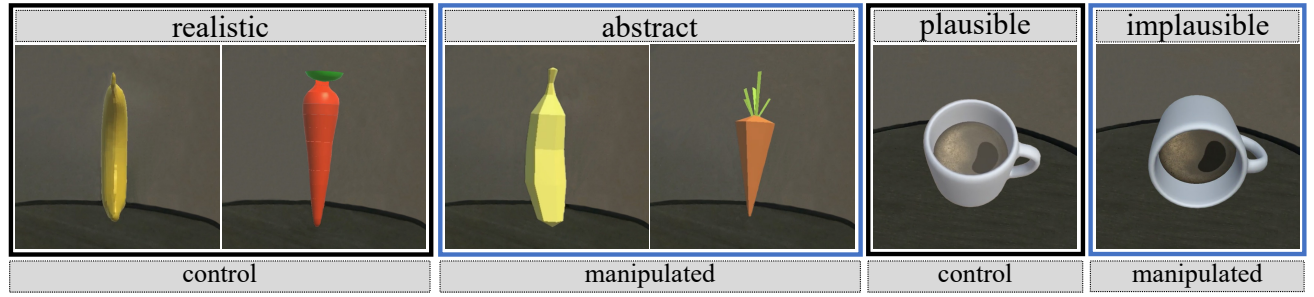
Fig. 2: *Realistic, abstract, plausible, and implausible virtual objects used in the experiments.*

straight up toward the ceiling, tap your finger down, and then quickly raise it back up again." This specific sequence of movements reduces the probability of mistriggering due to random motions. Additionally, in the scene depicted in Figure 3, there is no movement other than the participant performing an air tap in response to a cue. Therefore, the closed position of the participant's fingers in the image is not an error and does not pose a potential for inaccurate gesture recognition.

We designed a task that supports participants in knowing what to do (cue), knowing that the system is working (interaction/air tap), and knowing if their action was understood by the system (feedback). In the *realistic* vs. *abstract* scenario, the appearance of the object in the scene is the cue for the user to take action (air tap), and the disappearance of the object from the scene is the feedback to the user that their action was successful. In the *plausible* vs. *implausible* scenario, the cessation of change in the height of the coffee in the cup cues the user to initiate their action. The coffee cup changes color to provide feedback to the user about the success of their action (details in §4.6).

Next, we formally define the reaction time ($reaction_i$) in our study for the trial i as ($interaction_i - cue_i$). where trial i refers to one task iteration by the participant (cue, air tap, feedback). $interaction_i$ is the time when the air tap is recorded in trial i, and cue_i is the time of onset of the task cue in trial i.

4 USER STUDY

In this section, we detail the study measures, participants, equipment, and procedure for the user study.

4.1 Participants

We recruited 40 participants (23 male-identifying, 16 female-identifying, and one non-binary identifying) with a mean age of 26.6 years (standard deviation of 5.5 years). All participants volunteered and provided written informed consent. They received $25 for their participation. All but one participant had a technical background in computer science or engineering. All the participants had normal or corrected normal vision with contact lenses or glasses. Twelve participants had 1 to 4 days per week of MR experience, 24 participants had less than 1 hour per week of experience, and nine had never experienced MR before. Only 12 participants had used Hololens 2 before the study. The study was granted ethics clearance according to the ethics and privacy regulations of our Institutional Review Board (IRB).

4.2 Material

The study utilized an ergonomic, untethered, self-contained holographic device, Hololens 2 [4] equipped with a second-generation Holographic Processing Unit (HPU) for real-time computer vision and a Qualcomm Snapdragon 850 CPU for running applications. The virtual scenes were developed via Unity 3D (10.0.19362.0) game development engine with API for the Universal Windows Platform on Windows 10 PC.

The Hololens 2 accepts eye, spatial, and hand-tracking inputs with a field of view of $43°$ horizontal, $29°$ vertical, and $52°$ diagonal. Its dual see-through displays have a resolution of 1440×936 pixels each, a 60Hz refresh rate, and a tinted visor to minimize environmental light interference. We chose the HoloLens 2 for our experiment because its see-through setup allows the user a direct view of the real world. The other MR headsets that leverage video to show the physical world cannot be used for safety-critical applications like surgery.
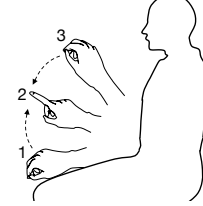

Fig. 3: *Illustration of the experimental task using air tap.*

4.3 Variables

In our study design, we change two variables to test our hypothesis, but we only manipulate one variable at a time. We chose *appearance* and *behavior* of virtual objects as our variables in the first and second experiment sets, respectively.

Experiment Set 1: Realistic vs. Abstract

In the *control* trials, all virtual objects in the scene will have a realistic appearance (textured and natural) and plausible behavior. In the rest of the paper, we refer to this block of trials as *realistic*. In the *manipulated* trials, all virtual objects will depict plausible natural behavior, but the appearance would be abstract (untextured and geometric). We refer to this block of trials as *abstract* in the remainder of the paper.

In the control trials, all of the virtual objects in the scene have a realistic appearance (textured and natural) and plausible behavior. In the rest of the paper, we refer to this block of trials as *realistic*. In the manipulated trials, all virtual objects depict plausible natural behavior, but their appearance is abstract (untextured and geometric). We refer to this block of trials as *abstract* in the remainder of the paper. For the *realistic* vs. *abstract* trials, we chose a scene that contains textured virtual objects. To satisfy our simplicity constraint, we modified the popular Fruit Ninja game [3]. To mimic the semantics of a multi-object environment, we use two objects that are similar in shape and size but different in textural properties. As we wanted a simple scene and one virtual object at a time, we made the virtual objects appear in the scene one after the other. Ultimately, we used one fruit (banana) and one vegetable (carrot) for this experiment. We removed complex interactions (slicing) and additional cognitive loads (scores) from the original game. In the *realistic* condition, we make the carrots and bananas appear as natural as possible in terms of color, texture, and shape. In the *abstract* condition, we render the carrots and bananas with muted colors, without texture, and with geometrically blocky shapes, as shown in Figure 2.

Experiment Set 2: Plausible vs. Implausible

In the *control* condition, the virtual objects in the scene have a natural appearance and plausible behavior. We refer to this block of trials as *plausible* in the remainder of the paper. In the *manipulated* condition, the virtual objects retain a natural appearance but behave implausibly from real-world standards (disobeying laws of physics). In the remainder of the paper, we refer to this block of trials as *implausible*.

For the *plausible* vs. *implausible* scenario, we chose a scene with elements that have physical constraints. We modified the popular Coffee Stack game [2] with a coffee mug on a surface that periodically fills up with coffee. Like Fruit Ninja, we have removed complex interactions and cognitive loads from this game. In the *plausible* scenario, the empty
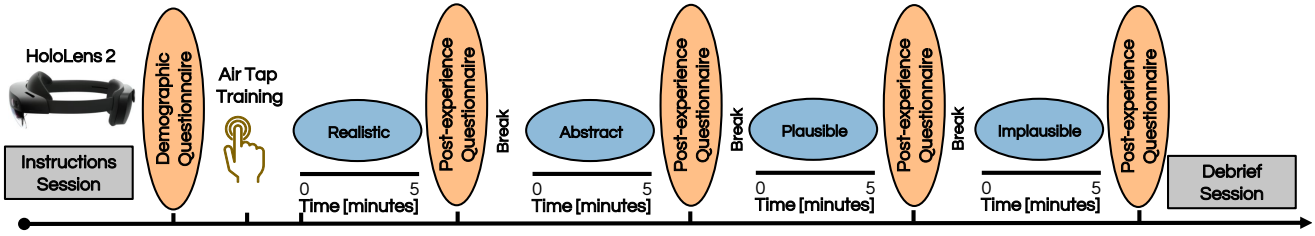
Fig. 4: *User study timeline consisting of pre-and post-questionnaire and four blocks across the two experiments.*

white porcelain coffee mug is upright, and the coffee pours into the mug periodically until the mug is full. Then, a new mug appears, and this cycle continues to repeat. In the *implausible* scenario, the empty white porcelain coffee mug is in a tilted position on the surface, and coffee continues to pour into and fill the mug even though gravity and volumetric constraints should not allow this (see Figure 2).

