Adaptive XR training systems design, implementation, and evaluation

by

Melynda Hoover

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Human Computer Interaction

Program of Study Committee:
Eliot Winer, Major Professor
Michael Dorneich
Stephen Gilbert
Jonathan Kelly
Elizabeth Shirtcliff

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

NOMENCLATURE

AR	Augmented Reality
DV	Dependent Variable
H/Hy	Hypothesis
HMD	Head Mounted Display
ITS	Intelligent Tutoring System
IV	Independent Variable
LOA	Levels of Automation
M	Mean
Mdn	Median
MR	Mixed Reality
n	Sample Size
RQ	Research Question
SD	Standard Deviation
SE	Standard Error
VE	Virtual Environment
VR	Virtual Reality
XR	Extended Reality

ACKNOWLEDGMENTS

I imagine that getting a Ph.D. is kind of like an ultra-marathon. (Not that I've ever run an ultra-marathon, but this is how I imagine it goes.) There's of course your trainer who whips you into shape, critiques your form, and makes sure you have all the skills you need to cross the finish line. I couldn't have found a better trainer than my committee chair, major professor, and advisor, Eliot Winer. He recognized my curiosity and enthusiasm for learning as a freshman in Mechanical Engineering and immediately hired me to do research in his lab. Without that opportunity and his guidance, I wouldn't be where I am today.

Next, you have a team of other veteran runners with unfathomable knowledge and skills to draw from and help hone your technique, AKA my committee. Thank you to Stephen Gilbert, Jon Kelly, Michael Dorneich, and Birdie Shirtcliff for exercising my mind by asking the tough questions. My dissertation and critical thinking skills are exponentially improved, thanks to you all.

Of course, during the race, you're slogging along with a bunch of other sorry saps who were just crazy enough to sign up to run too. Those would be my fellow grad students. Stacy MacAllister, thanks for taking me under your wing and teaching me everything I know about academic writing. Your patience with me as a clueless nineteen-year-old was otherworldly. There's a reason people used to call me Stacy junior and that's because you are truly a role model of mine. Jack Miller and Robert Philpott, thank you for sharing in my joys and frustrations during the last few years. Your support and friendship have meant the world to me. And of course, Kaitlyn Ouverson, thank you, for being my closest friend and "Productivity Nanny" for the last year while we slogged through writing dissertations, journal papers, and data analysis. I honestly don't know what I would have done without our daily accountability meetings during

the pandemic. Working with a fellow girl boss and getting things done together has been the most empowering experience. I can't wait to see you cross that finish line too!

Lastly, you have the team that rides along in the van to cheer you on, feed, and water you every few hours. That would be my generous and loving friends and family. Thank you to my mom and dad who raised me to be persistent (read stubborn) enough to cross that finish line. You always told me I could do anything I set my mind to, and that has stuck with me through this process which has so often come down to mind over matter. And of course, thank you to my amazing husband, Derek Kreidler for sticking by my side through the peaks and valleys. Your smile and laughter after a long day of writing make all my worries and fears melt away. You have been my inspiration for finishing this PhD each and every day. I can't wait to see what the next phase of life has in store for us.

To everyone listed here, and those who aren't, thank you for being on my team. We did it!

ABSTRACT

Extended reality (XR) is a continuum encompassing various current and future virtual technologies such as augmented reality (AR) and virtual reality (VR). These technologies have been shown to increase learning outcomes when applied to training. For example, virtual and augmented reality have been shown to improve performance during assembly task training over traditional methods. However, even an advanced technology like XR won't provide a benefit if the information is ill-timed or poorly suited to the problem. One potential way to solve this issue is with adaptive automation. Adaptive automation provides the ability to off-load tasks from a human to a computer. It has been shown to increase productivity with the added benefit of reducing human-labor costs. Previous research has shown that applying automation at the wrong levels or during the wrong stage of human information processing can result in decreased performance. Furthermore, if there are inaccuracies or gaps in the automated system, resulting in brittle automation, a human can develop an overreliance on the automation leading to costly mistakes.

One place where adaptive automation is already being applied to XR technology is in the field of intelligent tutoring systems (ITS). Designers of XR ITS are increasingly automating the role of a human instructor in an effort to decrease the instructor to learner ratio. While works touting the development of this type of adaptive XR training systems exist, few address the design and efficacy of adaptations that do not directly pertain to the learning materials and pedagogy such as troubleshooting software and hardware use. Without this information, it is impossible for today's researchers and developers to know how to design and implement these types of adaptive XR training systems. Therefore, this dissertation will identify and evaluate

effective triggers and adaptations for XR training environments so that they can be replicated for future applications.

