Diegetic Representations for Seamless Cross-Reality Interruptions

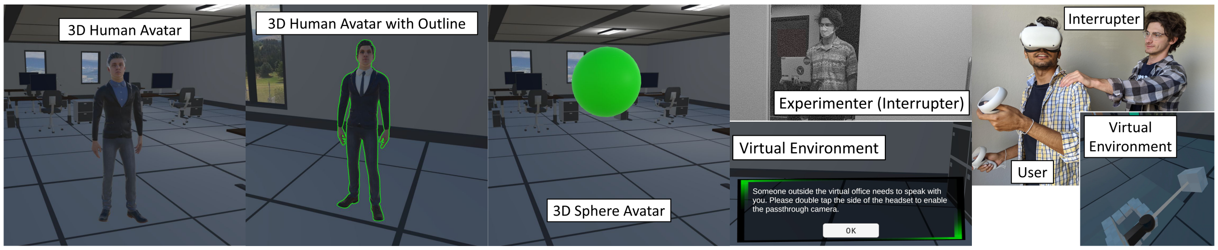
Matt Gottsacker*
SREAL, University of Central Florida

Nahal Norouzi†
SREAL, University of Central Florida

Kangsoo Kim‡
SREAL, University of Central Florida
University of Calgary

Gerd Bruder§
SREAL, University of Central Florida

Greg Welch¶
SREAL, University of Central Florida



(a) FDA (b) PDA (c) NDA (d) UIP (e) Baseline

Figure 1: Illustrations of the five experimental conditions in the experiment: (a) fully diegetic avatar, (b) partially diegetic avatar, (c) non-diegetic avatar, (d) user interface with passthrough camera view, and (e) non-mediated baseline interaction (tapping on one's shoulder without a visual representation).

ABSTRACT

Due to the closed design of modern virtual reality (VR) head-mounted displays (HMDs), users tend to lose awareness of their real-world surroundings. This is particularly challenging when another person in the same physical space needs to interrupt the VR user for a brief conversation. Such interruptions, e.g., tapping a VR user on the shoulder, can cause a disruptive break in presence (BIP), which affects their place and plausibility illusions, and may cause a drop in performance of their virtual activity. Recent findings related to the concept of *diegesis*, which denotes the internal consistency of an experience/story, suggest potential benefits of integrating registered virtual representations for physical interactors, especially when these appear internally consistent in VR. In this paper, we present a human-subject study we conducted to compare and evaluate five different diegetic and non-diegetic methods to facilitate cross-reality interruptions in a virtual office environment, where a user's task was briefly interrupted by a physical person. We created a Cross-Reality Interaction Questionnaire (CRIQ) to capture the quality of the interaction from the VR user's perspective. Our results show that the diegetic representations afforded the highest quality interactions, the highest place illusions, and caused the least disruption of the participants' virtual experiences. We found reasonably high senses of co-presence with the partially and fully diegetic virtual representations. We discuss our findings as well as implications for practical applications that aim to leverage virtual representations to ease cross-reality interruptions.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality;

*email: gottsacker@knight.ucf.edu
†e-mail: nahal.norouzi@knights.ucf.edu
‡e-mail: kkim.ucalgary@gmail.com
§e-mail: bruder@ucf.edu
¶e-mail: welch@ucf.edu

1 INTRODUCTION

Immersive virtual reality (VR) technologies, such as head-mounted displays (HMDs), afford users the multi-sensory illusion that they are "present" [22] in a computer-generated virtual space, indicated by a sense of "being there" in the virtual environment (VE), known as *place illusion*, and a sense of internal consistency and plausibility, known as *plausibility illusion* [41]. Practically, a high sense of presence in a virtual space often coincides with a reduced awareness of the VR user's physical surroundings. This is particularly challenging when another person in the same physical space intends to interact with the VR user, be them friends, family members, or co-workers. For instance, this may require tapping a VR user on the shoulder to make them aware of one's presence. Such cross-reality interruptions may cause severe breaks in presence (BIPs) for VR users, which may disrupt their place and plausibility illusions, but may also affect their virtual activities or task performance [40]. Some interruptions, such as unrelated tasks or high priority communications, may require the VR user's full attention outside the VE and thus call for a BIP. However, many other interruptions can be resolved by brief interactions in the VE and can benefit from a more seamless interaction experience that does not require substantial transitions in the user's place illusion. For example, a collaborative scenario with a VR user and one or more non-VR users may involve many interjections by the non-VR users as they suggest improvements or relay updates from the physical environment that need to be addressed in the VE. Reducing the negative effects of such cross-reality interruptions is an important topic for many applications as well as a challenging topic for basic research as these interruptions bridge Milgram's reality-virtuality continuum [26]. Such interruptions require VR users to either focus entirely on the real world or mentally synchronize the location and state of a physical interactor with their perceived VE.

A promising approach for more seamless interruptions across Milgram's continuum may be achieved through the notion of *diegesis*, which has recently gained interest in the fields of virtual/augmented reality and human-computer interaction [32]. Diegesis is defined as the internal consistency of an experience or story/narration, including the appearance and behavior of entities/objects and their environment [14, 32]. In the VR community, diegetic representations are receiving a growing interest, e.g., they are utilized in user interfaces and when a user's attention needs to be attracted/detracted with minimal disruption to a user's sense of presence [27, 32, 36, 37]. We believe that diegetic approaches may be adopted for seamless cross-reality interruptions by presenting a diegetic representation of a physical interrupter to a VR user. This kind of interaction may allow the VR user to keep the headset on, eliminating the disruption caused by doffing and donning the HMD, and enabling the user to both interact with that person and maintain a high sense of presence in VR.

This research fits into George et al.'s recently published *SeaT* design space of cross-reality interactions and transitions [11] in the following dimensions: Social Interaction & Collaboration Motivation, System-Triggered Availability, Visual & Audio Modality, and Gradual Act of Transitioning, which are lacking in effective solutions [11].

In this scope, we considered the following research questions:

- **RQ1**: How do different degrees of diegesis during cross-reality interruptions affect users' sense of presence in the physical and virtual environments before, during, and after the interaction?
- **RQ2**: How do different degrees of diegesis during cross-reality interruptions affect users' awareness of the physical and virtual environments, social behaviors, and task performance?
- **RQ3**: How do diegetic representations of a person during an interruption affect users' perception of them as real/virtual humans?

We addressed these research questions by performing a human-subject study to explore and compare different virtual representations from fully-diegetic to non-diegetic forms and evaluated them against a hardware solution based on a passthrough camera view, and a baseline in which participants took off the VR HMD (see Figure 1). Interaction is by nature two-sided; however, previous research has identified various research opportunities for creating a less disruptive and more seamless user experience for interrupted users [46]. Thus, we studied the effects of these conditions from the VR user's perspective. We measured participants' performance completing a virtual task as well as their sense of co-presence with the virtual representations, and we introduce a Cross-Reality Interaction Questionnaire (CRIQ) to capture participants' experience and perception of the interaction across realities. Our results show significant benefits of fully diegetic and partially diegetic representations for such interruptions across different measures.

The remainder of this paper is structured as follows. Section 2 provides an overview of related work on cross-reality interaction and diegetic representations. In Section 3 we describe our experiment. The results are presented in Section 4 and discussed in Section 5. Section 6 concludes the paper and discusses future research.

2 Related Work

In this section, we discuss related work on cross-reality interaction and diegetic representations.