4.4 Experimental Task

The experimental task in both sets can be divided into three sub-tasks, as shown in Figure 3. The experiment objects are placed in the participant's field of view as a prerequisite. The first step for participants is to view the object with their hand in a neutral position without raising their elbow, as illustrated in Figure 3(1) and captured in Figure 2. The second step for the participants is to lift their index finger and react upon cue by air-tapping the virtual objects, as shown in Figure 3(2). Finally, the third step for participants is to return their hands to the neutral position after seeing the feedback. To avoid double triggers, we instructed participants to perform a single air tap in response to a cue.

4.5 Measures

Presence scores for the questionnaires were obtained using 39 items (6 SUS, 14 IPQ, 19 WS) on a 7-point scale. We did not modify any of the questions. The reaction time is recorded by our software on HoloLens 2. On average, we collected 50 reaction time measurements per block of trials, totaling 200 data points per participant for the four blocks. The reaction time is measured in milliseconds (ms).

4.6 Pilot Study

Before starting the formal study, we conducted a pilot study with two participants to tune the parameters to minimize environmental variables and maximize reliability and objectivity. We use a talk-aloud protocol and ask participants questions to align the experiment for general comfort and ease for the participants but not to bias the experiment for a specific set of users. We tuned the following session-specific parameters.

Participant position. We experimented with standing, walking, and sitting positions. The participants reported feeling tired while standing. While walking, they reported that the scene kept changing around them, necessitating extra attention to locate the virtual object. Participants reported the sitting position as the most comfortable position for the experiment. We asked the participant to sit on a chair with relaxed shoulders, an arm on the lap or armrest, and feet flat on the floor. The chair with armrest and backrest was reported to be the most comfortable with air tap interaction [1].

Room lighting. Lighting affects a user's ability to see the virtual environment and its objects [20]. It also affects the rendering of virtual objects and user interaction with those objects, as they rely on the tracking module of Hololens 2. For optimal visuals through Hololens 2, lighting should be even and sufficiently bright so that a participant can see without effort but not so bright that a participant has difficulty looking into the environment. To compensate for the darkness of the visor, dim lights reflecting in the direction of the participant's head are deemed most effective because a tinted visor may cause a loss of contrast in the physical environment.

Experiment duration. How long will an experiment run with the same repetitive task? It should be long enough for the participant to feel involved, and we can collect sufficient data on *reaction time*. It should be short enough that the participant does not feel tired or disengaged, impacting the accuracy of their interactions. We tested for a duration

of 2 to 10 minutes. At 2 minutes, the participants reported that they could not get acquainted with the environment. At 10 minutes, the participants reported feeling disinterested after a while. We also wanted to test the recovery time of the presence or reaction time (discussed in detail in 5.1.2). We picked 5 minutes per experiment as it allowed us to obtain at least 60 reaction time readings per experiment, and the participant did not feel tired.

Wait time between blocks and experiments. After each block of an experiment, the participant was asked to complete the questionnaires. We explored 0-20 minutes of wait times after filling out the questionnaire and before starting the next block. Participants reported 5-minute wait times as sufficient, but a longer break was available upon request. In the main study, we explicitly asked participants if they needed more downtime before each experiment.

Experiment and block order. For an experiment, we could expose the participant to a control block and then to manipulated trails, or vice versa. In either case, there is a risk of obtaining better performance due to greater experience. Out of an abundance of caution, we decided to run all participants with the control block first and manipulated block second so that any underlying tendency for performance to improve over time would work against our hypotheses. However, in hindsight, we recognize that counterbalancing would have been a more appropriate way to control such potential effects.

Virtual object placement. We use Fitts's law [29] to calculate the expected time of motor movement for several different positions in the scene. We placed the virtual object parallel to a sitting participant's eye level. The object was placed at a 45 centimeter distance, as recommended by Hololens 2 intractable object guidelines with air tap interaction [5]. The object's size was tested between $1.4 \times 1.4cm$ and $3.5 \times 3.5cm$. Both participants felt comfortable with $1.4 \times 1.4cm$.

Event and task period. Our experiments involve repeating trials in each block. However, repeating events too quickly can make the task more difficult and decrease accuracy, while long breaks between events can break the presence [13]. To find the appropriate interval, we tested intervals from $1 - 15$ seconds (s) but found that periods shorter than $3s$ were too short and caused confusion, while periods longer than $8s$ were boring for participants. Therefore, we settled on a $5s$ interval between cue onsets to balance user comfort and task accuracy.

Cue appearance. We initially tested using a color change as a visual cue for the participant's interaction in both sets of experiments (see details in §3.2), but it was found to be distracting. We then tried other cues and settled on using objects appearing in the scene to initiate an air tap gesture in the fruit ninja game in realistic and abstract scenarios. A glowing button prompt was tested but found distracting for the coffee mug experiment. The cue was ultimately changed to filling the coffee in the mug, which participants found more engaging.

Proximity of interaction. We tested the air tap interaction at a close and far distance. The participant struggled with the far air tap. This could be because the virtual object was placed near the user, and the far-touch interaction moved the virtual object farther from the participant and created an unnecessary distraction. The participant found the near air-tap interaction is more natural, so we kept it as the mode of interaction for all of our experiments.

Accuracy of interaction. To assess the potential for mistriggers, we conducted experiments with pilot study participants. In this pilot, we asked participants to perform random movements with their hands for 10 minutes while avoiding the air tap gesture. Throughout the pilot, our system did not register any unintended interactions by the participants, indicating a low-to-no chance of mistriggers.
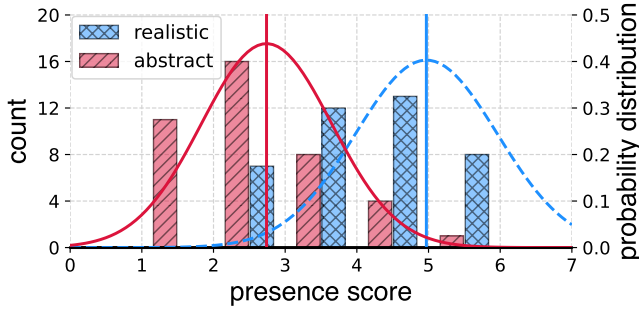
Fig. 5: *Histogram, means (left y-axis), and fitted Gaussian distribution (right y-axis) using mean questionnaire scores.*

No	realistic (μ, MAD)	abstract (μ, MAD)	t-test	p (<0.05)
ALL	4.97, 0.90	2.74, 0.74	19.44	yes
SUS	4.92, 1.09	2.60, 1.89	8.35	yes
IPQ	4.17, 1.58	2.31, 0.57	8.98	yes
PQ	5.58, 0.65	3.09, 0.99	10.03	yes
REAL	5.13, 0.54	3.04, 0.79	17.03	yes
REAL, GP, SP	4.99, 0.71	2.94, 0.59	6.63	yes
ALL - (REAL, GP, SP)	4.94, 1.40	2.94, 1.29	1.03	0.07

Table 1: *Mean, mean absolute deviation (MAD), and paired samples t-test results for all questionnaires.*

We took several steps to mitigate the possibility of erroneous gesture detection. First, before the experiment, we coached participants on performing a successful single air tap gesture. Second, we did not record any timing data for a trial where an air tap was undetected. Our data shows an average of 50 detected air taps out of 60 possible during the 5-minute interaction, as indicated in table 2 and table 4. Third, if a participant tapped twice in response to a single cue, we considered only the first measure as the response time.

