Evaluating Interaction Cue Purpose and Timing for Learning and Retaining Virtual Reality Training

Xinyu Hu
University of Central Florida
XinyuHu@knights.ucf.edu

Alec G. Moore
University of Central Florida
agm@knights.ucf.edu

James Coleman Eubanks
University of Texas at Dallas
j.coleman.eubanks@utdallas.edu

Afham Ahmed Aiyaz
University of Texas at Dallas
axa157730@utdallas.edu

Ryan P. McMahan
University of Central Florida
rpm@ucf.edu

ABSTRACT

Interaction cues inform users about potential actions to take. Tutorials, games, educational systems, and training applications often employ interaction cues to direct users to take specific actions at particular moments. Prior studies have investigated many aspects of interaction cues, such as the feedforward and perceived affordances that often accompany them. However, two less-researched aspects of interaction cues include the effects of their purpose (i.e., the type of task conveyed) and their timing (i.e., when they are presented). In this paper, we present a study that evaluates the effects of interaction cue purpose and timing on performance while learning and retaining tasks with a virtual reality (VR) training application. Our results indicate that participants retained manipulation tasks significantly better than travel or selection tasks, despite both being significantly easier to complete than the manipulation tasks. Our results also indicate that immediate interaction cues afforded significantly faster learning and better retention than delayed interaction cues.

CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI); Interaction paradigms; Virtual reality.

KEYWORDS

Interaction cues, Virtual reality, Training

ACM Reference Format:

Xinyu Hu, Alec G. Moore, James Coleman Eubanks, Afham Ahmed Aiyaz, and Ryan P. McMahan. 2020. Evaluating Interaction Cue Purpose and Timing for Learning and Retaining Virtual Reality Training. In *Symposium on Spatial User Interaction (SUI '20), October 31–November 01, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3385959.3418448

SUI '20, October 31–November 01, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-7943-4/20/10…$15.00
https://doi.org/10.1145/3385959.3418448

1 INTRODUCTION

Interaction cues provide information to users so that they are aware of their present interactive possibilities [1] and to guide users in the interaction process [32]. Tutorials and games commonly employ interaction cues to instruct users to perform specific actions at particular times [8]. For example, many consumer VR experiences, such as *The Lab* by Valve, visually and verbally prompt users to use the controllers and buttons to learn how to travel or how to manipulate objects.

Interaction cues, which convey an action to take, occur before the user's action or "pre-action" [9]. Interaction cues are often accompanied with perceived affordances and feedforward, which also occur pre-action [39]. Perceived affordances convey how to physically perform an available action [39], such as how to use inputs to manipulate an object [19]. On the other hand, feedforward communicates the purpose of an action and conveys what the result of that action will be [9], such as where to place an assembly piece [26]. Perceived affordances have been extensively investigated [12, 13, 27], and feedforward has received much more attention in recent years (e.g., [5–7, 9, 39, 41]).

In addition to perceived affordances and feedforward, Dillman et al. [8] recently presented a descriptive framework for visual interaction cues. They identified three additional aspects of interaction cues to consider: purpose, markedness, and trigger. More recently, Hu et al. [16] identified timing as another aspect of interaction cues. The *purpose* of an interaction cue is the task that the cue guides the user to take action on, such as traveling to a location or manipulating an object. The *markedness* of an interaction cue is the extent that the cue blends into the game environment. For example, interaction cues can be as subtle as contrasts in lighting or as distinguishable as integrated virtual arrows and labels. A cue's *trigger* is a predefined event or criteria that brings about its presentation, such as entering a room or activating an ability. Finally, once triggered, a cue's *timing* determines whether the cue is immediately presented or delayed.

These additional aspects of interaction cues have not been extensively researched. Yoshimura et al. [42] recently researched different styles of markedness for directing the user's attention with eye-gaze-triggered visual cues, but did not conduct a formal study. Hu et al. [16] conducted a between-subject user study to investigate the effects of cue timing at 0, 8, and 16 seconds for a VR training application. However, they did not find any significant differences among the three timing conditions, aside from the result that the immediate 0-second cues enabled users to complete the VR training significantly faster and with fewer errors than the delayed 16-second cues.

In order to advance knowledge about interaction cues and their aspects, we conducted a mixed-design user study investigating purpose within subject for three types of tasks (travel, selection, and manipulation) and timing between subjects at three levels (0, 8, and 16 seconds). We evaluated the effects of the two aspects on the learning and retention of tasks learned and assessed through a previously developed VR training application. Our results indicate that the travel tasks in our VR training application were easier to complete than the selection tasks, which were easier to complete than the manipulation tasks. Despite this, our results also indicate that participants better retained the manipulation tasks than the travel and selection tasks. Based on a post-hoc analysis, we hypothesize that this may have been due to some type of object-oriented, context-dependent memory [34]. Our results also indicate that the 0-second interaction cues afforded significantly faster learning and better retention than the 8- and 16-second cues, which implies that developers should avoid delayed cues.

2 RELATED WORK

In this section, we first discuss prior work relating to interaction cue purpose and differences among types of tasks. We then discuss examples of different interaction cue timings and a prior study investigating the effects of such timing differences.

2.1 Interaction Cue Purpose and Task Type

Dillman et al. [8] originally defined three purposes for interaction cues based on their ludography: discover, look, and go. Discover cues inform the user of objects or points of interest in the environment (e.g., [31]). Look cues inform the user where to put their visual attention (e.g., [3]). And, go cues provide navigational assistance (e.g., [14]). Based on their literature review, Hu et al. [16] identified four additional purposes for interaction cues: pick, manipulate, gesture, and speak. Pick cues inform the user to select a particular object or element (e.g., [26]). Manipulate cues inform the user to move an object to a new position (e.g., [18]). Gesture cues inform the user to perform specific movements with their body (e.g., [4]). Finally, speak cues inform the user to verbally communicate (e.g., [10]). In our study, we investigated go, pick, and manipulate cues, which we will refer to as travel, selection, and manipulation to match common VR terminology [20].