2.1 Cross-Reality Interaction

As computing devices become more pervasive, researchers have started to explore interactions across Milgram's reality-virtuality continuum [26] using sensors and displays to provide innovative means for collaboration and shared experiences [20, 21, 23, 31, 35]. In the context of VR and non-VR collaborators, the cross-reality design space is interesting because each interactor's reality has its own set of perspectives and affordances. There is substantial work on solutions for bridging the gap between a VR user's physical environment and the VE so that the user may maintain awareness of co-located people, objects, or events.

By default, some consumer headsets support features that can support cross-reality interaction. For instance, the HTC Vive's "Knock Knock" feature allows a person in the real world to press a button on a tethered VR computer's keyboard to notify the VR user that someone outside the VE would like to talk to them [9]. Zenner et al. extended this approach by providing adaptive notifications with varying priority to the VR user [56]. The Oculus Quest further has a passthrough feature that allows users to double tap the side of the headset to enable a grayscale, real-time video view of their physical surroundings [28]. Through this feature, they can keep the HMD on their head while they look at the person in the real world.

Most research has focused on providing VR users uni-directional awareness of other people and their proximity in the physical environment, especially during short interruptions, which was rated as the most important aspect of awareness by VR users in a survey conducted by McGill et al. [25]. They explored this usability challenge in part through a prototype that portrayed a bystander in the VE as a transparent, ghost-like representation and as a real-time video cut-out. They found that users appreciated having knowledge of the bystander, but wanted a warning or more abstract representation of the bystander, while the real-time video of the bystander significantly disrupted the user's sense of presence. Simeone [39] created a related motion tracking widget to visualize the movement of persons in the physical environment to VR users with 2D representations. Ghosh et al. completed several prototypes to represent persons from the physical environment in VR using different modalities, including a video cut-out of a person and visualizing their location as footsteps [13]. They further completed a design exercise about best methods for notifying and interrupting VR users from outside the VE, suggesting the benefits of 3D representations for nearby people and objects ranging in detail from a glowing orb to a realistic model generated from real depth data, and using animations such as a wave to differentiate them from other in-world elements. However, they did not examine how they affected users' sense of presence. Willich et al. [50] further examined three different virtual representations for a bystander: an abstract avatar, a 3D point cloud of the bystander, and a 2D video of the bystander and surrounding environment. They found the avatar afforded the best spatial awareness, but felt less like interacting with a real person than the 2D video to the participants. They also found the 3D point cloud and avatar conditions reduced distraction. However, their study was not aimed at collaborative scenarios in which the bystander seeks productive communication with the VR user. In their RealityCheck system, Hartmann et al. [16] blended people from the physical environment into the virtual one as graphics objects or by allowing users to use the controllers as "flashlights" into the real world. RealityCheck effected higher presence scores, but there were no significant results for communication with the co-located mixed-presence bystanders. Williamson et al. [53] explored ways to support cross-reality interactions in an airplane setting, proposing initiating bystander interruptions naturally (via a gesture or common phrase) or via a peripheral device, and a passthrough view for the user to observe their surroundings.

While there is substantial work on technical solutions for cross-reality interactions, there is a gap in understanding how these interactions affect users' sense of place and plausibility, and how brief interruptions affect virtual activities and task performance.

2.2 Diegetic Representations

Diegesis which is originally a Greek word meaning *narrative* has been historically used in literature and film theory [5, 17, 44, 54]. In these domains, various definitions have been provided for diegesis. Gorbman defined diegesis as "narratively implied spatiotemporal world of the action and characters" [14], which is inspired by Genette's definition as he described diegetic as "what relates, or belongs, to the story" (translated by Bunia [5, 12]). In video games, Galloway described diegesis as "...the game's total world of narrative actions" [10]. More recently, the concept of diegesis is being used in gaming and cinematic VR to denote an internally consistent appearance and behavior of virtual user interface elements that integrate well with the VE and application context [10, 32, 36, 37].

In VR, with the increasing popularity of 360-degree experiences, such as cinematic VR and 360-degree videos, the concept of diegesis has been used to characterize mechanisms to guide a viewer's attention [27, 32, 42, 47]. In such 360-degree VR experiences, viewers are usually presented with the opportunity to explore the VE freely, and in some cases, this added freedom results in viewers missing parts of the story [32]. Therefore, in recent years, finding appropriate attention guiding mechanisms that can effectively guide users while maintaining their sense of presence has received a lot of attention [6, 27, 32–34, 42, 51, 55]. In this area, diegetic representations of cues are part of the story and the environment of the VR experience, while non-diegetic representations of cues are external elements to the story and the environment. For instance, movements of characters and sounds in the VE that can be seen and heard by other virtual characters are considered to be diegetic. On the other hand, using mechanisms with primitive objects, such as arrows and spheres, to guide a viewer's attention is considered non-diegetic.

Compared to non-diegetic cues, diegetic representation of cues are usually associated with or hypothesized as bringing about a higher sense of presence and improved user experience [32]. In a cinematic VR experience where participants embodied a captive in a cabin in the woods, Nielsen et al. [27] observed that participants found a diegetic firefly programmed to guide them to important parts of the story more helpful with the potential for higher levels of presence than a non-diegetic alternative of forced rotation. In another cinematic VR experience where participants were immersed in a virtual castle with dragons, Cao et al. [6] studied the effects of diegetic and non-diegetic representations by comparing birds and arrows as attention guiding cues, finding that birds were preferred by the majority of the participants.

The concept of diegesis is studied in other areas of VR research, such as redirected walking where researchers introduce attractors/detractors to mask rotations of the VE or ensure the user's safety by redirecting them, as maintaining the user's sense of presence is considered a primary goal [8, 29, 43]. For instance, in work by Sra et al. [43], they introduced the concept of *embedded context sensitive attractors* as "coherent with the narrative, related user interactions, and the virtual environment" which aligns with the concept of diegesis [10, 32]. In this field, Peck et al. [29] studied the influence of different types of distractors aimed at masking the VE's rotation with respect to the participants' sense of presence. They found that using visually and thematically more consistent distractors such as a hummingbird in an outdoor nature environment can lead to an increased sense of presence.

The findings above point towards the fact that diegetic representations, irrespective of their use case, show an increased potential for user preference and an improved sense of presence. Therefore, we believe it is valuable to utilize the notion of diegetic representations in cross-reality interactions where factors such as the user's sense of presence are equally important.

3 EXPERIMENT

In this section we describe the human-subject study we performed to compare different diegetic and non-diegetic methods to improve cross-reality interruptions.

3.1 Participants

After initial pilot tests, we estimated the effect size of the expected strong effects, and based on a power analysis, we made the decision to recruit 24 participants from our university community (17 identified as male and 7 identified as female; ages between 18 and 46, $M = 24.1$, $SD = 7.0$). Our experimental procedure and recruitment of participants were approved by the institutional review board of our university under protocol number SBE-17-13446. All of the participants had normal or corrected-to-normal vision. None of the participants reported known visual or vestibular disorders, such as dyschromatopsia or a displacement of balance. One participant reported being color blind, but as our experiment was designed with sufficient luminance differences across all colors, we did not consider this a reason for exclusion. 23 participants had used a VR HMD before. The participants were either students or non-student members of our university, who responded to open calls for participation, and received monetary compensation for their participation.