Feedback. In the realistic vs. abstract experiment, we employed virtual objects disappearing as feedback for successful interaction, which participants found acceptable, similar to the original fruit ninja game. For the plausible vs. implausible experiment, we initially used a subtle coffee-poured sound as feedback, but one participant found it distracting. We then changed to using the mug color change as feedback, which was found to be more subtle and less distracting. Interestingly, participants preferred the "change in color" being used as feedback rather than as a prompt for a gesture.

4.7 Procedure

Figure 4 shows the complete timeline of the user study. Upon arrival in the study room, the experimenter welcomed the participants and provided them with verbal introductions and instructions. Participants subsequently read and signed the IRB-approved informed consent form and completed a pre-questionnaire about demographic characteristics (age, self-identified sex, and familiarity with the MR). We also briefed them on MR and HoloLens 2). We explained the study's flow, the types of experiments, the duration of each block and experiment, their order, and the questionnaires. We further clarified how to interact with the virtual objects, perform an air tap, and keep their hands in their laps when not air tapping. We explained what participants should expect while wearing the headset, how and when to interact with the virtual object, and the cues for air tapping and feedback. After the verbal training, we asked participants to wear the headset and perform five taps on average to select different menu items. The participants manually initiated HoloLens 2's calibration process, which improves visual quality and comfort for the participant. The calibration process is described in HoloLens 2 Manual [51].

As speeding could potentially override participants' natural instincts, we advised them to complete the task correctly without rushing or stalling. To avoid this instinctual bias, we deliberately kept this information (reaction time measurement) from the participant at the beginning and added this information to the debrief. Participants were also not informed of the frequency of tasks to minimize anticipation.
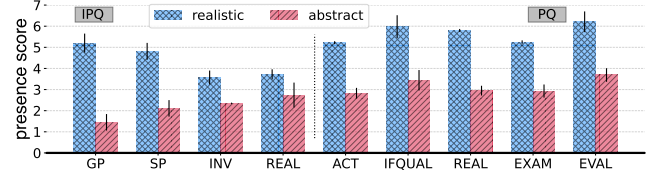

Fig. 6: *Subscale Scores. IPQ: general presence (GP), spatial presence (SP), involvement (INV), and realism (REAL); PQ: possibility to act (ACT), interface quality (IFQUAL), realism (REAL), possibility to examine (EXAM), and self-evaluation of performance (EVAL).*

Experiment	Reaction Time (ms) (μ | MAD)	Change (%)	No. of Interactions
realistic	954 | 170		50
abstract	1313 | 290	37.63	44

Table 2: *Average, MAD, %age change of user reaction times, and average number of interactions across the two sets of experiments.*

We also used the talk-aloud protocol and asked participants to think aloud as they performed the tasks to assess the varied reactions of the participants over time, which the subjective questionnaire does not capture. We also monitored the entire process through the first-person user view in Mixed Reality Capture [6]. Participants were also advised to stop at any point in the experiment if they felt uncomfortable.

We world-locked the virtual scene in the environment to ensure that the virtual objects always appeared at the same locations relative to objects in the physical environment rather than at a fixed position relative to Hololen's screen. The reason for world-locking the scene is to remove the environment variable, and the world-lock is adjusted at the participant's eye level by adjusting the participant's chair. We adjusted the headset to the participant's head and calibrated it to the participant's interpupillary distance for the best visual results. Before the first experiment, for training purposes, participants were asked to interact with the virtual object (an empty dialog box with the next button and cancel) using an air tap. The experimenter and the participant verbally confirmed the success of the first air tap.

We set the order of blocks for the participant as *realistic*, *abstract*, *plausible*, and finally *implausible*. In each block, the timer is started only when the participant is comfortable and verbally confirms readiness. At the end of the 5 minutes, the block ends by closing the scene (through the device portal [7]). We then remove the headset from the participant's head and ask them to fill out three post-experience questionnaires (see details in §3.3). After completing four blocks and the post-experience questionnaires, we thanked the participants for their participation and provided their compensation. The participants were then briefed on the study. After informing the participants about the purpose of the study, at this stage, we provided them with the option to withdraw from the study.

4.8 Hypothesis

Based on the related work, we developed the following hypothesis for our experiment sets.
H1: Manipulating the place illusion (virtual object appearance) or plausibility illusion (virtual object behavior) leads to a change in presence.
H2: Change in presence for a participant leads to a change in the participant's reaction time.
H3: Presence and reaction time are correlated.

5 RESULTS

This section presents the results quantifying *presence* using questionnaires and user *reaction time*.

5.1 Experiment Set 1: Realistic vs. Abstract

In this experiment, we changed the feelings of *presence* by manipulating the appearance of the object (i.e., *place illusion*). We get presence scores from the questionnaires and reaction time scores from HoloLens.

5.1.1 Presence Questionnaire Scores

We collect answers to all questionnaires on a 7-point scale. The *presence* score for all the questionnaires, individually or combined, is computed as the average of 7-point scores.
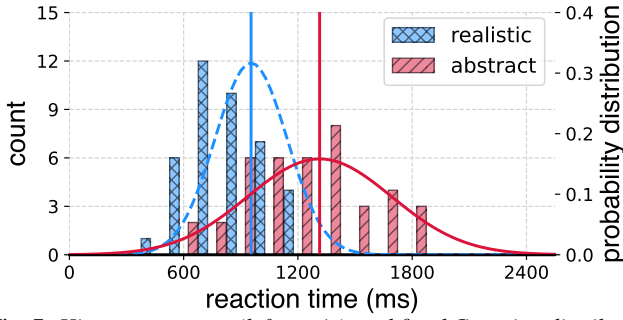
Fig. 7: *Histogram, means (left y-axis), and fitted Gaussian distribution (right y-axis) using the average reaction time across participants.*
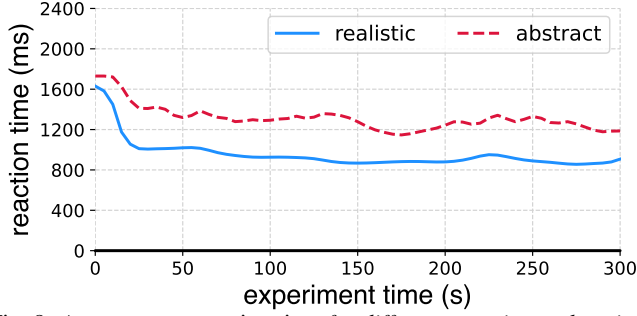

Fig. 8: *Average user reaction time for different experimental settings. User reaction time recovers (decreases) over time.*

Questionnaire scores. Our presence scores are collected from the same set of participants under two different conditions: realistic vs. abstract. We use a paired samples t-test, with a *null hypothesis* that *mean of two sets of experiments is equal*, to determine if the presence scores changed between realistic and abstract experiments. Before applying the t-test, we verified the normality of the difference between the presence scores for the two experiments. The results of this experiment are reported in Table 1. The difference in presence score between realistic experiment (M = 4.97; MAD = 0.90) and abstract experiment (M = 2.74; MAD = 0.74) was significant (t (40) = 12.85; p < 1.32e-15). Therefore, we can reject the null hypothesis and state that the presence of subjects changed across experiments. Figure 5 shows the histogram of the scores across users and its probability distribution.