This research sought to solve this problem in two stages. In the first stage, human XR instructors were used as models for potential adaptive automation schema. Since automation is the computer execution of a task once performed by a human, it was natural to choose human instructors as models for this work. Therefore, 11 semi-structured interviews were conducted with XR training and simulation experts. The questions posed during the interviews were crafted to determine how these instructors identified and mitigated confusion in learners. The information from these interviews was then analyzed for themes and synthesized with existing adaptive automation models. During the analysis phase, identifying and mitigating confusion were found to be analogous to the functions of triggers and adaptations in the adaptive automation model. These interviews resulted in two sets of recommendations. First, when designing adaptive triggers in XR, verbal triggers should be prioritized, followed by physical triggers. A head-mounted display (HMD) can be used to monitor verbal and physical triggers using a factory standard microphone and inertial measurement unit. Second, verbal and physical channels should also be prioritized when designing adaptation methods. This can be done through recorded audio messages, textual interfaces, and by providing demonstrative animations and models in an XR environment. Finally, adaptations with increasing levels of specificity and intrusiveness allow learners to solve problems independently.

The second phase of this research implemented these recommendations within an XR task. A simple triangle completion task was chosen, in which a learner must teleport to three different positions in a VE using two different techniques. Afterwards, the learner is tested on their ability to remember where they started by pointing at the origin and clicking using an XR

controller. Triggers and adaptations were developed to assist in the correction of five different erroneous behaviors during the task. These included incorrect button presses, inaction, and attempting to teleport to the wrong location. Finally, the resultant unsupervised, remote, adaptive XR simulation was evaluated to determine the efficacy of the adaptations and compared to a version of the simulation with no adaptations and a human instructor present to provide feedback, when necessary. The results of this experiment validated that the adaptations were sufficient at providing instructions to participants during the remote unmoderated study because the participant success rate was equal to that of the lab-based study (89%). In addition, participants' performance and completion times were not statistically different from those of the lab group. The adaptations triggered when expected and had the intended effect of helping learners correct their mistakes. Finally, participants gave feedback about the intrusiveness, helpfulness, and quantity of adaptive feedback they received. It was found that the quantity of feedback was adequate without being too intrusive, however, there were mixed reviews about the subjective helpfulness of the feedback. Ultimately, this research was successful at increasing the efficacy of adaptive XR systems, and reducing the time and cost associated with humans facilitating XR training.

CHAPTER 1. GENERAL INTRODUCTION

Motivation

The benefits of extended reality (XR) on training outcomes and procedural task performance are numerous. For example, augmented reality (AR) has been shown to reduce errors and time spent on assembly tasks by as much as 50% (Hou et al., 2015). This can result in notable cost savings for manufacturers. Another performance benefit of XR instruction is increased accuracy in high-precision tasks such as welding (Doshi et al., 2017). Furthermore, XR technology has been shown to increase presence and engagement during training, while simultaneously decreasing mental workload in some cases (Buttussi & Chittaro, 2018; Loch et al., 2016). Beyond these advantages, AR and virtual reality (VR) provide capabilities that cannot be achieved with traditional training. Specifically, AR can provide timely instructions to learners while in the field and VR can simulate experience that are too dangerous or expensive to recreate in real life (Caudell & Mizell, 1992; Lele, 2013). However, one disadvantage of XR training is consistent monitoring of a trainee (learner) by an expert instructor to maintain situational awareness and ensure that a learner completes tasks successfully (Gutwin & Greenberg, 2002). This creates a dynamic where an instructor must remain vigilant over a learner leading to inefficiencies.

Adaptive automation has the capability of supplanting or augmenting the capabilities of an instructor. In fact, adaptive training systems are essentially the automation of functions typically performed by a skilled instructor (Kelley, 1969). By replacing an instructor with adaptive automation, a human-automation team is established. However, simply introducing automation is not enough to ensure increased performance. Research has shown that high levels of automation can result in miscalculated trust and over-reliance on the system, thereby degrading performance (Smith et al., 1997). Over-reliance on automation is even riskier in training scenarios, as learners will be expected to perform tasks without the help of automation following training. One solution for this issue is to assign information acquisition and implementation tasks to the training system while allowing a human to be responsible for analytical thinking and decision making (Kaber et al., 2005). Applying this theory to adaptive XR training could result in simulations that train a learner in semantic knowledge and problem solving, while procedural knowledge and physical skills are mostly obtained in the field. Depending on the training goals, this may or may not be desirable. For example, if training on the job is inherently dangerous or expensive, it might make more sense to conduct all training in XR. In these cases, developers would need to determine if it is appropriate or not to allocate task execution to the computer or the trainee. Unfortunately, the terminology surrounding the combination of adaptive systems and XR technology is complex and varied due to its interdisciplinary nature, making it difficult to ascertain how to design these systems.