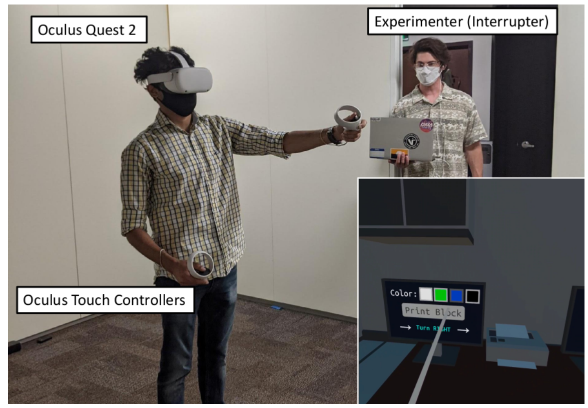


Figure 2: Annotated photo showing a participant in the experiment wearing the HMD and holding the controller during the task. The inset shows the virtual environment from the participant's view. The experimenter acted as the real interrupter in the experiment.

3.2 Material

In this section, we describe the experimental setup and virtual background task, during which the cross-reality interruptions occurred.

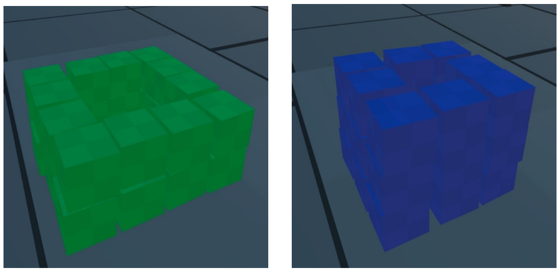
3.2.1 Physical and Virtual Environment

Figure 2 shows a participant in the physical environment for our study. For the experiment, we used a large open space in our laboratory dedicated for human-subjects studies. A desk was positioned off to the side for the participants to give their informed consent, receive study descriptions and instructions from the experimenter, and answer the questionnaires on one of the lab's laptops. We used an Oculus Quest 2 HMD and its default controllers for this experiment. The Quest 2 has a single fast-switch LCD display with a refresh rate of 90 Hz and a resolution of 1832×1920 pixels per eye.

We used the Unity Engine version 2019.4.15f1 LTS to create the experiment's VE and program its functionality. As shown in the inset in Figure 2, we created an office environment. The participant began in a training room with a single desk and several virtual screens providing instructions for controls and mechanics of the experiment. The participant then teleported down a hall and into a larger office space in which there were several desks with computers and chairs. The participant arrived at a desk with two virtual monitors and a 3D printer that allowed them to print virtual blocks with equal side lengths of 0.16 meters, which they then stacked on a floor-level platform behind them. A virtual human sat at one of the nearby desks performing an idle typing animation on their keyboard and muttered some canned, non-interactive statements throughout the experiment. Two other virtual humans walked up and down a hall at several points throughout the experiment. We included these virtual humans to simulate a more realistic virtual office environment.

3.2.2 Task: Stacking Virtual Blocks

The primary virtual task for the participants consisted of stacking virtual blocks in two different configurations and different colors (white, green, blue, gray) in the VE. First, the participants had to print a block via a virtual 3D printer by pressing virtual buttons for



(a) Configuration A (b) Configuration B

Figure 3: Examples for the two block stacking task configurations with randomized colors: (a) Configuration A (4 blocks long, 2 blocks high) and (b) Configuration B (3 blocks long, 3 blocks high). Both configurations comprise of 24 total blocks with comparable task difficulties.

color selection and block printing. Then, they had to turn around in alternating directions and place the block on a building platform located behind them in the configuration the current task prescribed. Configuration A was a structure with four walls, each of which were four blocks long and two blocks high. Configuration B called for four walls that were each three blocks long and three blocks high. The block configurations are shown in Figure 3. Both of these configurations comprise of 24 blocks in total. The configurations and colors were randomized between trials in the experiment. As these were cognitively and physically comparable tasks, we did not consider them as factors in our experiment and mainly included them to add some variance to the experiment trials.

We particularly chose this virtual block stacking task because it requires a consistent level of cognitive load, provides continuous visual feedback on progress, and attracts the participant's attention within a short amount of time. Although it might not be able to represent all VR experiences, the task had features that are required for our study purposes about task interruption and resumption.

When participants moved the 16th block in the 24-block configuration into the building area, an interruption began. The interruptions had a timed twelve-second onset delay during which the experimenter approached the participant's physical position. At the same time, in a synchronized position, and if the condition called for it, a virtual representation of the experimenter materialized in the VE and approached the participant, leading to a brief interruption of the virtual task. Synchronization between the experimenter and his virtual representation was achieved through a Wizard-of-Oz technique [15]. During this interruption, the experimenter explained the participant's subsequent task to work on after completing their current task. The experimenter then asked if the participant understood the next task and clarified if necessary. The participant was then free to complete their current task. Before the trials, the participant was informed they would be interrupted several times, but they were not given any instructions on how to handle the interruption.

The interruption and accompanying task instruction closely related to the primary task, a choice that corresponds to interruptions that may happen in cross-reality collaborative environments [30]. This choice also reduces some of the factors that contribute to the disruptiveness of the interruptions, including the interruption's duration, complexity, and similarity to the primary task [46]. Controlling these factors increases the likelihood that all interruptions are perceived as equally disruptive throughout a trial, yielding data that better shows the fundamental effectiveness of the interruption techniques and allows us to more reliably measure the participants' sense of presence in the VE over the course of the cross-reality interaction.

3.3 Methods

For this experiment, we used a within-subjects design with five different conditions. The order of the conditions was counter-balanced to reduce potential carryover effects.

3.3.1 Conditions

The five conditions tested different methods by which the experimenter interacted with the participants during the cross-reality interruptions (see Figure 1):

- Fully Diegetic Avatar (**FDA**): In this condition, a virtual human avatar walked toward the participant's location. A 3D audio source was used to render footstep sounds from the avatar's position, which were presented through the HMD's speakers. The appearance of this virtual avatar matched the other virtual humans and the surroundings.
- Partially Diegetic Avatar (**PDA**): In this condition, we used a thematically similar virtual avatar and the same footstep sounds as for the FDA condition, but we added an emissive material that created a virtual halo effect around the avatar. We included this halo effect to differentiate this avatar from the other virtual humans and the surroundings.
- Non-Diegetic Avatar (**NDA**): Instead of a human representation, this condition showed a sphere avatar, which hovered toward the participant. An atmospheric sound was rendered in 3D from the avatar's position and presented through the HMD's speakers.
- User Interface with Passthrough View (**UIP**): In this condition, a non-diegetic user interface notification appeared in front of the participant and a notification sound was presented through the HMD's speakers. The notification informed the participant that someone outside the virtual environment wanted to talk to them and to enable their passthrough camera by tapping twice on the side of the HMD.
- Non-Mediated Interaction (**Baseline**): In this condition, the experimenter tapped the participant on their shoulder[1] and then asked them to remove the HMD to talk to them.

Each condition presents different interpersonal cues related to social presence [22]. For the interaction conditions that are placed entirely in the virtual environment aurally and visually, the participant cannot observe their interaction partner's facial expression, style of dress, or body language. The passthrough camera displays these features, but in low fidelity. Of course, doffing the headset affords access to all features. In all cases, the participant can observe the experimenter's tone of voice. We used doffing the HMD as our baseline condition as opposed to an audio-only interruption to balance the level of embodiment of the interrupter to the participant across conditions.