Subscales. While our aggregate results demonstrate that the presence score changed as we altered the realism of the objects, we want to investigate the factors that contributed to the change in presence. The mean scores for all subscales and the aggregate realism scores across questionnaires are shown in Figure 6. First, the realism questions constitute 11 of the 39 questions and show a significant change in the presence score. This suggests that the realism significantly changed across the two experiments, also confirmed by the t-test. We also report the disaggregated presence scores for realism- and presence-related questions to analyze how factors other than realism and feeling of presence affect the overall presence score. While the null hypothesis is true, the p-value is very small, indicating that other aspects had a lesser impact but require further investigation.

User characteristic and presence. We performed a regression analysis using F-test[1] to check if age, gender, and familiarity with MR had any impact on *presence* scores. In this analysis, the null hypothesis is that a regression model based on a given variable is not a better fit than a simple intercept-only model. We observed that age, gender, and familiarity with MR did not have any effect on overall *presence* scores, as we obtained values of $F(1, 38) = 0.96, p = 0.34$, $F(1, 38) = 1.64, p = 0.96$, and $F(1, 38) = 2.19, p = 0.71$, respectively.

5.1.2 Reaction Time

To evaluate participants' reaction time, we record a time-stamp when the cue appears and a time-stamp when the user's action is recorded.

---

[1] $F(X, Y)$: X, Y are degrees of freedom between and within groups, respectively. $X = $ total groups-1, $X = $ group size$-$total groups.

No	plausible (μ, MAD)	implausible (μ, MAD)	t-test	*p* (<**0.05**)
ALL	5.17, 0.93	2.81, 0.91	8.05	yes
SUS	5.34, 1.09	2.70, 1.13	7.52	yes
IPQ	4.51, 1.35	2.63, 0.90	6.10	yes
PQ	5.61, 0.79	2.98, 1.30	7.88	yes
Plausible	4.73, 0.89	2.77, 0.57	5.23	yes
Plaus., GP, SP	4.94, 0.43	2.62, 0.77	9.05	yes
ALL - (Plaus., GP, SP)	5.36, 1.34	2.97, 1.16	4.18	0.06

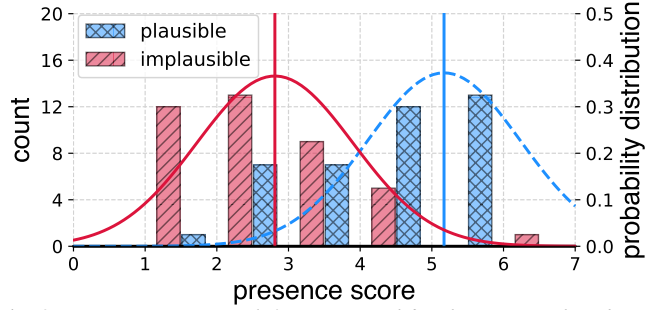Table 3: *Mean, MAD, and t-test results for all questionnaires.*

Fig. 9: *Histogram, means (left y-axis), and fitted Gaussian distribution (right y-axis) using mean questionnaire scores.*

We use the difference between these two timestamps as the reaction time. There are instances of "no-triggers" where either the participant does not respond, or the air tap does not register. We remove the corresponding cue from our observations if the air tap is not registered. As a result, our experiments recorded, on average, 50 responses out of the maximum possible 55-60, indicating that around 10-15% of the time, participants did not respond, or the air tap was not recorded. Removing this data ensures that "no-triggers" do not impact our findings. Table 2 presents the high-level results of the experiments.

Reaction time values. Figure 7 shows the distribution of average reaction time scores across the users. Similar to the presence scores, we used a t-test to evaluate the null hypothesis that the median reaction time across the two experiments is equal. Our results show a significant difference in average reaction times between the appearance conditions: 954ms in the realistic condition and 1313ms in the abstract condition with t-statistics of t(40) = 8.71 and $p < 1.09e^{-16}$. This rejects the null hypothesis and also presents a significant difference of 37.63%.

Reaction time recovery. Next, we examined the reaction time of users over time. In Figure 8, we observe that participants took significantly longer to respond to the cues at the start of the experiment. The reaction time drastically dropped and settled to a steady-state within the first 30 seconds. After that, there was only a modest improvement in the reaction time over the duration of the experiment.

Additional statistical analyses. We also measured the number of times a user was able to respond to the cue. Since the number of cues between experiments differed due to slight variations in the experiment duration, we only considered the first 60 cues to collect these statistics. We see that users could complete more interactions in the realistic appearance condition than in the abstract appearance condition.

5.1.3 Experiment Set 1: Discussion

Our results suggest that the place illusion part of our first hypothesis, "*H1: manipulating the place illusion (appearance of a virtual object) leads to change in presence*", is valid. We successfully altered the presence by manipulating the object's appearance from realistic to abstract. Prior work on manipulating the place illusion to alter the presence also supports our results [13]. Additionally, the high correlation between questionnaire scores between conditions suggests that the users who felt a greater presence in one condition also felt a lower presence in the second condition. Our second hypothesis, "*H2: Change in presence for a participant leads to change in participant's reaction time*" is also valid. Our first hypothesis confirms that the presence changed from the experiment with a realistic object to the experiment with an abstract object. Simultaneously, we observed a significant increase in the user
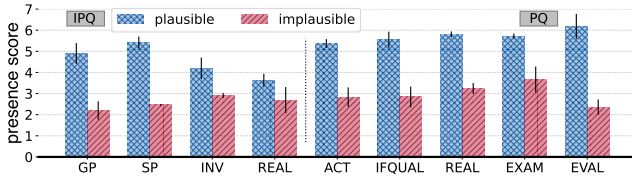
Fig. 10: *Subscale Scores. See Figure 3 caption for subscale acronyms.*

Experiment	Reaction Time (ms) ($\mu \mid MAD$)	Change (%)	No. of Interactions
plausible	930 | 240		54
implausible	1182 | 210	27.10	51

Table 4: *Average, MAD, %age change of user reaction times, and average number of interactions across the two sets of experiments.*

reaction time as participants moved from realistic to abstract object experiments. Finally, the recovery time analysis also yields interesting points. The reaction time for the manipulated experiment block is always higher than the control experiment block. Furthermore, the reaction time is steady after the initial few seconds. This means the participants initially took some time to get acquainted with the environment, leading to an increase in their reaction time. As time went on, they interacted with the virtual objects in a fairly consistent manner.

5.2 Experiment Set 2: Plausible vs. Implausible

As in the previous experiment set, we get presence scores from questionnaires and reaction times from HoloLens 2.

5.2.1 Presence Questionnaire Scores

In this experiment, we changed the feelings of *presence* by manipulating the behavior of the object (i.e., *plausibility illusion*). The experiment and analysis setup is the same as the previous experiment set.