Several researchers have studied different types of interaction cues for the same purpose by investigating different types of perceived affordances and feedforward [20]. For example, Khuong et al. [18] explored two different interaction cues for conveying where to position the next piece in an assembly task: a) an in-place wireframe visualization of the target position over the real-world assembly, and b) a nearby visualization of a virtual replica of the assembly with the object in its target position. They found that the nearby interaction cue afforded significantly better performance than the in-place interaction cue.

However, to the best of our knowledge, the effects of the purpose of interaction cues have not been previously investigated. The closest related works are a couple studies by Tyndiuk et al. [37, 38] comparing the effects of display size and task type on visual scanning scores. In both studies, they found that a larger display improved visual scanning scores for a naïve travel task and a manipulation task, but not a primed travel task. Additionally, their results indicate that the primed travel task was much easier than the manipulation task [37, 38].

In our study, we empirically evaluate the effects of three interaction cue purposes (travel, selection, and manipulation) on learning and retaining VR training.

2.2 Interaction Cue Timing

In general, most interaction cues are immediately presented [16]. However, there are examples within the literature of delayed interaction cues that are presented later to give the user a chance to act first. For example, Sanchez et al. [33] have a virtual agent walk through the scene as an interaction cue to ask for help, if the user is not making progress on a game mission. Similarly, Ordaz et al. [28] provide a prompt for completing an assembly step, if the user struggles to position the part or tool for 30 seconds.

To the best of our knowledge, the only evaluation of the effects of interaction cue timing was recently presented by Hu et al. [16]. They conducted a between-subject study investigating the effects of cue timing at three levels: 0 seconds (i.e., immediate), 8 seconds, and 16 seconds. Their rationale for choosing these timing differences was because attention spans have been reported to be as little as 8 seconds [11]. Their results indicated that the 0-second condition allowed participants to complete a VR training application significant faster, with significantly fewer physical errors, and with significantly fewer incorrect decisions than the 16-second condition. However, they found no significant differences among the three cue timing conditions on a knowledge test administered immediately after the VR training.

In our study, we also investigate the effects of interaction cue timing at the same three levels (0, 8, and 16 seconds). However, unlike Hu et al. [16], we assess retention by having participants return one week after the initial VR training with interaction cues to experience the same VR training application, but without interaction cues. Furthermore, we also simultaneously investigate the effects of interaction cue purpose (travel, selection, or manipulation), in addition to cue timing.

3 VIRTUAL REALITY LEARNING AND RETENTION STUDY

The goal of this study was to evaluate the effects of interaction cue purpose and timing on learning and retaining VR training. For the study, we used a previously developed VR training application designed to teach how to troubleshoot a surgical robot (see section 3.4). Participants first learned about the troubleshooting procedure by completing the VR training application with interaction cues provided. One week later, participants then completed the same application without interaction cue to assess their retention.

3.1 Research Questions

Our specific research questions were:

Q1. Do the purposes of interaction cues have a significant effect on performance while learning the relevant steps of a procedure? H1. Yes, we hypothesized that interaction cues for travel would afford significantly better performance of their corresponding procedural steps than interaction cues for selection or manipulation would. We made this hypothesis made on the prior studies that indicate travel is an easier task than manipulation [37, 38].

Q2. Do the purposes of interaction cues have a significant effect on retaining the relevant steps of a learned procedure? H2. Yes, we hypothesized that travel interaction cues would afford significantly better retention of procedural steps than selection or manipulation interaction cues. We made this hypothesis based on our previous hypothesis and the Cognitive Load Theory (CLT), which indicates that increases in cognitive load (such as those due to increased task difficulty) can lead to decreases in learning [29], which will lead to decreases in retention [22].

Q3. Does the timing of interaction cues have a significant effect on performance while learning the relevant steps of a procedure? H3. Yes, we hypothesized that interaction cues presented immediately at 0 seconds would afford significantly faster learning than delayed interaction cues presented after 8 and 16 seconds. We made this hypothesis based on similar results presented by Hu et al. [16].

Q4. Does the timing of interaction cues have a significant effect on retaining the relevant steps of a learned procedure? H4. No, we hypothesized that there would be no significant differences between the immediate (0-second) and delayed (8- and 16-second) interaction cues in terms of retention. We made this hypothesis based on the results of Hu et al. [16], who found no significant differences between immediate and delayed interaction cues on a knowledge posttest.

3.2 Independent Variables

Our independent variables were the study *session* (learning or retention), the *purpose* of the interaction cues (travel, selection, and manipulation), and the *timing* of the interaction cues (at 0, 8, or 16 seconds). The session varied within subject, as every participant experienced the VR training application twice, once for initial learning and then one week later for assessing retention. The purpose of the interaction cues also varied within subject, as the VR training application involved all three types of tasks and included interaction cues for each task (see section 3.4). Finally, the timing of the interaction cues was controlled between subjects, with three cohorts (i.e., one for each timing condition).

3.3 Dependent Variables

Our dependent variables were based on objective metrics automatically collected by the VR training application for each step of the troubleshooting procedure being trained. Our first dependent variable was the *mean task time* required to complete each step. Our second dependent variable was the *mean task errors* committed while completing each step, grouped according to their task type (i.e., travel, selection, or manipulation). Potential errors included mistakes such as grabbing the wrong object, selecting the wrong dialog option, or manipulating an object incorrectly. Our final dependent variable was the *mean task success* of completing each step without any errors and before a 30-second timeout. Whenever an error or timeout occurred, a virtual agent within the VR training application would verbally instruct the participant and the interaction cue for the relevant step would be presented (the first time in the retention version or again in the learning version).

Table 1: The purpose of each interaction cue presented in the evaluated VR training application.