3.3.2 Measures

Subjective Measures We utilized the following questionnaires to collect subjective responses from our participants.

- **Cross-Reality Interaction**: To examine the participants' perception of the cross-reality interaction, we devised our own questionnaire shown in Table 1. The sub-scales are *General*, *Place Illusion*, *Awareness*, *Behavioral Influence*, and *Plausibility Illusion*, each with multiple items on a 7-point scale. For this questionnaire, we modified relevant questions from the *Usability Experience Questionnaire* [19], *Temple Presence Inventory* [22], *Slater-Usoh-Steed Presence Questionnaire* [49], and *Godspeed Questionnaire* [2]. We based questions on Place Illusion and Plausibility on Slater's work [41]. The *General*

[1] This study was conducted during the COVID-19 pandemic. The experimenter followed all necessary guidelines, including using a stick to tap participants on the shoulder so they could maintain a social distance of 2 meters. We are confident that this caused no major differences in our participants' behaviors and results.

sub-scale captures elements of the user's experience and includes questions related to desirable qualities in brief cross-reality interactions. The *Awareness* sub-scale captures the extent to which the user felt aware of the physical/virtual environment. The *Behavioral Influence* sub-scale measures how much the physical/virtual environment influenced the user's behavior. The *Plausibility Illusion* sub-scale measures the extent to which the experience felt plausible. The *Place Illusion* sub-scale measures the user's sense of being in the VE instead of the physical environment before, during, and after the interaction, allowing us to analyze shifts in presence.

- **Co-Presence**: We used Basdogan et al.'s *Co-Presence Questionnaire* [3] for each interaction method to measure the extent to which the participants felt like they were interacting with a real human rather than a virtual representation. The questionnaire includes eight questions on a 1 to 7 scale, which are combined into a mean co-presence score (no sub-scales).

We also measured the participants' interruption methods preferences, and simulator sickness levels before and after the experience.

Objective Measures The temporal gap between the alert for an interruption to a primary task and the user's interaction with the interrupting content is known as the *interruption lag*, and the gap between the completion of the interruption interaction and the resumption of the primary task is known as the *resumption lag* [45].

- **Interaction Overhead**: We measured *Interaction Overhead* as the sum of the interruption lag and resumption lag. We considered the interruption alert to be when the virtual entities or notifications appeared in front of the user, and in the baseline condition when the experimenter tapped the participant on the shoulder. The primary task of stacking the virtual blocks was considered resumed when the interrupting interaction completed and the user moved a block onto the building platform. The interaction overhead measure provides insights into the effect of the interruption on the user's task performance.

3.3.3 Protocol

After reading a consent form and affirming their informed consent in the study, participants filled in a demographics questionnaire and pre-experiment simulator sickness questionnaire [18]. The experimenter gave a brief introduction to the study and guided the participants into the experiment's VE, which had two phases:

Phase 1 (Training): In the first phase, the participants were provided with basic instructions on how to use the HMD and handheld controllers to complete the virtual tasks. These instructions covered basic controls, including using a ray emitted from one of the controllers or directly using the other controller to pick up and place boxes, using the ray to interact with virtual controls, and teleporting. The participants were also instructed on how to enable the Oculus Quest 2's passthrough camera and given a chance to practice.

Phase 2 (Experiment Trials): In the second phase, participants were tasked with printing virtual blocks and stacking them in certain configurations (see Section 3.2.2). Once during each condition, the experimenter approached them while they were stacking the blocks, and interrupted them to give them instructions for their next task, which varied according to randomized colors and configurations shown in Figure 3. The methods used for these interruptions differed according to the conditions detailed in Section 3.3. The interruptions ended with the experimenter asking the participant to confirm they understood the instructions. If not, the experimenter clarified the instructions. This process repeated for all five conditions in a pre-determined, counter-balanced order. When the participants finished all conditions, the VE faded to black and the application exited.

After taking the HMD off, the participants filled out our post-experiment questionnaires. The participants then received a monetary compensation for taking part in this experiment.

Table 1: Cross-Reality Interaction Questionnaire (CRIQ). The sub-scales are: *General* (G1–G7), *Awareness* (A1–A2), *Behavioral Influence* (BI1–BI2), *Plausibility Illusion* (PSI1–PSI3), and *Place Illusion* (PI1–PI3). Each question is assessed on a 7-point scale *(1: not at all, 7: very much)*. Scales with an * are inverted for the analysis.

G1*	How *disruptive* was the interaction? (*Disruption*)
G2	How *predictable* was the interaction? (*Predictability*)
G3*	How *confusing* was the interaction? (*Confusion*)
G4	How *efficient* was the interaction? (*Efficiency*)
G5	How *realistic* was the interaction? (*Realism*)
G6	How *safe* did you feel during the interaction? (*Safety*)
G7	How *comfortable* did you feel during the interaction? (*Comfort*)
A1	During the interaction, how aware were you of the *virtual* environment/entities? (*Virtual Awareness*)
A2*	During the interaction, how aware were you of the *physical* environment/entities? (*Physical Awareness*)
BI1	During the interaction, how much was your behavior influenced by the *virtual* environment/entities? (*Virtual Influence*)
BI2*	During the interaction, how much was your behavior influenced by the *physical* environment/entities? (*Physical Influence*)
PSI1	How much did the interaction feel plausible to you? (*Plausibility*)
PSI2	How much did you have the perception of a synchronized interaction between the physical and virtual environment? (*Synchronicity*)
PSI3	How much did you feel like the experience was responding to you (interactive)? (*Responsiveness*)
PI1	*Before* the interaction, was your sense of being in the virtual environment stronger than your sense being in the physical environment? (*Place–Before*)
PI2	*During* the interaction, was your sense of being in the virtual environment stronger than your sense being in the physical environment? (*Place–During*)
PI3	*After* the interaction, was your sense of being in the virtual environment stronger than your sense being in the physical environment? (*Place–After*)

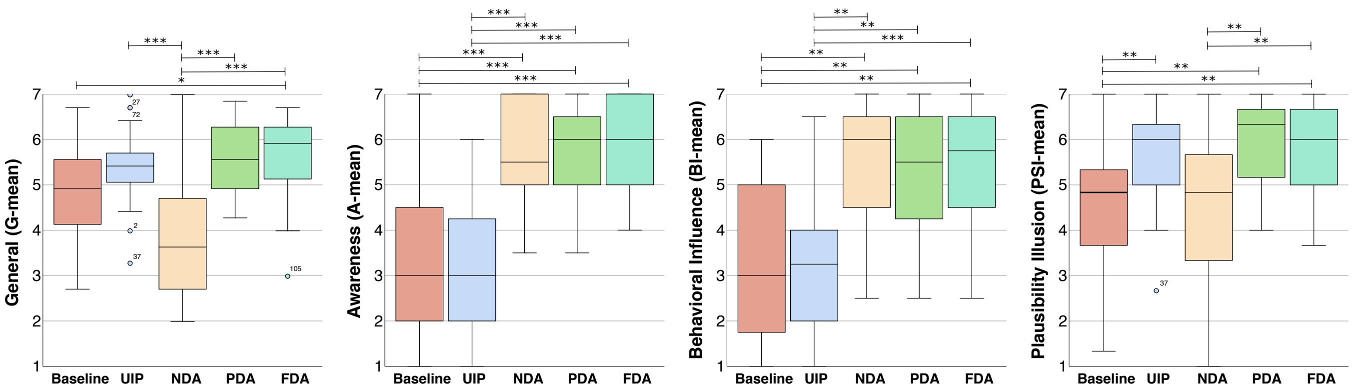
3.3.4 Hypotheses

Based on the literature in this field and our aforementioned research questions we arrived at the following hypotheses:

