The Effects of Delayed Interaction Cues in Virtual Reality Training

Xinyu Hu[1], Alec G. Moore[1], James Coleman Eubanks[2], Afham Ahmed Aiyaz[2], Ryan P. McMahan[1]

[1]University of Central Florida, Orlando, FL, USA

[2]University of Texas at Dallas, Richardson, TX, USA

Abstract

Interaction cues are commonly used in games and educational systems to direct users to take appropriate actions at particular moments, such as pressing a specific button to start a conversation with a virtual agent. Based on a literature review and a related framework, we identified delayed interaction cues as an under-explored research area. We present a controlled evaluation of the effects of delayed cues in a virtual reality training application. Our results indicate that immediate interaction cues are significantly more efficient than delayed cues. We discuss the implications of our results on the design of interaction cues for future games and educational systems.

Keywords: Interaction cues; timing; virtual reality; training.

Index Terms: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality; User Centered Design

1 Introduction

Games and educational systems commonly inform users to perform specific interactions at particular times. For example, many virtual reality (VR) experiences, such as *The Lab* by Valve, visually and verbally prompt users to use the handheld controllers and buttons to learn how to teleport to travel. Recently, Dillman et al. [1] have referred to these types of communication as "interaction cues".

The term "interaction cue" has been sparsely used in the literature with regard to human-computer interactions (note, most uses within the literature pertain to social interaction cues between social agents). In their research on video games, Dillman et al. [1] describe interaction cues as "used to communicate information about the game world to players". Similarly, Bardzell [2] defines an interaction cue as "a sign that provides information to players so that they are aware of their present interactive possibilities". In contrast, Rohs and Zweifel [3] have explained that "interaction cues guide users in the interaction process".

To better understand interaction cues, we conducted a literature review in the field of games and educational systems. We particularly focused on including publications involving 3D interactions, augmented reality (AR), mixed reality (MR), and VR, as these types of interactive systems often include actions different from traditional user interfaces (UIs). Based on this literature review, we have identified delayed interaction cues as an underexplored area compared to immediate interaction cues.

To better inform this area, we conducted a controlled evaluation of the effects of interaction cue timing comparing immediate cues (0 seconds) to two types of delayed cues (8 and 16 seconds). We hypothesized that the 8-second condition would result in the most effective knowledge acquisition because it would encourage more active exploration than the 0-second condition while also avoiding potential frustrations caused by the 16-second condition. However, the results of our study indicate that there are little, if any, significant differences among the conditions with regard to knowledge acquisition. Our study did provide evidence that immediate interaction cues are significantly more efficient in terms of time and errors than delayed cues. We discuss how these results suggest that immediate interaction cues should be generally used instead of delayed interaction cues.

[1] {XinyuHu, agm}@knights.ucf.edu, rpm@ucf.edu
[2] {j.coleman.eubanks, axa157730}@utdallas.edu

2 Related Work

Dillman et al. [1] developed a descriptive framework of visual interaction cues, based on the analysis of 49 contemporary video games. Their visual interaction cue framework consists of three dimensions: purpose, markedness, and trigger. In the intelligent tutoring systems (ITS) domain, we found that the timing of interaction cues is another concern, as students have been found to take advantage of cues to proceed through content with little effort [4]. Hence, below we discuss and provide examples for four cue dimensions: purpose, markedness, trigger, and timing.

2.1.1 Purpose

As discussed in the introduction, an interactions cue guides the user's goal or task with an action to take. That task is the purpose of the interaction cue. As Dillman et al. [1] noted, "interaction cues are purposely designed".

Dillman et al. [1] defined three original types of purpose based on their analysis of video games: discover, look, and go. *Discover* cues inform the user about objects or areas to interact with. In their MR collaborative game called *TROC*, Renevier et al. [5] used 3D localized sounds and a mini-map to inform users about digital objects to collect, in addition to the locations of other players. *Look* cues are similar but inform the user to focus their visual attention on a particular detail, object, group of objects, or area. For example, Buttussi and Chittaro [6] used a virtual agent to verbally invite the user to look at a PDA display in their *MOPET* fitness training system. ***Go*** cues inform the user to move to a new destination. Hernandez et al. [7] used bright cylinders to represent target destinations in *The Art Gallery*, their VR system for experiencing artworks.