Step	Purpose	Step	Purpose
1	Travel	13	Manipulation
2	Selection	14	Manipulation
3	Selection	15	Manipulation
4	Selection	16	Selection
5	Selection	17	Selection
6	Selection	18	Travel
7	Selection	19	Selection
8	Selection	20	Manipulation
9	Selection	21	Manipulation
10	Manipulation	22	Selection
11	Manipulation	23	Travel
12	Travel	24	Selection

3.4 VR Training Application

We had to choose whether to develop a VR training application designed specifically for our study or to use a preexisting, more ecologically valid, VR training application. We chose the latter to more closely approximate a practical, real-world VR training system (at the expense of more scientific control and statistical power) and to more quickly conduct our initial investigation. We ultimately decided to use a VR application that was previously developed for training how to troubleshoot a surgical robot [24].

The application's training scenario involves interacting with a virtual operating room (OR) team to troubleshoot the surgical robot after an instrument failure. To accomplish the scenario, the application requires the user to complete a variety of steps, including moving about the OR (i.e., travel via real walking [20]), selecting dialog options and picking up tools (i.e., selection via virtual hand [25]), and interacting with instruments and the surgical robot (i.e., manipulation via virtual hand [23]). The scenario involves a total of 24 steps, including 4 travel tasks, 13 selection tasks, and 7 manipulation tasks (see Table 1).

During the learning session, the VR training application provided an interaction cue for each step. Each interaction cue consisted of subtle verbal instructions from one of the OR virtual agents and integrated visual animations as perceived affordances and feedforward. The travel cues used semi-transparent green boots continuously linearly interpolated from the user's current position to a green "Stand Here" icon that represented the travel destination (see Figure 1). The selection cues employed a semi-transparent green controller model continuously linearly interpolated from the user's current hand position to the dialog option or object to select (see Figure 2). The manipulation cues used a semi-transparent green version of the object being manipulated and continuously linearly interpolated it from the object's current position to its desired position (see Figure 3). In the retention session, the VR training application did not provide these interaction cues, unless the participant committed

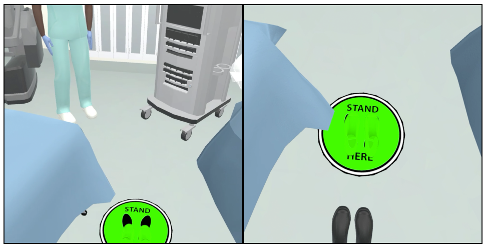

Figure 1: Example travel cues used in the application.

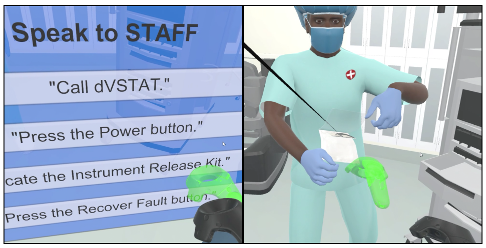

Figure 2: Example selection cues used in the application.

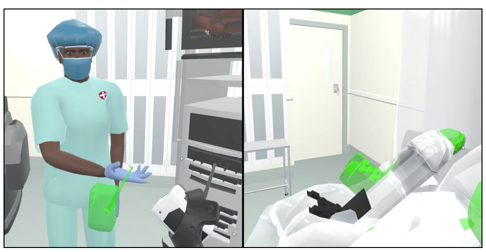

Figure 3: Example manipulation cues used in the application.

an error or was inactive for 30 seconds (i.e., a timeout). For both versions of the application, whenever an error or timeout occurred, a virtual agent would verbally instruct the participant what to do and the interaction cue for the step would be presented.

3.5 Materials

We used an HTC Vive system, including the head-mounted display (HMD) and both handheld controllers, to run the VR training application. The Vive HMD provided a 110° diagonal field of view with a display resolution of 1080 x 1200 pixels per eye and a 90 Hz refresh rate. The HMD was retrofitted with a Vive audio strap. The VR training application was developed using Unity and maintained framerates of 90 frames per second to match the Vive's refresh rate. The SteamVR plugin for Unity was used to process the Vive's input data.

3.6 Procedure

The following procedure was reviewed and approved by the University of Texas at Dallas Institutional Review Board.

The study consisted of two sessions for each participant: the learning session and the retention session. The learning session lasted approximately 60 minutes. The retention session occurred one week later and lasted approximately 30 minutes.

After informed consent, each participant would be randomly assigned to one of the three timing conditions (0, 8, or 16 seconds). The learning session began with a background survey regarding demographics, education, and technology experiences. The participant would then don the HTC Vive and experience the SteamVR tutorial to learn how to use the Vive. The participant would then experience the VR training application with interaction cues corresponding to their assigned timing condition. After successfully completing the VR training application, the participant would complete questionnaires regarding various aspects of the user experience.

The retention session took place one week after the learning session, restricted to the same day of the week to avoid potential confounds. The participant would experience the same VR training application but without interaction cues. After successfully completing the application, the participant would be given a free-response exit survey and compensated $15 USD.

3.7 Participants

A total of 67 participants (14 females, 53 males) were recruited through university mailing lists for this study. However, only 60 participants (11 females, 49 males) completed both sessions of the study. As an exclusion criterion, none of the participants had prior experience or knowledge of surgical robots. The overall average age was 22.6 ± 4.2 years, within a range of 18 to 45 years. Based on self-reported background data, 46 participants played video games on a regular basis (i.e., at least one hour per week) and 34 participants had prior VR experiences. Each participant was randomly assigned to one of the three timing conditions (0, 8, or 16 seconds). Table 2 shows the demographics of the participants.

3.8 Results

For each metric, we conducted a three-way (session, purpose, timing), mixed (session and purpose within, timing between), repeated measures analysis of variance (RM-ANOVA) at a 95% confidence level. The Shapiro-Wilk test of normality was used to ensure results were approximately normally distributed. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test of sphericity indicated that the assumption of sphericity had been violated. Bonferroni corrections and post hoc tests were used to correct for Type I errors in the repeated measures and to identify significantly different conditions when a significant main effect was found.