Questionnaire scores. Our presence scores are collected from the same set of participants under two different conditions: plausible vs. implausible. We use a paired samples t-test, with a *null hypothesis* that *mean of two sets of experiments is equal*, to determine if the presence scores changed between plausible and implausible experiments. Before applying the t-test, we verified the normality of the difference between the presence scores for the two experiments. The results of this experiment are reported in Table 3. The difference in presence score between plausible experiment (M = 5.17; MAD = 0.93) and implausible experiment (M = 2.81; MAD = 0.91) was significant (t (40) = 8.05; p < 8.11e-10). Therefore, we can reject the null hypothesis and state that the presence of subjects changed across experiments. Figure 9 shows the histogram of the scores across users and its probability distribution.

Subscales. While our aggregate results demonstrate that the presence score changed as we altered the plausibility of the objects, we want to investigate the factors that contributed to the change in presence. The mean scores for all subscales and the aggregate plausibility scores across questionnaires are shown in Figure 6. First, we filtered the questions relating to plausibility, comprising 8 out of 39 questions, which show a significant change in the presence scores. This suggests that plausibility significantly changed across the two experiments, also confirmed by the t-test. We also report the disaggregated presence scores for plausible- and presence-related questions and the rest of the questions to analyze how factors other than plausibility and feeling of presence affect the overall presence score. While the null hypothesis is true, the p-value is very small, indicating other aspects had a lesser impact but require further investigation.

5.2.2 Reaction Time

We collected 50 data points per user per experiment on average after post-processing data as described in Section 5.1.2 for the first experiment. Table 4 presents the results of the experiments.

Reaction time values. Figure 11 shows the distribution of average reaction time scores across the users. Similar to the presence scores, we used a t-test to evaluate the null hypothesis that the median reaction time across the two experiments is equal. Our results show a significant
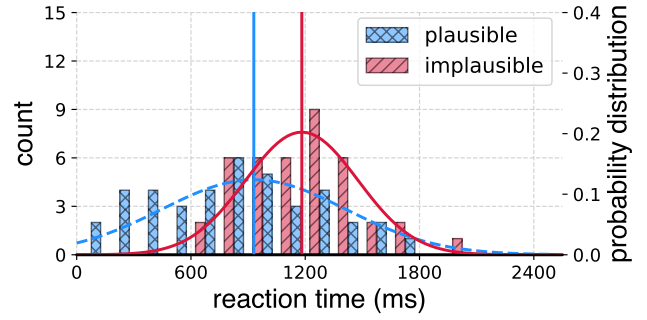

Fig. 11: *Histogram, means (left y-axis), and fitted Gaussian distribution (right y-axis) using the average reaction time across participants.*
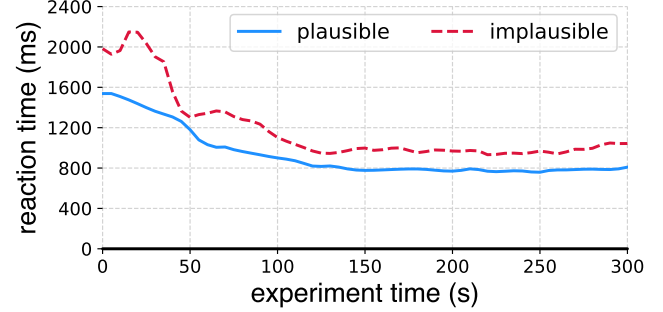

Fig. 12: *Average user reaction time for different experimental settings. User reaction time recovers (decreases) over time.*

difference in average reaction times between the two plausibility conditions: 930ms in the plausible condition and 1182ms in the abstract condition with t-statistics of t(40) = 11.67, and $p < 1.93e^{-13}$. This rejects the null hypothesis and also presents a significant difference of 27.10%. As in the first experiment set, we see decreased interactions after the manipulation.

Reaction time recovery. Next, we examined the reaction time of users over time. In Figure 12, participants took slightly longer than 60 seconds to reach a steady state for the plausible vs. implausible experiment, which was longer than the first experiment set.

5.2.3 Experiment Set 2: Discussion

This experiment aimed to establish that presence can be modified in ways other than manipulations to the place illusion done in the previous experiment. For this experiment, the plausibility illusion part of our hypothesis "*H1: manipulating the plausibility illusion (behavior of virtual object) leads to change in presence*" is valid. We altered the presence by manipulating the object's behavior from plausible to implausible. Gravity is a crucial aspect of our lives, and we expect objects to behave in specific ways under gravity. If their behavior is not plausible according to the laws of physics, it leads to degraded presence. Similar to the previous experiment, we saw a high correlation between questionnaire scores, suggesting that plausibility illusion was the factor in altering presence. The second hypothesis, H2, can be accepted based on our results. The only difference is that the change in presence was due to the change in plausibility illusion and not the place illusion. The recovery time analysis for this experiment set yields similar results as the previous experiment. However, users took longer to reach the steady state, and reaction times varied over time. This means that manipulating the appearance of the objects has less effect on participants than behavior manipulation. This is understandable as humans are more likely to see objects with non-standard appearance as opposed to observing objects that do not conform to gravity.

6 DISCUSSION

In this section, we discuss the implications of the quantitative results presented in the previous section.

Our first hypothesis, "*H1: Manipulating the place illusion (appearance of a virtual object) leads to change in the presence*" for the first experiment set can be accepted. The presence score of participants for the manipulated experiment was more than 2 points lower on a 7-point

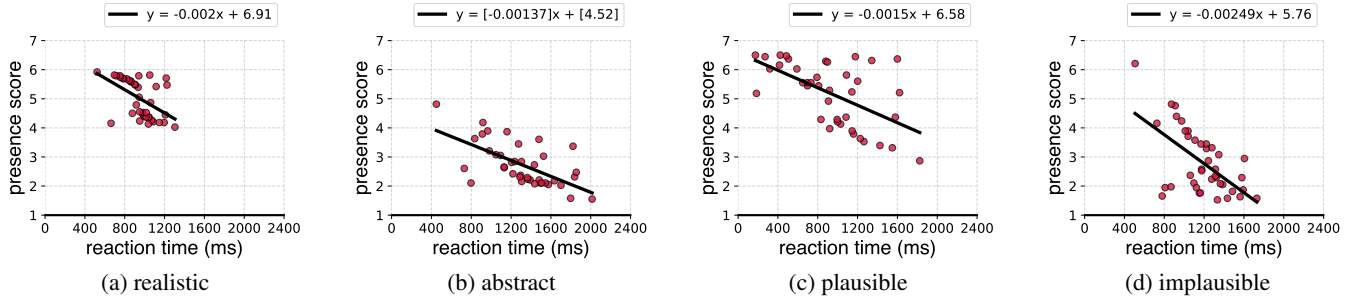| (a) realistic | (b) abstract | (c) plausible | (d) implausible |

Fig. 13: *Presence vs. Reaction Time: Presence decreases as reaction time increases. Reaction time and presence also show a modest correlation: overall (-0.65), realistic (-0.51), abstract (-0.63), plausible (-0.57), and implausible (-0.59). Each red circle represents a study participant. The black line is the linear regression fit for the data.*

scale than the control experiments. As we used everyday virtual objects, bananas in this case, a change in their texture significantly alters our strong experience-based prior. The results from the second experiment set further support this statement and result in the acceptance of H1. As all the participants experience gravity at all times, an object defying gravity is implausible and causes a break in presence. This statement is supported by prior work asserting that break-in plausibility, gravity-defying behavior, leads to a change in presence [13]. Finally, our hypothesis is acceptable for the individual questionnaires and their subscales, as the change in presence is the same as the overall results.