Based on our literature review, we have also identified four additional purposes for interaction cues: pick, manipulate, gesture, and speak. *Pick* cues inform the user to select a particular object or element. For example, Webster et al. [8] used a virtual image and textual description to inform users what the next correct piece was in their AR architectural construction system. *Manipulate* cues inform the user to move an object to a new position. In their AR assembly training system, Khuong et al. [9] explored two different interaction cues to convey where to position the next assembly piece: a) an in-place wireframe visualization of the target position on the real-world assembly, and b) a side visualization of the current assembly with the object in its target

978-1-7281-6532-5/20/$31.00 ©2020 IEEE
DOI 10.1109/VRW50115.2020.0-257

position. *Gesture* cues inform the user to perform specific body movements. For example, Chan et al. [10] used a virtual instructor to convey what body movements to perform for a VR dance training system. *Speak* cues inform the user to verbally communicate. Engwall [11] used a virtual articulation instructor called *ARTUR* and an augmented view of the instructor's tongue positions and movements to inform users how to verbally communicate specific words In another Speak cue example, Robb et al. [12] used a virtual surgical technician to prompt the user to speak up to the virtual surgeon about patient safety issues.

2.1.2 Markedness

Dillman et al. [1] described *markedness* as "the extent to which the cue blends into the game environment (or how it stands out from that environment)". They also defined four degrees of markedness for visual cues: subtle, emphasizing, integrated, or overlaid. We have adopted three of these degrees to address all of our modalities, in addition to visual interaction cues.

Subtle interaction cues blend into the virtual environment and "are difficult to distinguish from the environment itself" [1]. For example, Dinh et al. [13] used a coffee scent when the user was near a coffeemaker as an olfactory-based discover cue. *Emphasized* cues "highlight an existing object or surface" in the environment [1]. Visually highlighting objects [14] and de-emphasizing irrelevant environmental aspects, such as silencing police sirens [15], are examples of emphasized interaction cues. *Integrated* cues involve adding an object or sensory stimulus to the virtual environment. Text instructions [16], narrator instructions [17], 3D arrows [18], abstract sounds [5], and haptic guidance [19] are all integrated interaction cues. *Overlaid* cues, such as UI elements and mini-maps, "explicitly distinguish two different aspects of the player's viewport" [1].

2.1.3 Trigger

A cue's *trigger* is a predefined event or criteria that brings about the presentation of the interaction cue. Dillman et al. [1] defined four types of triggers in their visual interaction cue framework: player, context, other/agent, and persistent.

User-triggered (or player-triggered) interaction cues are "activated by an explicit action by the player" or user [1]. Allowing the user to control the progression of a virtual information tour by pressing a "Next" button is an example of a user-triggered interaction cue [20]. *Context-triggered* interaction cues "are activated by the player through implicit actions" and often associated with the user's location [1]. For example, the previously mentioned olfactory-based discover cue was activated by the player entering the virtual room with the coffeemaker [13]. *Agent-triggered* cues are activated by an agent other than the user. This agent could be another user in a multi-user system, such as a collaborative life support training simulator [21], or an automated agent within the virtual environment, such as the previously mentioned virtual surgical technician [12]. *Persistent* cues are always presented [1]. UI elements, such as an informative cursor that indicates the user should take additional actions [22], commonly serve as persistent cues.

2.1.4 Timing

As aforementioned, we have identified a fourth dimension of interaction cues based on when they are presented. *Immediate* interaction cues are presented the instant that they are triggered. The majority of interaction cues within the literature are immediate cues. *Delayed* cues, on the other hand, are purposefully presented later than when they are triggered. For example, if the player does not ask for help on a two-person task after a given amount of time, Sanchez et al. [23] have a virtual agent walk through the scene as an interaction cue to ask for help. Similarly, Baker et al. [24] developed their Latent Response Model to help with re-designing intelligent tutors to respond appropriately to students "gaming the system" to obtain correct responses.

3 Study of Delayed Interaction Cues

3.1 Motivation

Within our literature review, we only found examples of delayed interaction cues within the ITS domain (e.g., [4], [23]–[25]). However, we didn't find any direct comparisons of delayed and immediate cues. The purpose of our study was to evaluate and compare the effects of immediate and delayed interaction cues on a training application, to better understand uses of delayed cues.