3.8.1 Mean Task Time. We did not find a significant three-way interaction among the session, purpose, and timing on mean task time, $F(3.666, 78.822) = 1.518$, $p = 0.209$.

We did find a significant two-way interaction between the session and purpose, $F(1.833, 78.822) = 3.337$, $p = 0.045$, $\eta^2 = 0.072$. Post hoc tests indicated that the learning task times were significantly faster than the retention task times for all three purposes ($p < 0.001$). In the learning session, the travel task times were significantly faster than the selection ($p < 0.001$) and manipulation ($p < 0.001$) times, but the selection and manipulation times were not significantly different ($p = 1.000$). In the retention session, the manipulation task times were significant slower than the travel (p

Table 2: The demographics of the participants across the three, between-subjects, interaction cue timing conditions.

Demographic	0 Seconds	8 Seconds	16 Seconds
Age	23.6 ± 5.8	21.9 ± 3.0	22.4 ± 3.2
Gender	4 Females	5 Females	2 Females
	16 Males	15 Males	18 Males
Video Games	18 Yes	12 Yes	16 Yes
	2 No	8 No	4 No
VR Experience	12 Yes	14 Yes	8 Yes
	8 No	6 No	12 No

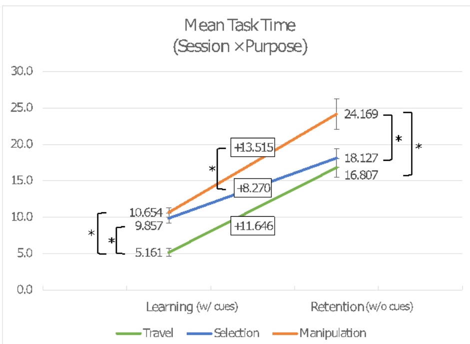

Figure 4: Mean task time with standard error bars. Asterisks indicate significantly different conditions and interactions.

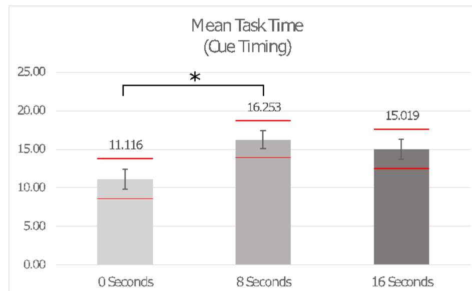

Figure 5: Mean task time with standard error bars. Red lines represent 95% confidence intervals. Asterisks indicate significantly different timing conditions.

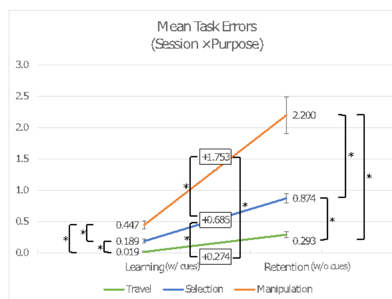

Figure 6: Mean errors with standard error bars. Asterisks indicate significantly different conditions and interactions.

= 0.007) and selection (p = 0.033) times, but the travel and selection times were not significantly different (p = 1.000). See Figure 4 for a depiction of the session and purpose interactions for mean time.

To further investigate the significant interaction between the session and purpose, we conducted a one-way (purpose) RM-ANOVA at a 95% confidence level on the slopes of the task times between the two sessions (i.e., retention task time minus learning task time) and found a significant main effect of purpose, F(1.819, 80.054) = 4.391, p = 0.018, η^2 = 0.091. Post hoc tests indicated that selection times increased significantly less from learning to retention than manipulation times (p = 0.038), and there was a statistical trend that selection times increased less than travel times (p = 0.087). There was not a significant difference between the increases in travel times and manipulation times (p = 0.973).

We did not find significant two-way interactions between session and timing, F(2, 43) = 0.296, p = 0.746, or between purpose and timing, (3.934, 82.426) = 0.495, p = 0.732.

We did find a significant main effect for timing, F(2, 43) = 4.516, p = 0.017, η^2 = 0.174. Post hoc tests indicated that the 0-second timing condition was significantly faster than the 8-second condition (p = 0.017) overall. The 16-second condition was not significantly different from the 0-second (p = 0.110) or 8-second (p = 1.000) conditions. See Figure 5 for a comparison of the timing conditions with regard to task times.

3.8.2 Mean Task Errors. We did not find a significant three-way interaction among the session, purpose, and timing on mean task errors, F(2.273, 55.681) = 0.321, p = 0.754.

We did find a significant two-way interaction between the session and purpose, F(1.136, 55.681) = 19.928, p < 0.001, η^2 = 0.289. Post hoc tests indicated that significantly more errors occurred during the retention session than the learning session for all three purposes (p < 0.001). In the learning session, there were significantly fewer travel errors than selection errors (p < 0.001) and manipulation errors (p < 0.001), and there were significantly fewer selection errors than manipulation errors (p = 0.005). In the retention session, there were also significantly fewer travel errors than selection errors (p < 0.001) and manipulation errors (p < 0.001), and again, there were significantly fewer selection errors than manipulation errors (p = 0.001). See Figure 6 for a depiction of the session and purpose interactions for task errors.

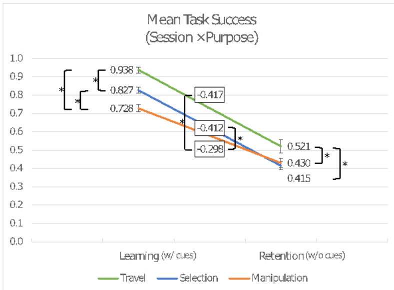

Figure 7: Mean success with standard error bars. Asterisks indicate significantly different conditions and interactions.