- **H1**: The FDA condition will score highest among the *General*, *Awareness*, *Behavioral Influence*, and *Plausibility Illusion* sub-scales of the CRIQ, followed by PDA, NDA, UIP, and Baseline.
- **H2**: The Baseline and UIP conditions will cause higher drops in virtual *task performance* compared to the diegetic methods.
- **H3**: The Baseline condition will have the highest associated *co-presence* scores, the UIP condition will score slightly lower, followed by FDA, PDA, and NDA.
- **H4**: The Baseline will cause a substantial drop on the *Place Illusion* sub-scale of the CRIQ from *before* to *during* and *after* the interaction, followed by the UIP condition and the diegetic conditions, depending on their level of diegesis.
- **H5**: The FDA condition will be *preferred* by participants, followed by PDA, NDA, UIP, and the Baseline.

4 Results

In this section, we present the results of our experiment. We analyzed the results with a repeated measures ANOVA and Tukey multiple comparisons at the 5% significance level with Bonferroni correction. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated. We confirmed the normality assumptions of the parametric analysis methods.



(a) General (G) (b) Awareness (A) (c) Behavioral Influence (BI) (d) Plausibility Illusion (PSI)

Figure 4: Results of the first four CRIQ sub-scales (means) for each of the first four experimental conditions and baseline (higher is better). The whiskers indicate post-hoc test results (* p<0.05, ** p<0.01, *** p<0.001). The points outside the box plot whiskers represent outliers.

4.1 Subjective Measures

Cross-Reality Interaction Figure 4 shows the combined results for the first four sub-scales, Figure 5(a) shows CRIQ total score, and Figure 5(d) shows the individual results for the fifth sub-scale. Table 2 shows the ANOVA results for the main effects and Cronbach's alpha values for the CRIQ sub-scales in parentheses next to each sub-scale, which indicate high internal consistency (cf. [1, 24, 48, 52]), suggesting that the sub-scale items may be combined. The results of our post-hoc tests are shown as part of the plots.

All of our CRIQ sub-scales and the aggregated mean score showed significant main effects of *Condition*. For the CRIQ mean score, the Baseline, UIP, and NDA showed significantly lower scores than the PDA and FDA conditions. For the *General* and *Plausibility Illusion* sub-scales, both the Baseline and NDA conditions showed significantly lower scores than some of the other conditions. For the *Awareness* and *Behavioral Influence* sub-scales, both the Baseline and UIP conditions received significantly lower scores than the other conditions. Our results give insights into the different shortcomings of these three methods, which we discuss in Section 5.

We performed a variation of this analysis for the *Place Illusion* sub-scale as we believe that its individual items deserve special consideration. Additionally to the results shown in Table 2, we analyzed the relative changes in scores from *before* the interruption to *during* and *after* the interruption. For both the Baseline and UIP conditions, we found a significant effect of *Condition* on the change in the *Place Illusion* scores from *before* the interaction to *during* the interaction (both $\boldsymbol{p} < \mathbf{0.001}$), and from *before* the interaction to *after* the interaction (both $\boldsymbol{p} < \mathbf{0.001}$). This indicates that the participants' place illusion decreased significantly as a result of the cross-reality interruption for the Baseline and UIP conditions, while it showed no noticeable effect for the virtual representation conditions.

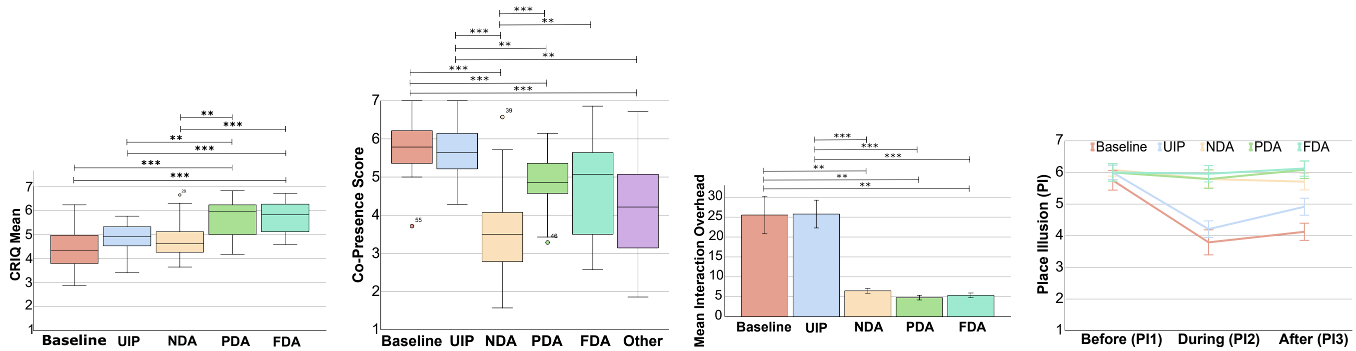
Co-Presence Figure 5(b) shows the pooled results of Basdogan et al.'s *Co-Presence Questionnaire* for the experiment. Additionally to our five experimental conditions, we also report co-presence scores for the non-interactive virtual agents that were shown as part of the VE. We found a significant main effect of *Condition* on the co-presence scores, $F(3.5, 81.2) = 20.36$, $\boldsymbol{p} < \mathbf{0.001}$, $\eta_p^2 = 0.470$. The results of our post-hoc tests are shown in the figure. Our results show that the Baseline and UIP conditions received significantly higher scores than the virtual representations and the non-interactive virtual agents. The PDA and FDA conditions scored fairly high as well, but the latter showed less evenly-distributed scores. The NDA condition was the worst in terms of perceived co-presence.

Preferences The aggregate rankings from most preferred to least preferred were: PDA, FDA, UIP, NDA, and Baseline. The PDA was rated best or second best by 17 out of our 24 participants. The FDA was rated best or second best 12 times. The UIP was rated best or second best 11 times. The NDA and Baseline conditions were rated best or second best each only 5 times, while the Baseline condition was rated worst a total of 11 times. We found a significant main effect of *Condition* on preference rankings, $F(2.6, 60.7) = 4.4$, $\boldsymbol{p} = \mathbf{0.01}$, $\eta_p^2 = 0.160$. The Bonferroni corrected pairwise comparisons showed no significant effects.

Simulator Sickness We measured a mean pre-SSQ sickness score of $M = 8.1$ ($SD = 16.5$) before the experiment and a mean post-SSQ score of $M = 24.9$ ($SD = 22.3$) after the experiment. On an absolute scale, these post-SSQ scores indicate a comparatively low/moderate amount of simulator sickness. As expected for an immersive VR experiment, the increase in simulator sickness symptoms was significant, $t(23) = 3.89$, $\boldsymbol{p} = \mathbf{0.001}$.

Table 2: ANOVA results for the individual CRIQ questions and the five sub-scales (means). The Cronbach α computed for the CRIQ sub-scales is presented next to each sub-scale.