3.2 Task

We used a previously developed VR application, designed to train users how to troubleshoot a surgical robot, to evaluate the effects of delayed interaction cues on the efficiency and effectiveness of the training application.

The application's training scenario involves interacting with a virtual surgeon and a virtual non-sterile staff member to troubleshoot a surgical robot with a faulted instrument arm by attempting to restart the system, using a special tool to open the instrument's jaws, removing the instrument, stowing the arm, and disabling it. To accomplish the scenario, the VR training application includes 24 steps that require various actions of the user (see Table 1). It is important to note that the application uses dialog buttons (pick tasks) for conversing with the virtual agents.

For each step, the application uses an interaction cue to guide the user through the troubleshooting scenario. Each cue employs subtle verbal instructions from one of the virtual agents for feedforward and integrated visual animations as perceived affordances (see Figure 1). All the cues are scenario-triggered.

In addition to the interaction cues, the VR training application employs verbal repercussions for mistakes. If the user makes an error or incorrect decision, one of the virtual agents will verbally inform the user of the mistake.

Step	Purpose	Step	Purpose
1	Go + Look	13	Manipulate
2	Pick (Dialog)	14	Manipulate
3	Pick (Dialog)	15	Manipulate
4	Pick (Dialog)	16	Pick (Dialog)
5	Pick (Dialog)	17	Pick (Dialog)
6	Pick (Dialog)	18	Go + Look
7	Pick (Object)	19	Pick (Object)
8	Pick (Dialog)	20	Manipulate
9	Pick (Object)	21	Manipulate
10	Manipulate	22	Pick (Dialog)
11	Manipulate	23	Go + Look
12	Look	24	Pick (Dialog)

Table 1. The purpose of each interaction cue for the 24 steps involved in the VR training application.

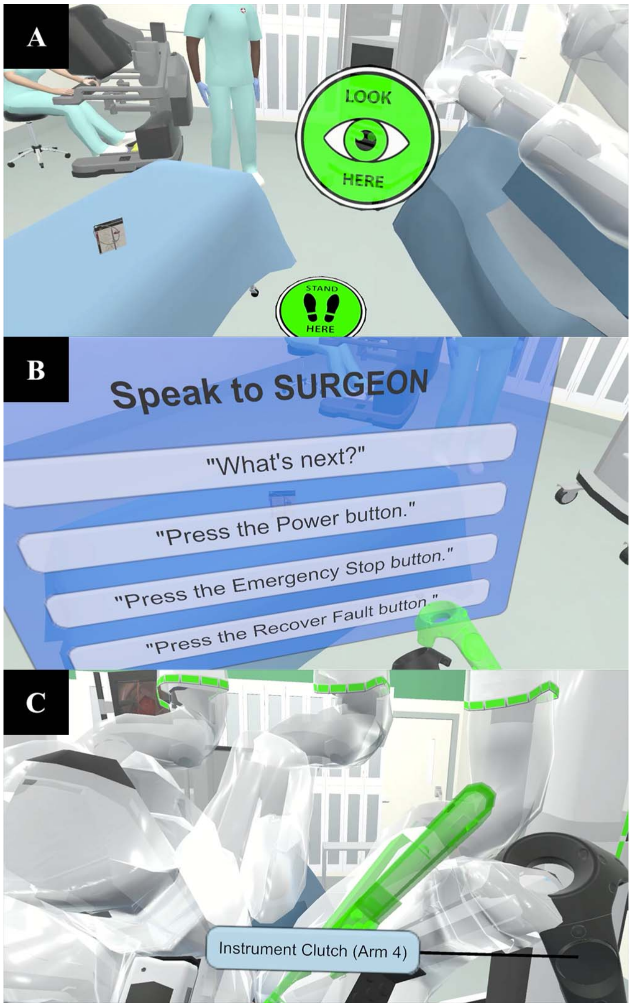

Figure 1: Examples of interaction cues from the VR training application: A) Green "Stand Here" and "Look Here" icons visually prompt Go + Look tasks. B) Green, semitransparent clones of the Vive controller are visually animated to prompt Pick tasks. C) Green, semi-transparent clones of virtual objects are visually animated to prompt Manipulate tasks.