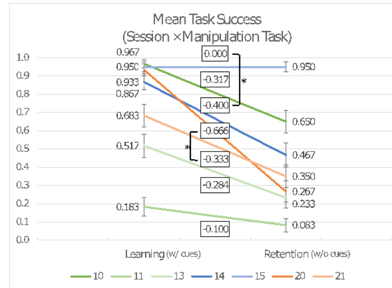

Figure 8: Mean success with standard error bars. Asterisks indicate significantly different interactions.

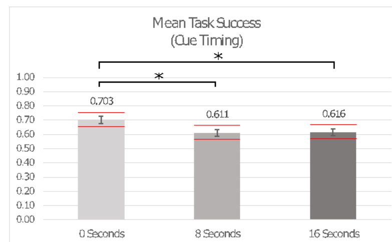

Figure 9: Mean task success with standard error bars. Red lines represent 95% confidence intervals. Asterisks indicate significantly different timing conditions.

To further investigate the significant interaction between the session and purpose, we conducted a one-way (purpose) RM-ANOVA at a 95% confidence level on the slopes of the task errors between the two sessions (i.e., retention task errors minus learning task errors) and found a significant main effect of purpose, $F(1.125, 56.231) = 20.238$, $p < 0.001$, $\eta^2 = 0.288$. Post hoc tests indicated that travel errors increased significantly less from learning to retention than selection errors ($p < 0.001$) and manipulation errors ($p < 0.001$), and selection errors increased significantly less than manipulation errors ($p = 0.001$).

We did not find significant two-way interactions between session and timing, $F(2, 49) = 0.027$, $p = 0.973$, or between purpose and timing, $F(2.276, 55.761) = 0.919$, $p = 0.416$. We also did not find a significant main effect for timing, $F(2, 49) = 1.673$, $p = 0.198$.

3.8.3 Mean Task Success. We did not find a significant three-way interaction among the session, purpose, and timing on mean task success, $F(3.560, 99.691) = 0.886$, $p = 0.466$.

We did find a significant two-way interaction between the session and purpose, $F(1.780, 99.691) = 6.204$, $p = 0.004$, $\eta^2 = 0.100$. Post hoc tests indicated that significantly more learning tasks were successfully completed than retention tasks for all three purposes ($p < 0.001$). In the learning session, significantly more travel tasks were successfully completed than selection ($p < 0.001$) or manipulation ($p < 0.001$) tasks, and significantly more selection tasks were successfully completed than manipulation tasks ($p < 0.001$). In the retention session, significantly more travel tasks were successfully completed than selection ($p = 0.025$) and manipulation ($p = 0.011$) tasks, but there were no significant differences between selection and manipulation tasks ($p = 1.000$). See Figure 7 for a depiction of the session and purpose interactions for task success.

To better understand the significantly better retention of manipulation tasks, we conducted a post-hoc two-way (session, manipulation task) RM-ANOVA at a 95% confidence level and found a significant two-way interaction between the session and manipulation task, $F(5.077, 299.560) = 10.181$, $p < 0.001$, $\eta^2 = 0.147$. Post hoc tests indicated several pairs of significantly different tasks and interactions. Most notably, the post hoc tests indicated that task 15 was retained significantly better than task 14 ($p = 0.001$) and that task 21 was retained significantly better than task 20 ($p = 0.021$). These particular pairs of significantly different interactions are notable because tasks 14 and 15 involve removing the faulted instrument (task 14) and handing it to one of the virtual agents (task 15), and tasks 20 and 21 involve collapsing the robotic arm (task 20) and stowing it away (task 21). In other words, subsequent manipulation tasks involving the same object appear to be recalled better than the preceding tasks initially involving the objects. A similar interaction effect can be seen between tasks 10 and 11, which both involve using the same tool. Note, task 13 involves that same tool, but is preceded by task 12, which is a travel task. See Figure 8 for these interactions.

We did not find significant two-way interactions between session and timing, $F(2, 56) = 1.227$, $p = 0.301$, and between purpose and timing, $F(3.310, 92.689) = 1.314$, $p = 0.274$.

We did find a significant main effect for timing, $F(2, 56) = 4.348$, $p = 0.018$, $\eta^2 = 0.134$. Post hoc tests indicate that the 0-second timing condition was significantly better than the 8-second ($p = 0.033$) and the 16-second ($p = 0.046$) conditions. The 8-second and 16-second conditions were not significantly different ($p = 1.000$). See Figure 9 for a comparison of the conditions.

4 DISCUSSION

4.1 Easier Travel Tasks

Our results indicate that, for our VR training application, travel tasks were easier to complete than selection tasks, which were easier to complete than manipulation tasks. Participants completed the travel tasks significantly faster and with significantly fewer errors than the selection and manipulation tasks during the learning and retention sessions. In turn, participants completed the travel tasks with significantly greater success (i.e., without committing errors or timing out) than the selection and manipulation tasks during both sessions. These results support our H1 hypothesis that the interaction cues for travel would afford significantly better performance of their corresponding procedural steps than the cues for selection and manipulation.

4.1.1 Precision and Degrees of Freedom. Another explanation for our results is the complexity of the tasks employed within our VR training application. Our travel tasks were essentially low-precision, two-degree-of-freedom (2-DOF) tasks, requiring the user to move to approximate positions within the horizontal plane. Many of our selection tasks were low-precision, 4-DOF tasks, requiring the user to select one large dialog option from a set of four available options (see Figure 2). Most of our manipulation tasks were high-precision, 6-DOF tasks, requiring the user to position and orient an object within a confined volume. Hence, the travel tasks were clearly less complex than the selection and manipulation tasks for our specific application. However, it is important to note that in other applications, travel might be worse than selection due to task difficulty. For example, training applications that involve complex travel tasks, such as military maneuvers, would likely result in travel tasks that are more difficult than their selection and manipulation counter parts. A systematic investigation of the precision and DOF of these types of tasks would provide a greater understanding of this issue and could possibly yield a predictive model of task performance.