Existing research combining adaptive automation and XR use a variety of terms to describe the nexus of these technologies, including intelligent tutoring systems (ITS), computer-assisted instruction (CAI), adaptive virtual learning environment (VLE), intelligent VLE, and adaptive tutor/training/simulator. These terms are not completely interchangeable, and often overlap in ambiguous ways, making literature review in this space difficult. One article attempts to bridge the semantic gap by discussing the application of ITS to AR technology. The authors called the combination of these technologies Adaptive Augmented Reality Tutors (ARAT) (Herbert et al., 2018). While the authors presented a limited taxonomy and guidelines for choosing authoring tools and a system architecture, no recommendations for the design of adaptation methods and triggers were mentioned. Furthermore, ITS do not encompass the entirety of adaptive XR systems, rather, the field of ITS focuses on the application of adaptive automation to education and curricula (Sleeman & Brown, 1982). However, adaptive systems can be applied to other aspects of XR training as well. For example, learning to use the hardware itself and troubleshooting software challenges are common issues during XR training (Hoover & Winer, n.d.). Since the majority of the existing work combining adaptive systems and XR technology centers around ITS, this leaves a gap in the body of work on adaptive XR systems where pedagogy is not the main concern.

Currently there is little to no literature that focuses on the implementation of adaptive system as it pertains to helping a user execute tasks successfully in VR and AR. The current gap in this area of research means that there are no existing standards for the design of non-pedological adaptations within XR simulations. By not having adequate guidelines for designing these types of XR systems, developers are left guessing which triggers and adaptations are most suitable. This uncertainty will lead to mistakes and negative training outcomes. This research seeks to fill these gaps by interviewing experts, identifying design recommendations, and testing them on an example procedural task.

Dissertation Organization

The research will be presented as follows. Chapter 2 will introduce existing literature on topics related to this research including XR training and adaptive automation. Chapter 3 describes the research questions addressed in this dissertation. Chapter 4 is a research paper published in the proceedings of the 2020 Interservice/Industry Training, Simulation and Education Conference. The paper is entitled, "Situational Awareness Methods in Virtual Reality Training: A Scoping Review" and provides perspective on the instructor's role in VR training. Chapter 5 is a journal paper describing the methodology, results, and conclusions of the first phase of this research. It is entitled “Designing Adaptive Extended Reality Training Systems Based on Expert Instructor Behaviors” and is currently under review in the IEEE Access journal. Chapter 6 contains a research paper entitled “Integrating Adaptive Feedback into an Online VR Research Task” which is currently under review as a submission to IEEE Transactions on Visualization and Computer Graphics. The paper describes the methods and results of a remote, unsupervised VR study with adaptive feedback. Chapter 7 contains supplemental results from the same study that were not included in the journal paper. Finally, Chapter 8 summarizes the results from this research and how they addressed the research questions as well as future related work.

References

Buttussi, F., & Chittaro, L. (2018). Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. IEEE Transactions on Visualization and Computer Graphics, 24(2), 1063–1076. https://doi.org/10.1109/TVCG.2017.2653117

Caudell, T. P., & Mizell, D. W. (1992). Augmented reality: an application of heads-up display technology to manual manufacturing processes. Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences, 659–669 vol.2. https://doi.org/10.1109/HICSS.1992.183317

Doshi, A., Smith, R. T., Thomas, B. H., & Bouras, C. (2017). Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. The International Journal of Advanced Manufacturing Technology, 89(5–8), 1279–1293. https://doi.org/10.1007/s00170-016-9164-5

Gutwin, C., & Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. Computer Supported Cooperative Work, 11(3–4), 411–446. https://doi.org/10.1023/A:1021271517844

Herbert, B., Barrett, E., Weerasinghe, A., Billinghurst, M., & Wigley, G. (2018). Design considerations for combining augmented reality with intelligent tutors. Computers & Graphics, 77, 166–182. https://doi.org/10.1016/j.cag.2018.09.017

Hoover, M., & Winer, E. (n.d.). Designing Adaptive Extended Reality Training Systems Based on Expert Instructor Behaviors. IEEE Access.