3.3 Independent Variable

Our study's independent variable was the delay at which the VR training application's interaction cues were presented after being triggered. We controlled this delay at 0, 8, and 16 seconds, which yielded one immediate-cue condition and two delayed-cue conditions. We chose 8 seconds for the first delayed cue condition because attention spans have been reported to be as little as eight seconds [26]. Similarly, we chose 16 seconds, twice the delay, to better understand how attention span may affect cues.

3.4 Dependent Variables

As indicated above, the purpose of our study was to evaluate the effects of delayed cues on the efficiency and effectiveness of the VR training application. We were also interested in the effects of delayed cues on common subjective responses.

3.4.1 Efficiency

Efficiency is concerned with the time and resources used during training to reach mastery [27]. The VR training application was previously designed to capture several efficiency metrics, including the number of physical errors, sterility breaks, incorrect decisions, periods of inactivity, and total training time. Physical errors occur when the user incorrectly picks or manipulates a virtual object. Sterility breaks are a special case of physical errors that involve the user touching a non-sterile portion of the environment. Incorrect decisions occur when the user picks the wrong dialog option for interacting with the virtual surgeon or non-sterile staff member. Finally, periods of inactivity are recorded when the user takes no action, correct or incorrect, for 30 seconds. Note, in the VR application, periods of inactivity are addressed with the non-sterile staff member providing additional verbal instructions of what to do.

3.4.2 Effectiveness

Effectiveness is concerned with the acquisition, retention, or transfer of skills and knowledge from training to practice [27]. We designed our study to focus on knowledge acquisition based on a post-training knowledge test. To avoid lengthy procedures for participants, we did not employ a post-training VR practice (skills acquisition).

3.4.3 Subjective Responses

After the VR training experience, we administered the Simulator Sickness Questionnaire (SSQ) [28], the Spatial Presence Experience Scale (SPES) [29], the System Usability Scale (SUS) [30], and questionnaires regarding the scenario fidelity [31] and body ownership aspects [32] of the VR training experience.

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 90Hz 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 lasted approximately 60 minutes for each participant. After informed consent, each participant would be randomly assigned to one of the three delayed cue conditions (0, 8, or 16 seconds). The first session began with a background survey on participant's demographics, education, and technology experience. 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 a version of the VR training application corresponding to their assigned study condition. After successfully completing the VR training application, the participant would be administered the SSQ, SPES, SUS, and questionnaires regarding scenario fidelity and body ownership. The study would then conclude with a knowledge test consisting of 20 multiple-choice questions pertaining to the training scenario.

3.7 Research Questions

Q1. Which cue condition will be the most efficient? H1. We hypothesized that the 0-second condition (i.e., the immediate cue) would result in the most efficient training (i.e., the least training time and fewest training errors) because users would not need to wait for the interaction cues and would not make mistakes while waiting.

Q2. Which cue condition will be the most effective? H2. We hypothesized that the 8-second condition would result in the most effective training (i.e., best knowledge test scores) based on prior results suggesting that active exploration and error encouragement are effective training elements [33]. We believed that the 8-second condition would allow more active exploration and would encourage more errors than the 0-second condition. We also believed that the 16-second condition would result in losing the participant's attention, and even causing anxiety, when the participant could not find the correct action to take on their own.

Q3. Which cue condition will yield the best subjective responses? H3. We hypothesized that the delayed cue conditions (i.e., 8 and 16 seconds) would reduce presence, perceived usability, and perceived scenario fidelity due to users waiting on the cues. We hypothesized that there would not be significant differences between the conditions in terms of simulator sickness or body ownership.

3.8 Participants

A total of 60 participants (11 females, 49 males) completed 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 cue conditions. Table 2 shows the demographics of the participants for each condition.

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

Table 2. Demographics of participants across all three conditions.

3.9 Results

For each objective metric, we conducted a one-way (cue condition) analysis of variance (ANOVA) at a 95% confidence level. The Shapiro-Wilk test of normality was used to ensure results were approximately normally distributed, and Levene's test of homogeneity of variances was used to ensure the conditions had similar variabilities. For each subjective metric, we conducted a Kruskal-Wallis H test as a non-parametric alternative to an ANOVA.