4.1.2 Egocentric versus Exocentric. There are a few potential explanations for why the travel tasks were easier to complete than the selection and manipulation tasks. For one, travel tasks are generally regarded as egocentric tasks while selection and manipulation are considered exocentric tasks [38], and egocentric tasks have been demonstrated to improve performance [40] and reduce cognitive workload [17], compared to exocentric counterpart tasks. Hence, by their nature, egocentric travel tasks may be easier to complete than exocentric selection and manipulation tasks. However, more research and evidence are necessary to make such a claim and to better understand the nuances of egocentric and exocentric tasks in VR.

4.1.3 Varying Markedness. Another potential explanation, though much less likely, is the fact that the travel cues used continuously linearly interpolated 2D icons while the selection and manipulation cues used continuously linearly interpolated semi-transparent 3D objects. It is possible that the 2D icons were easier to understand than the 3D objects. However, we doubt that this minor confound is the reason for the significant differences among travel, selection, and manipulation. Additional research could be conducted to use semi-transparent 3D avatars for the travel cues, in order to eliminate the minor confound.

4.2 Better Retention for Manipulation Tasks

Our results indicate that participants retained manipulation tasks better than travel or selection tasks. While task success decreased from learning to retention for all three types of tasks, the decrease in manipulation task success was significantly less than the decreases in travel success and selection success. These results do not support our H2 hypothesis that the travel interaction cues would afford significantly better retention of the training steps.

4.2.1 Object-Dependent Memory. One potential explanation for the better retention of manipulation tasks is that participants leveraged an object-oriented form of context-dependent memory. Context-dependent memory refers to the improved retrieval of information when the context of retrieval matches the context in which the information was encoded [34]. These contexts are often considered to be environmental in nature [22], such as the environmental odors present during encoding [15]. However, based on a post-hoc analysis of our manipulation tasks, we believe that the target of a manipulation task can serve as the memory context (i.e., "object-dependent memory").

More specifically, our post-hoc analysis revealed that for each manipulation target (the tool, the instrument, and the robotic arm, respectively) the initial manipulation task was less successful than subsequent manipulation tasks involving the same target. We believe that, during learning, manipulation tasks with the same targets are likely to be encoded within the same memory region and sub-network of neurons. Hence, during the retention session, once a participant starts manipulating a target (often after receiving an interaction cue due to an error or timeout), the same memory region and sub-network of neurons will be activated, making it easier to recall and execute subsequent manipulations involving the same target.

4.2.2 Training Implications. Many VR training applications combine travel, selection, and manipulation tasks. As such, the ratio of these three types of tasks may have an impact on the overall usability and training effectiveness for each application. Our results suggest that manipulation-heavy training applications, such as surgical simulators, may yield better retention than selection-heavy training applications, such as assembly-based applications. While further research is needed to confirm this hypothesis, if supported by future results, it would serve as a guideline for choosing what types of training would be best supported by VR applications.

4.3 Faster and Better Immediate Cues

Our results indicate that the 0-second cue timing condition afforded significantly faster learning and better retention than the delayed 8- and 16-second conditions. Participants in the 0-second condition completed their tasks significantly faster than those participants in the 8-second condition. This result partially supports our H3 hypothesis that immediate interaction cues would afford significantly faster learning. More importantly, participants in the 0-second condition were significantly more successful during the retention assessment than participants in either the 8- or 16-second conditions. This result does not support our H4 hypothesis that there would be no significant differences among the timing conditions in terms of retention.

Another interesting result was that participants took less time on average to complete the 16-second condition than the 8-second condition. In fact, the mean task time for the 16-second condition was 15.019 seconds (i.e., less time than when the interaction cue was presented). Based on our observations, this interesting result was due to participants in the 16-second condition attempting to complete the task without guidance, which would often result in an error and the activation of the verbal instructions and interaction cue for the step, as described in section 3.3. On the other hand, we observed that most participants in the 8-second condition would simply wait for the interaction cue to be presented before attempting to complete the step.

4.3.1 Comparison to Prior Results. As described in section 2.2, Hu et al. [16] did not find any significant differences among their three timing conditions with regard to a knowledge test administered immediately after the VR training. This was the basis for our H4 hypothesis. However, there are a few explanations for why we found significant differences in terms of retention while they did not. For one, participants are more likely to remember the relevant information immediately after their VR experience, when Hu et al. [16] administered their knowledge tests, regardless of their assigned timing condition, based on similar results for other VR recall studies [35, 36]. Furthermore, the effectiveness of using a knowledge test as a validated instrument is dependent upon the reliability of the test [21], which is dependent upon the construction of its questions [2]. Hence, the knowledge test used by Hu et al. [16] may not have been a useful instrument for finding significant differences between the three timing conditions. Finally, instead of having participants demonstrate their retention through a declarative knowledge test, we had participants to demonstrate their procedural knowledge by executing the same tasks that they performed during learning. Hence, we believe the results of our study are a truer indication of the effects of interaction cue timing.

4.3.2 Cue Timing Design Guidelines. Based on our results and the results of Hu et al. [16], we recommend that developers should generally avoid delayed interaction cues and focus on providing interaction cues as soon as they are relevant. By doing so, users should be able to complete their tasks faster and are more likely to retain task knowledge.

4.4 Limitations

There are limitations to the generalizability of our results.

4.4.1 Application Specific. It is important to note that our results are based on the use of a specific, previously developed, VR training application, which introduced its own potential confounds. First, as described above, the travel cues employed 2D icons while the selection and manipulation cues employed 3D shapes, which may have been more difficult to understand. Another limitation of our VR application is that it did not require an equal number of travel, selection, and manipulation tasks. Instead, it required 4 travel tasks, 13 selection tasks, and 7 manipulation tasks. The low number of travel tasks could be another explanation for why the travel tasks were significantly easier to complete.