Hou, L., Wang, X., & Truijens, M. (2015). Using Augmented Reality to Facilitate Piping Assembly: An Experiment-Based Evaluation. Journal of Computing in Civil Engineering, 29(1), 05014007. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000344

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. Human Factors, 47(4), 730–741. https://doi.org/10.1518/001872005775570989

Kelley, C. R. (1969). What is Adaptive Training? Human Factors: The Journal of the Human Factors and Ergonomics Society, 11(6), 547–556. https://doi.org/10.1177/001872086901100602

Lele, A. (2013). Virtual reality and its military utility. Journal of Ambient Intelligence and Humanized Computing, 4(1), 17–26. https://doi.org/10.1007/s12652-011-0052-4

Loch, F., Quint, F., & Brishtel, I. (2016). Comparing video and augmented reality assistance in manual assembly. Proceedings - 12th International Conference on Intelligent Environments, IE 2016, 147–150. https://doi.org/10.1109/IE.2016.31

Sleeman, D., & Brown, J. (1982). Intelligent tutoring systems. https://halshs.archives-ouvertes.fr/hal-00702997/

Smith, P. J., McCoy, C. E., & Layton, C. (1997). Brittleness in the design of cooperative problem-solving systems: The effects on user performance. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans., 27(3), 360–371. https://doi.org/10.1109/3468.568744

CHAPTER 2. BACKGROUND

Effects of Extended Reality Training

Virtual Reality (VR) allows a user to experience a Virtual Environment (VE) that can completely differ from their current reality, as every object in their field of view is computer generated. Similarly, Augmented Reality (AR) lets a user see a version of their environment supplemented with additional computer-generated information. Both methods typically use digital displays and location tracking to render the VE. The virtual reality continuum, also known as Extended Reality (XR), or Mixed Reality (MR), encompasses both Virtual and Augmented Reality (Milgram et al., 1994). The uses for XR technology continue to grow as more hardware becomes available and additional software functionality is developed. However, one of the earliest use cases for XR was for training applications. As early as the 1930s, XR was used to train U.S. military pilots via the Link flight trainer (Link, 1931). This device combined an artificial cockpit with motion controls to create a flight experience on the ground. Later, in the 1950s, screens were added to the simulators to mimic the pilot's view of the outside world (Page, 2000). XR technology is often chosen for military and other training applications because of its ability to simulate situations that are dangerous, difficult, or expensive to stage in real life (Lele, 2013). Other industries have chosen this technology because of its benefits to performance and mental workload (De Crescenzio et al., 2011; Hou et al., 2015). Regardless of the industry, stakeholders must weigh the benefits and costs of XR technology before employing it as a training medium.

Performance Improvements from XR-enabled Training

The promise of performance outcomes can greatly influence the decision to adopt XR technology. The lengthy history of research in VR and AR technology has shown that it can deliver training that is as effective as traditional training mediums such as manuals and computer-based systems, but with the added advantage of time and cost savings (Kaplan et al., 2020). Additionally, some studies have concluded that XR technology results in greater performance benefits than traditional training.

As early as 1999, researchers were realizing the advantage XR technology posed for reducing errors in assembly-related training tasks. More specifically, Baird and Barfield published work that found that participants made fewer errors on a motherboard assembly task when using AR Head Mounted Displays (HMDs) to deliver instructions rather than paper or computer-based instructions. The authors also concluded that the increased performance of AR HMD users could be attributed to the phenomenon of recognition over recall. Since the information was constantly in a user's field of view, they did not have to reread, or put forth as much effort to memorize information (Baird & Barfield, 1999). More recently, a study by Hou et al. in 2015 compared the use of AR instructions to traditional isometric drawings when putting together a pipe assembly. They found that the use of AR instructions resulted in 50% reduction in assembly errors compared to the control group (Hou et al., 2015). The performance advantages of AR training were further upheld by a 2017 study which found AR to increase precision in a welding task by 52% (Doshi et al., 2017). Furthermore, in a 2020 study at Iowa State University, assembly instructions for a mock aircraft wing provided on the Microsoft HoloLens HMD were found to result in fewer assembly errors than instructions provided on a tablet or desktop computer (Hoover et al., 2020).

A similar pattern of increased performance was discovered when studying VR technology. In 2002, Vora et al. developed a VR HMD system which simulated the inspection of an aircraft cargo bay. They compared their VR trainer to a desktop-based trainer that showed 2D instructions and found that subjects were able to identify more defects and perform the task faster when using VR than when using the PC trainer (Vora et al., 2002). However, research suggests that the training benefits of VR technology are smaller in magnitude compared to those resulting from AR. Specifically, a study in 2015, which compared both VR and AR training tools to traditional training methods, found that AR resulted in fewer errors than the control, while VR training did not (Gavish et al., 2015). Another study by Buttussi and Chittaro compared the effects of three different computer / hardware setups on aviation safety training: 1) a traditional computer monitor, 2) a 3-DOF VR HMD, and 3) a modern 6-DOF VR HMD (Buttussi & Chittaro, 2018). They found that all of the systems resulted in similar training benefits, but the higher fidelity HMD resulted in more trainee engagement and presence (Buttussi & Chittaro, 2018).