3.9.1 Efficiency

We found a significant main effect of cue condition on ***total training time***, $F(2, 57) = 6.118$, $p = 0.004$, $\eta^2 = 0.177$. Bonferroni post hoc tests indicated that the 0-second condition was significantly faster than the 16-second condition ($p = 0.004$). There was a trend between the 0 and 8-second conditions ($p = 0.064$) and no significant difference between the 8 and 16-second conditions ($p = 0.895$). See Table 3 for the descriptive statistics.

We found a significant main effect on ***physical errors*** during training, $F(2, 57) = 7.057$, $p = 0.002$, $\eta^2 = 0.198$. Bonferroni post hoc tests indicated that both the 0-second condition ($p = 0.002$) and the 8-second condition ($p = 0.019$) yielded significantly fewer physical errors than the 16-second condition. However, we did not find a significant main effect on *sterility breaks* during training, $F(2, 57) = 0.445$, $p = 0.643$, $\eta^2 = 0.015$.

We found a significant main effect on ***incorrect decisions*** during training, $F(2, 57) = 3.251$, $p = 0.046$, $\eta^2 = 0.102$. Bonferroni post hoc tests indicated that the 0-second condition yielded significantly fewer incorrect decisions than the 16-second condition ($p = 0.041$). There were no significant differences between the 0 and 8-second conditions ($p = 0.801$) and the 8 and 16-second conditions ($p = 0.481$).

We did not find a significant main effect on *periods of inactivity* during training, $F(2, 57) = 1.617$, $p = 0.207$, $\eta^2 = 0.054$. See Table 3 for the descriptive statistics.

Metric	Delay	Mean	Std. Dev
Total Time	**0^A**	**634.8s**	**87.6s**
	8^{AB}	772.1s	179.5s
	16^B	**833.0s**	**247.3s**
Physical Errors	**0^A**	**4.90**	**3.01**
	8^A	**6.15**	**4.21**
	16^B	**11.15**	**8.14**
Sterility Breaks	0^A	0.30	0.47
	8^A	0.35	0.59
	16^A	0.50	0.95
Incorrect Decisions	**0^A**	**1.55**	**2.59**
	8^{AB}	2.85	4.15
	16^B	**4.50**	**4.06**
Periods of Inactivity	0^A	0.65	1.31
	8^A	1.70	2.25
	16^A	1.85	3.01

Table 3. Descriptive statistics for the efficiency metrics. Bold font indicates statistical significance ($p < 0.05$). Conditions with the same letter are not significantly different.

3.9.2 Effectiveness

We did not find a significant main effect of cue condition on the *post-training knowledge test*, $F(2, 57) = 0.110$, $p = 0.896$, $\eta^2 = 0.004$. See Table 4 for the descriptive statistics.

Metric	Delay	Mean	Std. Dev
Knowledge Test	0^A	8.70	2.43
	8^A	8.40	3.15
	16^A	8.25	3.58

Table 4. Descriptive statistics for the knowledge test metrics.

3.9.3 Subjective Responses

We did not find a significant main effect on *simulator sickness*, $\chi^2(2) = 0.115$, $p = 0.944$. We did not find a significant main effect on *presence*, $\chi^2(2) = 0.393$, $p = 0.822$. We did not find a significant main effect on *perceived usability*, $\chi^2(2) = 1.442$, $p = 0.486$. We did not find a significant main effect on *perceived scenario fidelity*, $\chi^2(2) = 0.479$, $p = 0.787$. Finally, we also did not find a significant main effect on *body ownership*, $\chi^2(2) = 0.760$, $p = 0.684$. See Table 5 for the descriptive statistics.

Metric	Delay	Mean	Std. Dev
Simulator Sickness	0^A	19.8	14.8
	8^A	20.8	19.4
	16^A	21.7	21.3
Presence	0^A	3.93	0.70
	8^A	3.99	0.72
	16^A	4.06	0.69
Perceived Usability	0^A	76.5	13.1
	8^A	72.8	14.8
	16^A	71.9	12.3
Scenario Fidelity	0^A	3.93	0.65
	8^A	3.76	0.81
	16^A	3.96	0.43
Body Ownership	0^A	3.51	0.81
	8^A	3.59	0.81
	16^A	3.79	0.72

Table 5. Descriptive statistics for the subjective responses.