While our results are application specific, our study is a good starting point for this line of research into the effects of interaction cue purpose and timing. We found significant differences for both aspects of interaction cues, which may indicate how these aspects will affect other applications. In addition to re-conducting the current work with another ecologically valid VR training application to see if the same results are obtained, we believe it is important that controlled studies are conducted to systematically investigate the effects and interactions of each aspect and to eliminate any confounding variables.

4.4.2 Limited Participant Sample. While we successfully conducted 60 total participants through our two-session, week-long study, our participant sample was limited. First, our sample lacked an equal representation of females and across all three between-subject conditions. Peck et al. [30] have recently found that changes in simulator sickness in VR studies are associated with the proportion of female participants. If simulator sickness systematically varies with the proportion of female participants, it is possible that learning and retaining VR training do as well. Another limitation of our sample is the disproportionate number of participants with prior VR experiences across the three conditions. In particular, the 16-second condition only had 8 participants with prior VR experience, while the 0- and 8-second conditions respectively had 12 and 14 participants with prior VR experience. However, we do not believe that the low number of participants in the 16-second condition had an influence on our results, as the majority of significant differences among the three conditions were between the 0- and 8-second conditions. The overall number of participants was also a limitation of our sample. While we found several significant differences, a larger sample size may have yielded more insights.

5 CONCLUSION AND FUTURE WORK

We conducted a mixed-design user study investigating the effects of the purpose and timing of interaction cues on learning and retaining VR training. We investigated three interaction cue purposes (travel, selection, and manipulation) and three interaction cue timings (0, 8, and 16 seconds). Our results indicate that the travel tasks in our VR training application were significantly easier to complete than the selection and manipulation tasks. However, our results also indicate that participants retained the manipulation tasks significantly better than the travel and selection tasks. We contribute this result to an object-oriented type of context-dependent memory. Our results also indicate that the immediate interaction cues afforded significantly faster learning and better retention than either cue delay. Based on this result, we recommend that developers avoid delayed interaction cues and focus on presenting immediate cues.

For future work, we plan to begin controlled, systematic investigations of interaction cue purpose. We specifically plan to study how the required precision and DOF of a task affects the effectiveness of its corresponding interaction cue and the retention of the task itself.

ACKNOWLEDGMENTS

This material is based on work partially supported by the National Science Foundation under Grant No. 2021607 – "CAREER: Leveraging the Virtualness of Virtual Reality for More-Effective Training."

REFERENCES

[1] Shaowen Bardzell. 2008. Systems of Signs and Affordances: Interaction Cues in 3D Games. *Extending Experiences. Structure, analysis and design of computer game player experience*: 191–209.

[2] Jared Breakall, Christopher Randles, and Roy Tasker. 2019. Development and use of a multiple-choice item writing flaws evaluation instrument in the context of general chemistry. *Chemistry Education Research and Practice* 20, 2: 369–382.

[3] Fabio Buttussi and Luca Chittaro. 2008. MOPET: A context-aware and user-adaptive wearable system for fitness training. *Artificial Intelligence in Medicine* 42, 2: 153–163.

[4] Jacky C.P. Chan, Howard Leung, Jeff K.T. Tang, and Taku Komura. 2011. A Virtual Reality Dance Training System Using Motion Capture Technology. *IEEE*

Transactions on Learning Technologies 4, 2: 187–195.

[5] Kwangsu Cho, Ju-Hwan Lee, Byeong-Taek Lee, and Eunil Park. 2015. Effects of Feedforward in In-Air Remote Pointing. *International Journal of Human–Computer Interaction* 31, 2: 89–100.

[6] Sven Coppers, Kris Luyten, Davy Vanacken, David Navarre, Philippe Palanque, and Christine Gris. 2019. Fortunettes: Feedforward about the future state of GUI widgets. *Proceedings of the ACM on Human-Computer Interaction* 3, EICS.

[7] William Delamare, Thomas Janssoone, Céline Coutrix, and Laurence Nigay. 2016. Designing 3D Gesture Guidance: Visual Feedback and Feedforward Design Options. *Proceedings of the International Working Conference on Advanced Visual Interfaces*, Association for Computing Machinery, 152–159.

[8] Kody R Dillman, Terrance Tin Hoi Mok, Anthony Tang, Lora Oehlberg, and Alex Mitchell. 2018. A Visual Interaction Cue Framework from Video Game Environments for Augmented Reality. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 140:1–140:12.

[9] Tom Djajadiningrat, Kees Overbeeke, and Stephan Wensveen. 2002. But How, Donald, Tell Us How?: On the Creation of Meaning in Interaction Design Through Feedforward and Inherent Feedback. *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, ACM, 285–291.

[10] Olov Engwall. 2012. Analysis of and feedback on phonetic features in pronunciation training with a virtual teacher. *Computer Assisted Language Learning* 25, 1: 37–64.

[11] Chris Galloway. 2017. Blink and they're gone: PR and the battle for attention. *Public Relations Review* 43, 5: 969–977.

[12] Rex Hartson. 2003. Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology* 22, 5: 315–338.

[13] Rex Hartson and Pardha S Pyla. 2012. *The UX Book: Process and guidelines for ensuring a quality user experience*. Elsevier.

[14] Luis Hernandez, Javier Taibo, Antonio Seoane, Ruben Lopez, and Rocio Lopez. 2004. The experience of the Empty Museum: displaying cultural contents on an immersive, walkable VR room. *Proceedings Computer Graphics International, 2004.*, 436–443.

[15] Michael J. Howell, Nicolas S. Herrera, Alec G. Moore, and Ryan P. McMahan. 2016. A reproducible olfactory display for exploring olfaction in immersive media experiences. *Multimedia Tools and Applications* 75, 20: 12311–12330.

[16] Xinyu Hu, Alec G. Moore, James Coleman Eubanks, Afham Ahmed Aiyaz, and Ryan P. McMahan. 2020. The Effects of Delayed Interaction Cues in Virtual Reality Training.*2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, IEEE, 63–69.

[17] Rajiv Khadka and Amy Banić. 2020. Effects of Egocentric Versus Exocentric Virtual Object Storage Technique on Cognition in Virtual Environments.*2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 205–209.