Measure (Cronbach α)	*df*	*dfe*	*F*	*p*	η_p^2
G1*	3.6	81.9	6.3	< **0.001**	0.214
G2	3.6	83.1	2.7	**0.042**	0.105
G3*	2.8	64.4	10.9	< **0.001**	0.322
G4	3.1	72.0	6.3	**0.001**	0.215
G5	3.0	69.1	16.7	< **0.001**	0.420
G6	3.0	68.5	7.1	< **0.001**	0.236
G7	2.4	55.4	10.2	< **0.001**	0.308
G-Mean (0.795)	3.2	73.6	13.6	< **0.001**	0.371
A1	2.4	55.0	22.5	< **0.001**	0.495
A2*	2.4	55.1	16.6	< **0.001**	0.420
A-Mean (0.700)	2.1	48.9	25.3	< **0.001**	0.524
BI1	1.8	40.8	22.4	< **0.001**	0.494
BI2*	2.2	50.8	8.8	< **0.001**	0.277
BI-Mean (0.656)	1.7	38.7	18.5	< **0.001**	0.446
PSI1	2.9	67.7	10.6	< **0.001**	0.315
PSI2	3.0	68.2	6.6	**0.001**	0.223
PSI3	2.5	58.3	10.3	< **0.001**	0.309
PSI-Mean (0.775)	2.9	65.8	10.6	< **0.001**	0.315
PI1	2.8	65.3	0.6	0.609	0.250
PI2	2.3	53.3	14.5	< **0.001**	0.386
PI3	2.6	60.6	20.7	< **0.001**	0.474
PI-Mean (0.742)	2.8	64.1	20.3	< **0.001**	0.468
CRIQ-Mean (0.826)	2.5	58.0	18.1	< **0.001**	0.440



(a) CRIQ-Mean (b) Co-presence (c) Interaction Overhead (d) Place Illusion

Figure 5: Results showing: (a) CRIQ-Mean showing total CRIQ scores for the five conditions (higher is better), (b) co-presence scores for the five conditions and also the results for the non-interactive virtual agents on the far right (higher is better), (c) mean interaction overhead indicating the time (in seconds) from the beginning of the interruption to resuming the task (lower is better), and (d) place illusion (PI) sub-scale of the CRIQ showing the scores before (PI1), during (PI2), and after (PI3) the interruption (higher is better). The whiskers indicate post-hoc test results (* $p<0.05$, ** $p<0.01$, *** $p<0.001$). The box plots were generated with SPSS, including the outlier point notation in (a) and (b).

4.2 Objective Measures

Interaction Overhead Figure 5(c) shows the *Interaction Overhead* results. We found a significant main effect of *Condition* on interaction overhead, $F(1.7, 39.8) = 17.7$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.435$. The results of our post-hoc tests are shown as part of the plot. Our results show that both the Baseline and UIP conditions resulted in significantly higher delays than the conditions with virtual representations.

5 Discussion

In this section, we summarize the main findings and discuss implications for the use of diegetic or non-diegetic methods during interruptions, while also addressing qualitative feedback we received from our participants, and limitations of our experiment.

5.1 Diegetic Representations Provide Better General User Experience, Awareness, Behavioral Influence, Plausibility, and Task Performance

In partial support of our Hypothesis **H1**, the seamless virtual avatar representations (FDA, PDA, and NDA) received higher scores in the *Awareness* and *Behavioral Influence* sub-scales of the CRIQ, and the diegetic representations (FDA and PDA) received higher scores in *General* and *Plausibility Illusion* sub-scales. In partial support of our Hypothesis **H2**, the seamless conditions also showed the lowest cost of interruption in terms of *Interaction Overhead*.

When the experimenter interrupted the participants in the Baseline condition, they experienced significantly lower awareness and behavioral influence of the VE than the interaction methods in which they were allowed to keep the HMD on. These results are expected as the participants had to visually and aurally leave the VE in order to interact with the experimenter. Also, these interaction methods caused large temporal disruptions in the participants' task performance—about 20 seconds more than other seamless interaction methods. It is also important to note that the VR user is able to continue working on the task at hand when the interrupter is represented in the VE.

The more abrupt transitions between the physical and virtual environments have more overhead than seamless interactions for several reasons. When doffing and donning the headset, the participant often experiences a need to re-adjust the HMD to ensure it fits properly, a process that is exacerbated if the participant wears glasses. In the UIP method, the notification UI did not provide any information about the experimenter's location, so the participant experienced additional reorientation steps if they were not already facing in the direction of the experimenter. Also, the participants had to execute a proper double tap on the side of the Oculus Quest 2, which proved difficult for some even after pre-experiment training. However, the *Plausibility Illusion* scores for the UIP condition were high and similar to the two diegetic avatar conditions, suggesting that all three supported a higher level of synchronicity and interactivity between the physical and virtual environments.

5.2 Diegetic Representations Provide Better Co-Presence than Non-Diegetic Representations, but Less than the Baseline or Passthrough Views

In line with our Hypothesis **H3**, we found that the Baseline and UIP conditions received the highest scores for *co-presence* with the interrupter. Several participants mentioned observing the interrupter's body language and facial expressions as an advantage of these mechanisms. We found no significant difference between these conditions and the fully diegetic avatar condition, but we found support for the partial diegetic and non-diegetic avatar conditions receiving lower co-presence scores. Interestingly, the non-diegetic avatar in the form of the glowing sphere scored the lowest for co-presence, similar to the scores reported for the passive virtual agents that afforded no interaction capabilities. Many participants commented that the sphere was odd, confusing, and made them feel unsafe. Others were unsure of its intentions and thought it might even collide with their block structures, suggesting that since it did not fit with the rest of the VE, it appeared ignorant/unaware of their context.

5.3 Diegetic Representations Allow Continuously High Place Illusion in the Virtual Environment

In line with our Hypothesis **H4**, we observed a significant reduction in the *Place Illusion* scores from before to during/after the interruption for the Baseline and UIP conditions. This is interesting as it demonstrates that both the task of doffing and donning the HMD in the Baseline condition as well as the task of switching on and off the passthrough camera view in the UIP condition had a significant effect on the participants' place illusion that lasted even after the initial break in presence and interaction occurred. These results indicate that the virtual representations and in particular the diegetic ones preserve the highest sense of place illusion between the physical and virtual environments. When considered together with the other CRIQ and co-presence results, this data indicates that interactors' diegetic avatars afford VR users meaningful interactions in their physical environment, allowing them to maintain a sense of "being there" in the VE when the interaction concludes with no penalty.

5.4 Diegetic Representations are Ranked Highest Among the Cross-Reality Interruption Methods

While the Bonferroni corrected comparisons for the ranking data were not significant, the significant main effect and the qualitative data from the participants' comments led us to see a potential for future research and explore these findings.