3.10 Discussion

3.10.1 Immediate Cues are More Efficient

In our study, we found multiple results indicating that immediate interaction cues are more efficient than delayed interaction cues. We found that the 0-second condition was significantly faster than the 16-second condition. There was also a statistical trend that the 0-second condition was faster than the 8-second condition. We found that the 0-second and 8-second conditions afforded significantly fewer physical errors than the 16-second condition. We also found that the 0-second condition afforded significantly fewer incorrect decisions than the 16-second condition.

These results support our H1 hypothesis that the 0-second condition would be the most efficient because users would not need to wait for the interaction cues and would not make mistakes while waiting. The implications of these results are that designers should avoid delayed interaction cues when efficiency is a priority. In the case of non-educational applications, such as video game tutorials, we recommend only using immediate cues.

3.10.2 Delayed Cues are Not More Effective

In terms of training effectiveness, we found no significant differences among the three cue conditions. We did not find a significant main effect of cue condition on the post-training knowledge test (i.e., knowledge acquisition).

These results do not support our H2 hypothesis that the 8-second condition would be the most effective. The descriptive statistics actually indicate that this hypothesis is incorrect, as the 0-second condition yielded the best results for our effectiveness metric. Hence, delayed cues are not more effective for training than immediate cues.

3.10.3 No Subjective Differences due to Cue Timing

In terms of subjective responses, we also found no significant differences among the three cue conditions. We did not find a significant main effect of cue condition on simulator sickness, presence, perceived usability, perceived scenario fidelity, or body ownership.

These results do not support our H3 hypothesis that the delayed cue conditions would reduce presence, perceived usability, and perceived scenario fidelity due to users waiting on the cues.

3.10.4 Implications for Training

Contrary to our hypothesis, delayed interaction cues were not more effective than the immediate cues. Our hypothesis was based on prior training research that indicated active exploration and error encouragement are effective training elements and should be used instead of error-avoidant training [33].

Furthermore, it is important to consider the efficiency of interaction cues. In our study, we found the 0-second condition to be significantly more efficient than the 16-second condition with a trend of also being more efficient than the 8-second condition. While efficiency clearly does not equate to effectiveness, it does affect how quickly the training will become effective. For example, in the surgical domain, surgical skills curriculums are designed for training to proficiency, not for training based on a specified number of repetitions or time [34]. In such curriculums, immediate interaction cues are likely to reduce the overall time required to train to proficiency, as opposed to delayed cues.

3.11 Limitations

Two limitations of our study were the lack of evaluating skills acquisition immediately after training and the lack of evaluating skills transfer to the real world. We purposefully chose not to evaluate skills acquisition to avoid a lengthy first session for users. However, we were unable to evaluate skills transfer due to a lack of access to a surgical operating room and robot. Hence, future research needs to be conducted to determine how the timing of interaction cues affect skill acquisition and transfer.

Another limitation of our study is the number of completed participants (60 total; 20 per condition). While this number was sufficient for finding significant differences in terms of efficiency, we did not find any significant differences in terms of effectiveness and subjective responses. However, a post hoc power analysis of the post-training knowledge test results revealed that approximately 1382 participants would need to be recruited for each condition to obtain statistical power at the 0.80 level recommended by Cohen [35]. Hence, if there are effectiveness differences among the conditions, those differences would be extremely small and would like indicate that immediate cues are more effective than delayed ones.

Finally, another limitation of our study is the underlying VR training application used. The inherent effectiveness of this application may have a greater statistical power effect than the investigated cue conditions, which would explain the lack of significant differences found. However, in order to investigate such timing effects, some training application must be used or created, which is just as likely to have a greater statistical power than the delayed cues. In other words, research must start somewhere, and we have started with our VR training application.

4 Conclusion

Interaction cues are commonly used by game designers and instructional designers to inform users about actions to take. We have used a substantial literature review pertaining to games and educational systems to expand upon a prior framework of visual interaction cues to develop a robust taxonomy of multimodal interaction cues, based on their purpose, modality, markedness, presentation, trigger, and timing. We have also presented one of the first controlled evaluations of delayed interaction cues, which indicated that immediate cues are significantly more efficient than delayed ones. We have also discussed the implications of our results on future games and training systems.