Our user *reaction time* results suggest that our second hypothesis "*H2: Change in presence for a participant leads to change in participant's reaction time*" can be accepted. As the users became more familiar with the environment and setup, the reaction times from one set of experiments to the other decreased. We also observed an improvement in the number of interactions across the two sets. However, this improvement could not overcome the increase in reaction times due to our manipulations within a set of experiments. Within a block, across both experiments, the reaction time drastically improved at the start but quickly reached a steady state. The difference in reaction time between the blocks of an experiment remained constant over time. This indicates that the recovery effect is consistent across blocks, and our manipulations influenced the change in reaction times of the participants.

6.1 Presence - Reaction Time Correlation

Up to this point, we have established that we were able to change the presence, and participants' reaction times changed as their presence changed. Figure 13 shows the scatter plots of participants' *presence* scores and their *reaction times* for all the experiments. We observe that *reaction time* has an inverse correlation with *presence* score, albeit with different slopes for the linear regression model across experiments. Our results highlight that appearance and behavior manipulations impacted the presence questionnaire scores and reaction time values similarly when aggregated across users. We also observed a moderate correlation between the presence scores and reaction times of individual users.

This is a significant result and suggests that *presence* impacts *reaction time*, at least within the confines of our study. This means we can accept our hypothesis that "*H3: presence and reaction times are correlated*". However, the correlation values lie in the modest range between -0.5 and -0.65. This means that further analysis is needed to fully establish that reaction time can be used to quantify presence. This is because a given change in *presence* score or *reaction time* does not elicit the same change in the other variable across all experiments. Also, the individual differences between different experiments are significant enough to warrant that the use of this relationship should be scenario-specific. Having said that, we can at least state that if the presence of a user increases or decreases, the user reaction time will also decrease or increase respectively.

6.2 Modeling Reaction Time–Presence Relationship

Our results in the previous section demonstrate a high correlation between presence and reaction time. We leverage this correlation to build a classification model that takes the user's reaction time as input
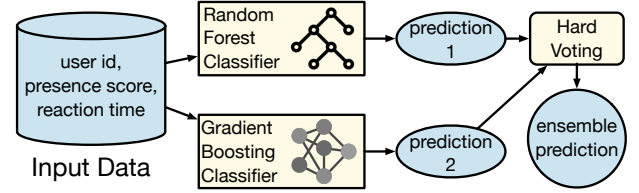

Fig. 14: *The architecture of reaction time–to–presence model.*

and outputs the level of presence for the user. We modeled the problem as a classification task, as the presence is measured in discrete states. Next, we describe the experimental setup and the training process.

6.2.1 Experimental Setup

Models. To perform the classification task, we utilize an ensemble machine learning model incorporating a hard voting classifier [71]. The model integrates a random forest classifier [70] and a gradient boost classifier [69]. Figure 14 shows the overall architecture of the classification model. We also tested several other classification models, such as K-nearest neighbors, support vector machines, and naive Bayes. However, they did not perform better than the ensemble model we chose for the classification task.

Dataset. Our dataset consists of reaction time and presence values for 40 users. For each user, we have reaction times and presence ratings under 4 conditions: realistic, abstract, plausible, and implausible. In total, our dataset contains 160 data points corresponding to the 4 conditions experienced by 40 users.

Training. We use an 80-20 dataset split for training and testing, respectively. We perform the split at the user-level data, meaning that all the data for a given user is used in training or testing. As a result, we use data from 32 randomly selected users for training and 8 for testing. We use grid search with 5-fold cross-validation for training the individual classifiers with the parameters specified in Table 5. Finally, we use a voting classifier with a hard voting configuration to choose between the two classifiers.

Output variable. Our presence questionnaire produced ratings of presence on a continuous scale from 0 to 7. However, it does not necessarily follow that we should design our model to estimate 7 discrete levels of presence. In many applications, knowing if presence is high or low could be sufficient. For this reason, we evaluated the efficacy of our model in estimating presence using either 2, 3, or 7 partitions of the continuous data. When using 2 levels, scores below 3.5 were assigned to level 1 and scores of 3.5 or above to level 2. When using 3 levels, scores between 0-2.33 were assigned to level 1, 2.34-4.66 to level 2, and 4.67-7 to level 3.

6.2.2 Model Evaluation

The accuracy of our model depends on two factors: the number of presence classes and the amount of data used for model training. Next, we evaluate the effect of these parameters on the model's accuracy.

Effect of Number of Presence Classes. Figure 15(a) shows the effect of number of presence levels (on x-axis) and model's accuracy in percentage (on y-axis). We use 80% of the data, 32 users, for training and 20% of the data, eight users, for testing. We observe that the accuracy

Random Forest Parameters	Values
no. of estimators	[50, 100, 150, 200, 500]
maximum depth	[None, 5, 10, 15, 20]
min. samples split	[2, 5, 10]
min. samples leaf	[1, 2, 4]
Gradient Boosting Parameters	**Values**
no. of estimators	[50, 100, 150, 200, 500]
maximum depth	[3, 5, 7]
learning rate	[0.01, 0.1, 0.2]

Table 5: *Parameters for the grid search with cross-validation.*

(a) Effect of Presence Levels   (b) Effect of Training Size

Fig. 15: *The effect of the number of presence levels (a) and the training data size (b) on the model's accuracy. The error bars represent the min and max values observed across 15 experiment runs.*

decreases as the number of presence classes increases. However, our model shows good accuracy despite the small data set of 128 observations. The accuracy is significantly higher than a purely random predictor with 1/(no. of presence levels), e.g., 50% for two levels, 33% for three levels, and 14.28% for seven levels.

Effect of Training Data Size. Figure 15(b) demonstrates the effect of training data size (on x-axis) on classification accuracy (on y-axis). In this experiment, we set the number of classes to two. We observe that model accuracy improves as the training data size increases. However, even with a very small training data size, the model performs quite well and achieves an accuracy of 73.09%. The accuracy improves to 79% when the number of users increases to 36. The upward trend shows that more data can further improve the model performance.
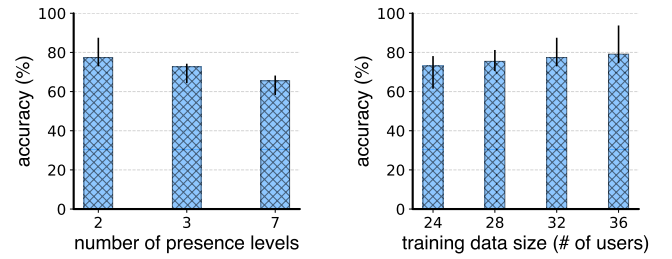
Key Takeaway. Our ensemble classification model estimates presence levels using the reaction time values with high accuracy, which depends on the number of presence levels and training data size. However, the accuracy can be further improved by using data from more users.

7   LIMITATIONS AND FUTURE WORK

For this study, we used two scenes consisting of simple one-object scenarios, a design choice to minimize the effect of variables other than presence on the reaction time. This study establishes the relationship between reaction time and presence when presence is altered by changing the realism and plausibility of virtual objects. While we do not have any indications that suggest this relationship will not hold if its presence is altered by any other method in a more complex environment, there is a possibility that the setup may not be sensitive to the broader effects of varying the feelings of presence on reaction time in a multi-object virtual scene. In future work, we plan to experimentally investigate how presence is affected by factors other than realism and plausibility and how it relates to reaction time with different degrees of scene complexity, cognitive load, and dynamic physical environment.