Although we cannot accept our Hypotheis **H5**, participants more often *preferred* the diegetic representations, while the Baseline condition was ranked lowest. Interestingly, the participants ranked the partially diegetic avatar higher than the fully diegetic avatar, even though we found no significant differences between these conditions among the CRIQ sub-scales, co-presence scores, or performance results. Regarding the partially diegetic avatar, one participant noted:

> *"I preferred the outlined avatar... It felt like I was in a video game, and that I was supposed to pay attention to the character because it had something important to say. I preferred that over the regular avatar."*

Several participants shared similar thoughts and indicated that the halo/outline of this avatar made it stand out in the virtual office space because it was clearly separate from the passive virtual humans. Several participants also thought the outline made the avatar more attention-getting and noticeable than the fully diegetic avatar. Regarding the fully diegetic avatar, one participant noted:

> *"The avatar without the outline was strange. I thought it was just an NPC [non-player character], and I wasn't sure if it was going to talk to me. I heard the footsteps but wasn't sure that it was going to approach me since other avatars walked by at different times."*

This confusion, which other participants shared, is possibly explained by related work on VR notifications/interruptions that suggests elements from the physical environment that demand the user's attention should be marked as different from other virtual scene elements [13]. These comments also align with research on the *Uncanny Valley* effect in virtual characters [38], where the fully diegetic avatar could be inherently uncanny [7]. This feature was amplified because it served as a communication medium for a physical person but gave no indication that it was separate from the rest of the VE.

Further, it is interesting that despite the technological friction in using the UI with passthrough camera view, participants still ranked it higher than the non-diegetic avatar method. One participant commented on this by saying:

> *"I liked the passthrough the most because it gave me control over the interaction. It felt like getting an email, and I could control when the task communication update began."*

This comment suggests that giving the interruptee some control over the interruption may reduce its disruptiveness [46]. Another participant noted for the passthrough condition:

> *"This was my favorite method of interaction because I didn't have to remove the headset, and I like being able to see the physical body language of the person talking to me. If the passthrough cameras were better (didn't make it look like I was underwater) it would be perfect for me."*

5.5 Limitations and Future Work

The methods for cross-reality interruptions we compared and evaluated in this experiment showed interesting effects and practical insights for possible applications. However, there are also a few limitations, which may lead to additional study ideas for future work.

One limitation of our experiment is that our avatars were registered to the interrupter's position, but not to their hand gestures, facial expressions, etc. With additional sensors, or by using image segmentation and classification on the video feed from the passthrough cameras [4], these features could be included, supporting more complex cross-reality interactions without having to remove the HMD. This may also improve the social fidelity of the interaction. Second, our interruptions were rather short and conceptually related to the participants' primary task. Longer or more disruptive interruptions may cause users to prefer a more substantial break with the VE and thus a less diegetic interruption method than a seamless integration of the interrupter. Third, we did not consider more varied or complex bi-directional interaction between the interrupter and the VR user, which may be worthwhile to explore in future work. Further, we believe it would be informative to examine these interactions from the interrupter's perspective to more fully understand the social dynamics involved. Examining these considerations could lead to a system that supports a variety of cross-reality contexts.

Additionally, our interruptions were triggered at pre-planned times, which is not how most interruptions occur. A full interruptions system needs to detect an interrupting person. There are existing technologies that could achieve this, including the HTC Vive "Knock Knock" [9] and additional algorithms run on an HMD's camera data. Displaying the interrupting person as a human avatar like in our experiment should generalize to most virtual contexts, but the specific appearance of the avatar needs to be diegetic to diverse and application-specific VEs. For example, a game set in a fantasy world may need to represent an interrupter in non-human ways in order for the interrupter to be diegetic. A deeper qualitative exploration is necessary to more fully understand the user experience and the relationships between the context of the experience (e.g., office, fantasy world, etc.) and different interruption mechanisms (e.g., diegetic vs. non-diegetic representations). Such qualitative analysis is an avenue for future work, as it could inform application developers and creators to identify and specify different interruption mechanisms for different experience contexts in their systems.

It would also be worthwhile to study these interactions in an augmented reality (AR) setting, e.g., where a user is wearing an AR HMD and may be interrupted by the virtual avatar of another user. We hypothesize that the user can benefit from an increased sense of co-presence when similar partially/fully diegetic representations are used to realize the interrupter.

6 Conclusion

In this paper, we described a human-subject study where we compared and evaluated five different diegetic and non-diegetic methods to facilitate brief cross-reality interruptions. Our results show that the diegetic avatar representations led to the highest subjective responses in terms of general user experience, awareness, behavioral influence, plausibility, place illusion and task performance scores, while affording a reasonably high sense of co-presence, high objective task performance, and were ranked highest among the tested methods. Of these, the partially diegetic avatar was ranked highest, with many participants noting that its outline made this virtual representation appear more useful and effective for facilitating the cross-reality interruptions than even our tested fully diegetic avatar condition. We discussed limitations of our study and avenues for future research such as exploring cross-reality AR interactions.

Acknowledgments

This material includes work supported in part by the National Science Foundation under Award Number 1564065 (Dr. Ephraim P. Glinert, IIS) and Collaborative Award Numbers 1800961, 1800947, and 1800922 (Dr. Ephraim P. Glinert, IIS) to the University of Central Florida, University of Florida, and Stanford University respectively; the Office of Naval Research under Award Numbers N00014-21-1-2578 and N00014-17-1-2927 (Dr. Peter Squire, Code 34); and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch).

REFERENCES

[1] S. F. Alfalah, J. F. Falah, T. Alfalah, M. Elfalah, N. Muhaidat, and O. Falah. A comparative study between a virtual reality heart anatomy system and traditional medical teaching modalities. *Virtual Reality*, 23(3):229–234, 2019.

[2] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.

[3] 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(4):443–460, 2000.

[4] G. Bruder, F. Steinicke, K. Rothaus, and K. Hinrichs. Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras. In *Proceedings of the International Conference on CyberWorlds*, pp. 43–50, 2009.

[5] R. Bunia. Diegesis and representation: beyond the fictional world, on the margins of story and narrative. *Poetics Today*, 31(4):679–720, 2010.

[6] C. Cao, Z. Shi, and M. Yu. Automatic generation of diegetic guidance in cinematic virtual reality. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, pp. 600–607, 2020.

[7] D. Chattopadhyay and K. F. MacDorman. Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. *Journal of Vision*, 16(11):7–7, 2016.

[8] Z.-H. Chen, Y.-C. Deng, C.-Y. Chou, and T.-W. Chan. Motivating learners by nurturing animal companions: My-pet and our-pet. In *AIED*, pp. 136–143, 2005.

[9] J. Durbin. Vive 'knock knock' lets you know people in the real world still care about you. `https://uploadvr.com/new-knock-knock-feature-lets-you-know-someone-needs-attention/`, May 16, 2016. [Accessed 2021-03-15].

[10] A. R. Galloway. *Gaming: Essays on algorithmic culture*, vol. 18. University of Minnesota Press, 2006.

[11] C. George, A. N. Tien, and H. Hussmann. Seamless, bi-direcational transitions along the reality-virtuality continuum: A conceptualization and prototype exploration. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, pp. 412–424, 2020.

[12] G. Gérard. Discours du récit. *Figures III*, pp. 65–278, 1972.

[13] S. Ghosh, L. Winston, N. Panchal, P. Kimura-Thollander, J. Hotnog, D. Cheong, G. Reyes, and G. D. Abowd. Notifivr: Exploring interruptions and notifications in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1447–1456, 2018.

[14] C. Gorbman. *Unheard melodies: Narrative film music*. Indiana University Press, 1987.

[15] P. Green and L. Wei-Haas. The Rapid Development of User Interfaces: Experience with the Wizard of Oz Method. In *Proceedings of the Human Factors Society—29th ANNUAL MEETING*, pp. 470–474, 1985.

[16] J. Hartmann, C. Holz, E. Ofek, and A. D. Wilson. Realitycheck: Blending virtual environments with situated physical reality. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, pp. 347:1–12, 2019.