5 ACKNOWLEDGMENTS

This material is based on work partially supported by the National Science Foundation under Grant No. 1552344 – “CAREER: Leveraging the Virtualness of Virtual Reality for More-Effective Training.”

6 REFERENCES

[1] K. R. Dillman, T. T. H. Mok, A. Tang, L. Oehlberg, and A. Mitchell, “A Visual Interaction Cue Framework from Video Game Environments for Augmented Reality,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 140:1--140:12.

[2] S. Bardzell, “Systems of Signs and Affordances: Interaction Cues in 3D Games,” *Extending Exp. Struct. Anal. Des. Comput. game Play. Exp.*, pp. 191–209, 2008.

[3] M. Rohs and P. Zweifel, “A Conceptual Framework for Camera Phone-Based Interaction Techniques BT - Pervasive Computing,” 2005, pp. 171–189.

[4] I. Arroyo, H. Mehranian, and B. P. Woolf, “Effort-based tutoring: An empirical approach to intelligent tutoring,” in *Educational Data Mining 2010*, 2010.

[5] P. Renevier, L. Nigay, J. Bouchet, and L. Pasqualetti, “Generic Interaction Techniques for Mobile Collaborative Mixed Systems BT - Computer-Aided Design of User Interfaces IV,” 2005, pp. 309–322.

[6] F. Buttussi and L. Chittaro, “MOPET: A context-aware and user-adaptive wearable system for fitness training,” *Artif. Intell. Med.*, vol. 42, no. 2, pp. 153–163, 2008.

[7] L. Hernandez, J. Taibo, A. Seoane, R. Lopez, and R. Lopez, “The experience of the Empty Museum: displaying cultural contents on an immersive, walkable VR room,” in *Proceedings Computer Graphics International, 2004.*, 2004, pp. 436–443.

[8] A. Webster, S. Feiner, B. MacIntyre, W. Massie, and T. Krueger, “Augmented reality in architectural construction, inspection and renovation,” in *Proc. ASCE Third Congress on Computing in Civil Engineering*, 1996, vol. 1, p. 996.

[9] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. La Viola, T. Mashita, and H. Takemura, “The effectiveness of an AR-based context-aware assembly support system in object assembly,” in *2014 IEEE Virtual Reality (VR)*, 2014, pp. 57–62.

[10] J. C. P. Chan, H. Leung, J. K. T. Tang, and T. Komura, “A Virtual Reality Dance Training System Using Motion Capture Technology,” *IEEE Trans. Learn. Technol.*, vol. 4, no. 2, pp. 187–195, 2011.

[11] O. Engwall, “Analysis of and feedback on phonetic features in pronunciation training with a virtual teacher,” *Comput. Assist. Lang. Learn.*, vol. 25, no. 1, pp. 37–64, Feb. 2012.

[12] A. Robb *et al.*, “Training Together: How Another Human Trainee’s Presence Affects Behavior during Virtual Human-Based Team Training ,” *Frontiers in ICT* , vol. 3. p. 17, 2016.

[13] H. Q. Dinh, N. Walker, L. F. Hodges, C. Song, and A. Kobayashi, “Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments,” in *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, 1999, pp. 222–228.

[14] J. W. Burke, M. McNeill, D. Charles, P. Morrow, J. Crosbie, and S. McDonough, “Serious Games for Upper Limb Rehabilitation Following Stroke,” in *2009 Conference in Games and Virtual Worlds for Serious Applications*, 2009, pp. 103–110.

[15] D. S. Vincent, A. Sherstyuk, L. Burgess, and K. K. Connolly, “Teaching Mass Casualty Triage Skills Using Immersive Three-dimensional Virtual Reality,” *Acad. Emerg. Med.*, vol. 15, no. 11, pp. 1160–1165, Nov. 2008.

[16] J. C. Eubanks, V. Somareddy, R. P. McMahan, and A. A. Lopez, “Full-Body Portable Virtual Reality for Personal Protective Equipment Training BT - Virtual, Augmented and Mixed Reality,” 2016, pp. 490–501.