In this study, we investigated the effects of varying the feelings of presence with only periodic tasks and active interaction (air tap). It's worth investigating the relationship between presence and reaction time under different conditions, such as with non-periodic tasks, or with different response measures, such as eye gaze. Participants reported their presence levels using questionnaires that leverage Likert scales. Other questionnaires that rely on different assessment mechanisms, such as open-ended questions, might reveal additional insights. We also acknowledge that our proposed approach depends on user interactions to measure the reaction time. Our technique may fail to produce any measurement in virtual scenes with little or no interaction. Future work should consider using other backup mechanisms like eye-gaze tracking to combat low-interaction scenarios.

We have not found any effect of gender, age, or familiarity with MR on this study's presence or reaction time. This can be due to our purposefully simple design, but in a complex scenario with some applications, these user characteristics may affect the presence or reaction time. In addition, we have not tested the effect of the break in the presence, but it may impact the presence and reaction time. In the future, the break in the presence can be tested with the reaction time as a measure and may add to the discussion of the presence-reaction-time relationship. This study did not ask participants to complete a cybersickness questionnaire. However, during the study with participants, we asked participants to report any discomfort they felt during or at the end of the experiments. None of the participants reported any feeling of discomfort. This may be due to the limited exposure time and a wider adjustable interpupillary distance range that Hololens 2 offers [4]. However, this variable can be tested in isolation to isolate the potential impact of cybersickness in varying the feelings of presence.

Despite the limitations of the work and the opportunity for improvements, we argue that our results present sufficient evidence of a relationship between *presence* and *reaction time* to justify a further discussion of whether a performance-based metric such as *reaction time* can be used to describe *presence*. Post-experience questionnaires are the most commonly used measures of *presence* in previous work. However, a significant disadvantage of such questionnaires is that they are based on the subjects' memories of the experience. Such memories can reflect an inconsistent and incomplete picture of the experience. *Reaction time*, on the other hand, is a passive and objective measure that does not depend on the comprehension of the user and the memory of the experience. In our work, we have developed a preliminary model that maps the reaction time to presence. In the future, with additional investigation, this model can serve as a measure of *presence* and as a feedback loop that developers can use to improve the run-time experience since it measures the phenomenon when it is perceived.

Additionally, a model could be developed that takes the presence as the input and yields reaction time as the output, which could describe how much of an ill effect a decrease in presence might cause. However, it must be noted that reaction time is an objective metric that can be measured, and presence is a subjective sensation that even people themselves have difficulty reliably quantifying. This is why our paper aims to use reaction time as an alternative and potentially more robust mechanism for assessing the presence, not for predicting the effect of lower presence on reaction time.

8   CONCLUSION

We presented a user study (N=40) to understand the relationship between *presence* and *reaction time*. We changed the sense of *presence* of the participants by manipulating *appearance (place illusion)* and the non-task-relevant *behavior (plausibility illusion)* of the virtual object and systematically measured the *reaction time* of the participants in response to visual stimulus. Our post-experience questionnaires show a significant change in the *presence* across experiments. Similarly, we see a significant change in user *reaction time* as we vary feelings for *presence*. Our analysis shows a negative correlation between the presence and *reaction time*. Furthermore, our study provides insight into the considerations for using *reaction time* as a possible measure of *presence*, as well as preliminary recommendations on the possibilities of future research to understand better the relationship between *presence* and *reaction time*. We found that as the average *presence* score for the two illusions decreased from 4.97 to 2.74 and 5.17 to 2.81 (on a 7-point scale), the average *reaction time* increased by 37.63% and 27.10%, respectively. We developed a model that estimates a user's presence level using reaction time values with high accuracy of up to 80%. While our study suggests that reaction time can be used as a measure of presence, further investigation is needed to improve the accuracy of the model.

REFERENCES

[1] Air tap gesture. https://docs.microsoft.com/en-us/dynamics365/mixed-reality/guides/operator-gestures-hl2.

[2] Coffee Stack Game. https://rollicgames.com/games.

[3] Fruit ninja game. https://www.halfbrick.com/games/fruit-ninja.

[4] Hololens 2. https://www.microsoft.com/en-us/hololens.

[5] Intractable objects. https://docs.microsoft.com/en-us/windows/mixed-reality/design/interactable-object.

[6] Mixed reality capture. https://docs.microsoft.com/en-us/windows/mixed-reality/develop/advanced-concepts/mixed-reality-capture-overview.

[7] Windows Device Portal. https://learn.microsoft.com/en-us/windows/uwp/debug-test-perf/device-portal.

[8] R. M. Baños, C. Botella, A. Garcia-Palacios, H. Villa, C. Perpiñá, and M. Alcaniz. Presence and reality judgment in virtual environments: a unitary construct? *CyberPsychology & Behavior*, 3, 2000.

[9] W. Barfield. Musings on presence twenty-five years after" being there". *Presence Teleoperators Virtual Environ.*, 25, 2016.

[10] W. Barfield and S. Weghorst. The sense of presence within virtual environments: A conceptual framework. *Advances in Human Factors Ergonomics*, 19, 1993.

[11] C. Basdogan, C.-H. Ho, M. A. Srinivasan, and M. Slater. An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction*, 7, 2000.

[12] T. Baumgartner, L. Valko, M. Esslen, and L. Jäncke. Neural correlate of spatial presence in an arousing and noninteractive virtual reality: An eeg and psychophysiology study. *Cyberpsychology & Behavior : the Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 9, 2006.

[13] 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.

[14] K.-E. Bystrom, W. Barfield, and C. Hendrix. A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 8, 1999.

[15] 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.

[16] N. Cooper, F. Milella, C. Pinto, I. Cant, M. White, and G. Meyer. The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PloS one*, 13, 2018.

[17] R. P. Darken, D. Bernatovich, J. P. Lawson, and B. Peterson. Quantitative measures of presence in virtual environments: the roles of attention and spatial comprehension. *CyberPsychology & Behavior*, 2, 1999.

[18] P. De Boeck and M. Jeon. An overview of models for response times and processes in cognitive tests. *Frontiers in psychology*, 10, 2019.

[19] 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.

[20] A. Erickson, K. Kim, G. Bruder, and G. F. Welch. Exploring the limitations of environment lighting on optical see-through head-mounted displays. In *Symposium on Spatial User Interaction*, 2020.

[21] W. M. Felton and R. E. Jackson. Presence: A review. *International Journal of Human Computer Interaction*, 38, 2022.

[22] 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.

[23] J. Freeman and S. E. Avons. Focus group exploration of presence through advanced broadcast services. In *Human Vision and Electronic Imaging V*, vol. 3959. SPIE, 2000.

[24] J. Freeman, S. E. Avons, R. Meddis, D. E. Pearson, and W. IJsselsteijn. Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators & Virtual Environments*, 9, 2000.

[25] 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.

[26] M. Giesel, A. Nowakowska, J. Harris, and C. Hesse. Perceptual uncertainty and action consequences independently affect hand movements in a virtual environment. *Scientific Reports*, 10: 22307, 12 2020.

[27] F. Graham. Attention: The heartbeat, the blink and the brain. *Attention and Information Processing in Infants and Adults: Perspectives from Human and Animal Research (Eds: B.A. Campell, H. Hayne, R. Richardson)*, 1992.

[28] S. Grassini and K. Laumann. Questionnaire measures and physiological correlates of presence: A systematic review. *Frontiers in Psychology*, 11, 2020.