[17] I. Iacovides, A. Cox, R. Kennedy, P. Cairns, and C. Jennett. Removing the hud: the impact of non-diegetic game elements and expertise on player involvement. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pp. 13–22, 2015.

[18] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993.

[19] B. Laugwitz, T. Held, and M. Schrepp. Construction and evaluation of a user experience questionnaire. In A. Holzinger, ed., *HCI and Usability for Education and Work*, pp. 63–76. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[20] M. Lee, N. Norouzi, G. Bruder, P. J. Wisniewski, and G. F. Welch. The physical-virtual table: Exploring the effects of a virtual human's physical influence on social interaction. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pp. 25:1–11, 2018.

[21] J. Lifton, M. Laibowitz, D. Harry, N. Gong, M. Mittal, and J. A. Paradiso. Metaphor and manifestation cross-reality with ubiquitous sensor/actuator networks. *IEEE Pervasive Computing*, 8(3):24–33, 2009.

[22] M. Lombard, T. B. Ditton, and L. Weinstein. Measuring presence: the temple presence inventory. In *Proceedings of the 12th Annual International Workshop on Presence*, pp. 1–15, 2009.

[23] C. Mai, S. A. Bartsch, and L. Rieger. Evaluating shared surfaces for co-located mixed-presence collaboration. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*, p. 1–5, 2018.

[24] G. Makransky and G. B. Petersen. Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers I& Education*, 134:15–30, 2019. doi: 10.1016/j.compedu.2019.02.002

[25] M. Mcgill, D. Boland, R. Murray-Smith, and S. Brewster. A dose of reality: Overcoming usability challenges in vr head-mounted displays. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, pp. 2143–2152, 2015.

[26] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77(12):1321–1329, 1994.

[27] L. T. Nielsen, M. B. Møller, S. D. Hartmeyer, T. C. Ljung, N. C. Nilsson, R. Nordahl, and S. Serafin. Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 229–232, 2016.

[28] Oculus. Oculus release notes. `https://support.oculus.com/release-notes/`, March 23, 2020. [Accessed 2021-03-15].

[29] T. C. Peck, H. Fuchs, and M. C. Whitton. Evaluation of reorientation techniques and distractors for walking in large virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 15(3):383–394, 2009.

[30] N. Peters, G. Romigh, G. Bradley, and B. Raj. When to interrupt: A comparative analysis of interruption timings within collaborative communication tasks. In I. L. Nunes, ed., *Advances in Human Factors and System Interactions*, pp. 177–187. Springer International Publishing, Cham, 2017.

[31] D. F. Reilly, H. Rouzati, A. Wu, J. Y. Hwang, J. Brudvik, and W. K. Edwards. Twinspace: An infrastructure for cross-reality team spaces. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, p. 119–128, 2010.

[32] S. Rothe, D. Buschek, and H. Hußmann. Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction*, 3(1):19, 2019.

[33] S. Rothe and H. Hußmann. Guiding the viewer in cinematic virtual reality by diegetic cues. In *Proceedings of the International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pp. 101–117, 2018.

[34] S. Rothe, H. Hußmann, and M. Allary. Diegetic cues for guiding the viewer in cinematic virtual reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–2, 2017.

[35] A. P. Ríos, V. Callaghan, M. Gardner, and M. J. Alhaddad. Using mixed-reality to develop smart environments. In *Proceedings of the International Conference on Intelligent Environments*, pp. 182–189, 2014.

[36] P. Salomoni, C. Prandi, M. Roccetti, L. Casanova, and L. Marchetti. Assessing the efficacy of a diegetic game interface with oculus rift. In *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pp. 387–392. IEEE, 2016.

[37] P. Salomoni, C. Prandi, M. Roccetti, L. Casanova, L. Marchetti, and G. Marfia. Diegetic user interfaces for virtual environments with hmds: a user experience study with oculus rift. *Journal on Multimodal User Interfaces*, 11(2):173–184, 2017.

[38] V. Schwind, K. Wolf, and N. Henze. Avoiding the uncanny valley in virtual character design. *interactions*, 25(5):45–49, 2018.

[39] A. L. Simeone. The VR motion tracker: visualising movement of non-participants in desktop virtual reality experiences. In *Proceedings of the IEEE Workshop on Everyday Virtual Reality*, pp. 1–4, 2016.

[40] M. Slater. Presence and The Sixth Sense. *Presence: Teleoperators and*

Virtual Environments, 11(4):435–439, 2002.

[41] 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(1535):3549–3557, 2009.

[42] M. Speicher, C. Rosenberg, D. Degraen, F. Daiber, and A. Krúger. Exploring visual guidance in 360-degree videos. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, pp. 1–12, 2019.

[43] M. Sra, X. Xu, A. Mottelson, and P. Maes. Vmotion: designing a seamless walking experience in vr. In *Proceedings of the ACM Conference on Designing Interactive Systems*, pp. 59–70, 2018.

[44] R. J. Stilwell. The Fantastical Gap between Diegetic and Nondiegetic. *Beyond the Soundtrack: Representing Music in Cinema*, 2007.

[45] J. G. Trafton, E. M. Altmann, D. P. Brock, and F. E. Mintz. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies*, 58(5):583–603, 2003.

[46] J. G. Trafton and C. A. Monk. Task interruptions. *Reviews of Human Factors and Ergonomics*, 3(1):111–126, 2007.

[47] S. Unseld. 5 lessons learned while making lost. `https://www.oculus.com/story-studio/blog/5-lessons-learned-while-making-lost/`, July 2015. [Accessed 2021-03-15].

[48] G. Ursachi, I. A. Horodnic, and A. Zait. How reliable are measurement scales? external factors with indirect influence on reliability estimators. *Procedia Economics and Finance*, 20:679–686, 2015.

[49] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, 9(5):497–503, 2000.

[50] J. von Willich, M. Funk, F. Müller, K. Marky, J. Riemann, and M. Mühlhäuser. You invaded my tracking space! using augmented virtuality for spotting passersby in room-scale virtual reality. In *Proceedings of the ACM Conference on Designing Interactive Systems*, p. 487–496, 2019.

[51] J. O. Wallgrün, M. M. Bagher, P. Sajjadi, and A. Klippel. A comparison of visual attention guiding approaches for 360° image-based vr tours. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 83–91, 2020.

[52] W. Wei, R. Qi, and L. Zhang. Effects of virtual reality on theme park visitors' experience and behaviors: A presence perspective. *Tourism Management*, 71:282–293, 2019. doi: 10.1016/j.tourman.2018.10.024

[53] J. R. Williamson, M. McGill, and K. Outram. Planevr: Social acceptability of virtual reality for aeroplane passengers. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, p. 1–14, 2019.

[54] B. Winters. The non-diegetic fallacy: Film, music, and narrative space. *Music and Letters*, 91(2):224–244, 2010.

[55] Q. Xu and E. D. Ragan. Effects of character guide in immersive virtual reality stories. In *Proceedings of the International Conference on Human-Computer Interaction*, pp. 375–391. Springer, 2019.

[56] A. Zenner, M. Speicher, S. Klingner, D. Degraen, F. Daiber, and A. Krüger. Immersive notification framework: Adaptive & plausible notifications in virtual reality. In *Extended Abstracts of the ACM CHI Conference on Human Factors in Computing Systems*, p. 1–6, 2018.