[17] J. Bertrand *et al.*, “The role of dimensional symmetry on bimanual psychomotor skills education in immersive virtual environments,” in *2015 IEEE Virtual Reality (VR)*, 2015, pp. 3–10.

[18] J. Wang, Y. Feng, C. Zeng, and S. Li, “An augmented reality based system for remote collaborative maintenance instruction of complex products,” in *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, 2014, pp. 309–314.

[19] L. Marchal-Crespo, S. McHughen, S. C. Cramer, and D. J. Reinkensmeyer, “The effect of haptic guidance, aging, and initial skill level on motor learning of a steering task,” *Exp. Brain Res.*, vol. 201, no. 2, pp. 209–220, 2010.

[20] R. P. McMahan, S. Schafrik, D. Bowman, and M. Karmis, “Virtual environments for surface mining powered haulage training,” in *Extracting the Science: A Century of Mining Research*, J. Brune, Ed. SME, 2010, pp. 520–528.

[21] P. Khanal *et al.*, “Collaborative virtual reality based advanced cardiac life support training simulator using virtual reality principles,” *J. Biomed. Inform.*, vol. 51, pp. 49–59, 2014.

[22] C. Lai, R. P. McMahan, M. Kitagawa, and I. Connolly, “Geometry Explorer: Facilitating Geometry Education with Virtual Reality BT - Virtual, Augmented and Mixed Reality,” 2016, pp. 702–713.

[23] R. P. Sanchez, C. M. Bartel, E. Brown, and M. DeRosier, “The acceptability and efficacy of an intelligent social tutoring system,” *Comput. Educ.*, vol. 78, pp. 321–332, 2014.

[24] R. S. Baker, A. T. Corbett, and K. R. Koedinger, “Detecting Student Misuse of Intelligent Tutoring Systems BT - Intelligent Tutoring Systems,” 2004, pp. 531–540.

[25] O. Gaggi *et al.*, “Serious Games for Early Identification of Developmental Dyslexia,” *Comput. Entertain.*, vol. 15, no. 2, pp. 4:1--4:24, Apr. 2017.

[26] C. Galloway, “Blink and they’re gone: PR and the battle for attention,” *Public Relat. Rev.*, vol. 43, no. 5, pp. 969–977, 2017.

[27] R. Hernández-Irizarry, B. Zendejas, S. M. Ali, and D. R. Farley, “Optimizing training cost-effectiveness of simulation-based laparoscopic inguinal hernia repairs,” *Am. J. Surg.*, vol. 211, no. 2, pp. 326–335, 2016.

[28] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, “Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness.,” *Int. J. Aviat. Psychol.*, vol. 3, no. 3, pp. 203–220, 1993.

[29] T. Hartmann *et al.*, “The Spatial Presence Experience Scale (SPES),” *J. Media Psychol.*, vol. 28, no. 1, pp. 1–15, Jan. 2015.

[30] J. Brooke, “SUS: A Retrospective,” *J. Usability Stud.*, vol. 8, no. 2, pp. 29–40, Feb. 2013.

[31] R. P. McMahan and N. S. Herrera, “AFFECT: Altered-Fidelity Framework for Enhancing Cognition and Training,” *Frontiers in ICT*, vol. 3. p. 29, 2016.

[32] A. Maselli and M. Slater, “The building blocks of the full body ownership illusion ,” *Frontiers in Human Neuroscience* , vol. 7. p. 83, 2013.

[33] N. Keith and M. Frese, “Effectiveness of error management training: A meta-analysis.,” *Journal of Applied Psychology*, vol. 93, no. 1. American Psychological Association, Keith, Nina: Justus-Liebig University of Giessen, Department of Work and Organizational Psychology, Otto-Behaghel Street 10 F, Giessen, Germany, 35394, nina.keith@psychol.uni-giessen.de, pp. 59–69, 2008.

[34] J. Bric, M. Connolly, A. Kastenmeier, M. Goldblatt, and J. C. Gould, “Proficiency training on a virtual reality robotic surgical skills curriculum,” *Surg. Endosc.*, vol. 28, no. 12, pp. 3343–3348, 2014.

[35] J. Cohen, “A power primer.,” *Psychological Bulletin*, vol. 112, no. 1. American Psychological Association, US, pp. 155–159, 1992.