[29] J. Grosbois, M. Heath, and L. Tremblay. Augmented feedback influences upper limb reaching movement times but does not explain violations of fitts' law. *Frontiers in psychology*, 6, 2015.

[30] C. Heeter. Reflections on real presence by a virtual person. *Presence: Teleoperators & Virtual Environments*, 12, 2003.

[31] H. G. Hoffman, T. Richards, B. Coda, A. Richards, and S. R. Sharar. The illusion of presence in immersive virtual reality during an fmri brain scan. *CyberPsychology & Behavior*, 6, 2003.

[32] M. P. Huang and N. E. Alessi. Presence as an emotional experience. *Studies in health technology and informatics*, 1999.

[33] W. IJsselsteijn, I. Bierhoff, and Y. Slangen-de Kort. Duration estimation and presence. In *Presence Conference*, 2001.

[34] W. Ijsselsteijn, H. de Ridder, R. Hamberg, D. Bouwhuis, and J. Freeman. Perceived depth and the feeling of presence in 3dtv. *Displays*, 18, 1998.

[35] W. A. IJsselsteijn. Presence in depth. *Technische Universiteit Eindhoven Eindhoven, The Netherlands*, 2004.

[36] D. P. Jang, I. Kim, S. Nam, B. Wiederhold, M. Wiederhold, and S. Kim. Analysis of physiological response to two virtual environments: Driving and flying simulation. *Cyberpsychology & Behavior : the Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 5, 2002.

[37] T. F. I. J.D. Prothero, D.E. Parker. Towards a robust, quantitative measure for presence. *Proceedings of the Conference on Experimental Analysis and Measurement of Situation Awareness*, 1995.

[38] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3, 1993.

[39] 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.

[40] T. Kim and F. Biocca. Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of computer-mediated communication*, 3, 1997.

[41] J. Laarni, N. Ravaja, and T. Saari. Using eye tracking and psychophysiological methods to study spatial presence. In *Annual International Workshop on Presence*, 2003.

[42] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm. Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30, 1993.

[43] P. Larsson, D. Västfjäll, and M. Kleiner. The actor-observer effect in virtual reality presentations. *CyberPsychology & Behavior*, 4, 2001.

[44] 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.

[45] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3, 1997.

[46] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3, 1997.

[47] A. Maneuvrier, L. M. Decker, H. Ceyte, P. Fleury, and P. Renaud. Presence promotes performance on a virtual spatial cognition task: impact of human factors on virtual reality assessment. *Frontiers in Virtual Reality*, 1, 2020.

[48] M. Marucci, G. Di Flumeri, G. Borghini, N. Sciaraffa, M. Scandola, E. F. Pavone, F. Babiloni, V. Betti, and P. Aricò. The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Nature Publishing Group Scientific Reports*, 11, 2021.

[49] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks. Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics*, 21, 2002.

[50] Microsoft. Hololens 2 gestures for authoring and navigating in dynamics 365 guides. https://learn.microsoft.com/en-us/dynamics365/mixed-reality/guides/authoring-gestures-hl2, 2021.

[51] Microsoft. Improve visual quality and comfort. `https://learn.microsoft.com/en-us/hololens/hololens-calibration#calibrating-your-hololens-2`, 2022.

[52] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. Augmented Reality: A Class of displays on the Reality-Virtuality Continuum. In *Telemanipulator and Telepresence Technologies*. SPIE, 1995.

[53] E. B. Nash, G. W. Edwards, J. A. Thompson, and W. Barfield. A review of presence and performance in virtual environments. *International Journal of Human–Computer Interaction*, 12, 2000.

[54] J. R. J. Neo, A. Won, and M. Shepley. Designing immersive virtual environments for human behavior research. *Frontiers in Virtual Reality*, 2, 2021.

[55] K. Nesbitt, S. Davis, K. Blackmore, and E. Nalivaiko. Correlating reaction time and nausea measures with traditional measures of cybersickness. *Displays*, 48, 2017.

[56] S. Nichols, C. Haldane, and J. R. Wilson. Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies*, 52, 2000.

[57] 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.

[58] 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.

[59] 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.

[60] X. Retaux. Presence in the environment: Theories, methodologies and applications to video games. *PsychNology J.*, 1, 2003.

[61] 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.

[62] 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.

[63] M. Sanchez-Vives and M. Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6, 2005.

[64] T. Schubert. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, 15, 2003.

[65] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10, 2001.

[66] V. Schwind, P. Knierim, L. Chuang, and N. Henze. "where's pinky?": The effects of a reduced number of fingers in virtual reality. In *ACM CHI Conference on Human Factors in Computing System in Play*, 2017.

[67] V. Schwind, P. Knierim, N. Haas, and N. Henze. Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019.

[68] V. Schwind, P. Knierim, C. Tasci, P. Franczak, N. Haas, and N. Henze. "these are not my hands!": Effect of gender on the perception of avatar hands in virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017.

[69] Scikit-Learn. Gradient Boosting for Classification. `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html`, Accessed August 2023.

[70] Scikit-Learn. A Random Forest Classifier. `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html`, Accessed August 2023.

[71] Scikit-Learn. Soft Voting/Majority Rule Classifier for Unfitted Estimators. `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html`, Accessed August 2023.

[72] T. B. Sheridan. Musings on telepresence and virtual presence. *Presence Teleoperators & Virtual Environment*, 1, 1992.

[73] 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.

[74] M. Slater, V. Linakis, M. Usoh, and R. Kooper. Immersion, presence and performance in virtual environments: An experiment with tri-dimensional chess. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 1996.

[75] M. Slater and A. Steed. A virtual presence counter. *Presence*, 9, 2000.

[76] M. Slater, M. Usoh, and Y. Chrysanthou. The influence of dynamic shadows on presence in immersive virtual environments. *Eurographics Workshops on Virtual Environments*, 1999.

[77] M. Slater, M. Usoh, and A. Steed. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction*, 2, 1995.

[78] M. Slater and S. Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6, 1997.

[79] A. Spagnolli, D. Varotto, and G. Mantovani. An ethnographic, action-based approach to human experience in virtual environments. *International Journal of Human-Computer Studies*, 59, 2003.

[80] M. Speicher, B. D. Hall, and M. Nebeling. What is mixed reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–15. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300767

[81] A. Steed and J. McDonnell. Experiences with repertory grid analysis for investigating effectiveness of virtual environments. *The Online Proceedings of Presence*, 2003.

[82] J. Steuer. Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42, 1992.

[83] B. Stevens and J. Jerrams-Smith. The sense of object-presence with projection-augmented models. In *International Workshop on Haptic Human-Computer Interaction*, 2000.

[84] 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.

[85] K. Szczurowski and M. Smith. Measuring presence: hypothetical quantitative framework. In *23rd International Conference on Virtual System & Multimedia*, 2017.

[86] S. Turner, P. Turner, F. Carroll, S. O'Neill, D. Benyon, R. McCall, and M. Smyth. Re-creating the botanics: towards a sense of place in virtual environments. In *The 3rd UK Environmental Psychology Conference*, 2003.

[87] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence*, 9, 2000.

[88] 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.

[89] B. K. Wiederhold, R. Davis, and M. D. Wiederhold. The effects of immersiveness on physiology. *Studies in Health Technology and Informatics*, 1998.

[90] 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.

[91] J. R. Wilson, S. Nichols, and C. Haldane. Presence and side effects: Complementary or contradictory? In *HCI (2)*. Elsevier, 1997.

[92] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators & Virtual Environments*, 7, 1998.