[18] Bui Minh Khuong, Kiyoshi Kiyokawa, Andrew Miller, Joseph J. LaViola Jr., Tomohiro Mashita, and Haruo Takemura. 2014. The effectiveness of an AR-based context-aware assembly support system in object assembly.*2014 IEEE Virtual Reality (VR)*, 57–62.

[19] Chengyuan Lai, Ryan P. McMahan, Midori Kitagawa, and Iolani Connolly. 2016. Geometry Explorer: Facilitating Geometry Education with Virtual Reality. *International Conference on Virtual, Augmented and Mixed Reality (VAMR)*, Springer, 702–713.

[20] Joseph J. LaViola Jr., Ernst Kruijff, Ryan P. McMahan, Doug Bowman, and Ivan Poupyrev. 2017.*3D User Interfaces: Theory and Practice (2nd ed.)*. Addison-Wesley Professional, Boston.

[21] Ximena López, Jorge Valenzuela, Miguel Nussbaum, and Chin-Chung Tsai. 2015. Some recommendations for the reporting of quantitative studies. *Computers & Education* 91: 106–110.

[22] Ryan P. McMahan and Nicolas S. Herrera. 2016. AFFECT: Altered-Fidelity Framework for Enhancing Cognition and Training. *Frontiers in ICT* 3: 29:1–15.

[23] Ryan P. McMahan, Regis Kopper, and Doug A. Bowman. 2014. Principles for Designing Effective 3D Interaction Techniques. In K.S. Hale and K.M. Stanney, eds., *Handbook of Virtual Environments*. CRC Press, Boca Raton, 299–325.

[24] Alec G. Moore, Xinyu Hu, James Coleman Eubanks, Afham Ahmed Aiyaz, and Ryan P. McMahan. 2020. A Formative Evaluation Methodology for VR Training Simulations.*2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, IEEE, 125–132.

[25] Alec G. Moore, Marwan Kodeih, Anoushka Singhania, Angelina Wu, Tassneen Bashir, and Ryan P. McMahan. 2019. The Importance of Intersection Disambiguation for Virtual Hand Techniques.*2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, IEEE, 450–457.

[26] Maria Murcia-López and Anthony Steed. 2018. A Comparison of Virtual and Physical Training Transfer of Bimanual Assembly Tasks. *IEEE Transactions on Visualization and Computer Graphics* 24, 4: 1574–1583.

[27] Donald A Norman. 1999. Affordance, Conventions, and Design. *interactions* 6, 3: 38–43.

[28] Néstor Ordaz, David Romero, Dominic Gorecky, and Héctor R Siller. 2015. Serious Games and Virtual Simulator for Automotive Manufacturing Education & Training. *Procedia Computer Science* 75: 267–274.

[29] Fred Paas, Alexander Renkl, and John Sweller. 2003. Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist* 38, 1: 1–4.

[30] Tabitha C. Peck, Laura E. Sockol, and Sarah M. Hancock. 2020. Mind the Gap: The Underrepresentation of Female Participants and Authors in Virtual Reality Research. *IEEE Transactions on Visualization and Computer Graphics* 26, 5: 1945–1954.

[31] Philippe Renevier, Laurence Nigay, J Bouchet, and L Pasqualetti. 2005. Generic Interaction Techniques for Mobile Collaborative Mixed Systems BT - Computer-Aided Design of User Interfaces IV. Springer Netherlands, 309–322.

[32] Michael Rohs and Philipp Zweifel. 2005. A Conceptual Framework for Camera Phone-Based Interaction Techniques BT - Pervasive Computing. Springer Berlin Heidelberg, 171–189.

[33] Rebecca Polley Sanchez, Chelsea M Bartel, Emily Brown, and Melissa DeRosier. 2014. The acceptability and efficacy of an intelligent social tutoring system. *Computers & Education* 78: 321–332.

[34] Steven M Smith and Edward Vela. 2001. Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review* 8, 2: 203–220.

[35] Evan A Suma, Samantha L Finkelstein, Seth Clark, Paula Goolkasian, and Larry F Hodges. 2010. Effects of travel technique and gender on a divided attention task in a virtual environment.*3DUI 2010 - IEEE Symposium on 3D User Interfaces 2010, Proceedings*, 27–34.

[36] Evan Suma, Samantha Finkelstein, Myra Reid, Sabarish Babu, Amy Ulinski, and Larry F. Hodges. 2010. Evaluation of the cognitive effects of travel technique in complex real and virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 16, 4: 690–702.

[37] F. Tyndiuk, V. Lespinet-Najib, G. Thomas, and C. Schlick. 2004. Impact of Large Displays on Virtual Reality Task Performance. *Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, Association for Computing Machinery, 61–65.

[38] F. Tyndiuk, G. Thomas, V. Lespinet-Najib, and C. Schlick. 2005. Cognitive Comparison of 3D Interaction in Front of Large vs. Small Displays. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, Association for Computing Machinery, 117–123.

[39] Jo Vermeulen, Kris Luyten, Elise van den Hoven, and Karin Coninx. 2013. Crossing the Bridge over Norman's Gulf of Execution: Revealing Feedforward's True Identity. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1931–1940.

[40] Jia Wang and Robert W. Lindeman. 2015. Object impersonation: Towards effective interaction in tablet- and HMD-based hybrid virtual environments.*2015 IEEE Virtual Reality (VR)*, 111–118.

[41] S. A. G. Wensveen, J. P. Djajadiningrat, and C. J. Overbeeke. 2004. Interaction Frogger: A Design Framework to Couple Action and Function through Feedback and Feedforward. *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, Association for Computing Machinery, 177–184.

[42] Andrew Yoshimura, Adil Khokhar, and Christoph W. Borst. 2019. Eye-gaze-triggered Visual Cues to Restore Attention in Educational VR.*2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1255–1256.