Knowledge Transfer & Retention in Training Performance

Knowledge transfer and retention is another measure of training performance and can be used to evaluate the efficacy of XR training systems. For example, a study by Botden et al., found that in a direct comparison on AR and VR laparoscopic surgery simulators, the AR system resulted in higher construct validity than the VR training system. This means that performance on AR training system better reflected the learner’s real expertise in the task (Botden et al., 2007). In the laparoscopic surgery training example, the authors attribute the higher performance level of the AR system to its realistic haptic feedback. This shows that some tasks, such as those requiring fine motor skills, may be more suited to AR, as VR can't replicate the touch and feel of the real-world accurately enough. In another study comparing VR to traditional training, Waller and Miller studied subjects who were trained to complete circuit board tasks using one of three training methods: desktop VR, paper, or video tutorials (Waller & Miller, 1998). They found that the VR training condition took more time to complete, but led to better performance retention over the course of two-weeks than the traditional training methods (Waller & Miller, 1998). Contrastingly, a small-scale study by Hall, Stiles & Horwitz, found that knowledge lost over time was similar for VR and 2D learners (Hall et al., 1998). These studies show that the use of XR training systems perform equally, if not better than some traditional training methods in knowledge retention.

Mental Workload during Training

Another benefit of some XR-enabled training systems is the ability to reduce mental workload. Mental workload is the amount of cognitive effort a learner must put forth to execute a task (Wickens et al., 1998). AR training systems tend to benefit from reduced mental workload over traditional training materials because they can provide critical information in the moment it is most needed without dividing a learner's attention. A common way of measuring mental workload during a task is by using the NASA Task Load Index (NASA-TLX) Questionnaire (Hart & Staveland, 1988). Loch et al. used the NASA-TLX to gauge mental workload in a Lego brick assembly task comparing video and AR instructions (Loch et al., 2016). They found that the AR instructions resulted in a lower mental workload in participants than traditional instructions. Similarly, Hou at al., found that their AR instructions yielded significantly lower mental workload values that traditional schematic drawings for the assembly of pipe components (Hou et al., 2015).

Reduction of Training Time

A system that can reduce the time to produce a fully trained learner is critical, as direct reductions in cost often result. One way to reduce overall training time is to reduce the necessary duration of a training task. Several studies point to the time saving advantages of using XR training, especially when AR is used (Boud et al., 1999; Hou et al., 2015). In their 2015 study, Hou et al. found that the use of AR as a training medium could reduce assembly training time by as much as 50%. Similar results have been published by other authors, which indicate AR reduces time spent searching for objects of interest during tasks. Weaver et al. found that users were able to perform a part picking task faster when using AR versus paper instructions (Weaver et al., 2010). Similarly, Henderson and Feiner found that military maintenance personnel could locate critical systems in need of servicing faster when using AR than with 2D digital instructions (Henderson & Feiner, 2009). Although these studies focused on providing on the job instruction rather than training, the advantages and lessons learned from this research can be applied to training as well. One such lesson is that AR may have limited improvement over traditional instructional methods when applied to overly simple tasks (Wiedenmaier et al., 2003).

While AR is more suited to presenting instructions on the job, the effects of both AR and VR have been studied as well. However, results pertaining to their effect on training time have been mixed. A study by Boud et al., researched the effect of AR, VR, and 2D training methods on unassisted task completion time after the training (Boud et al., 1999). They found that AR resulted in the fastest completion times after training, followed by the VR condition. Both AR and VR were faster than conventional training methods in this case (Boud et al., 1999). A study by Peniche et al. used a different approach that studied the combined effects of VR and AR together on training outcomes. Their method used five VR training sessions followed by three AR training sessions. This AR/VR training schema resulted in similar reductions in task completion time over the course of the training as traditional training methods. The authors concluded that AR/VR training could be a viable replacement for traditional manufacturing training methods (Peniche et al., 2012). In another study, by Hořejší, a simple augmented reality system using a webcam and a computer monitor display were used as a training tool for novice users assembling plumbing pieces. The study showed that participants who used the AR instructions during training learned to assemble the parts faster, and in fewer attempts, than their counterparts who used paper instructions (Hořejší, 2015). Contrastingly, a 2015 study by Gavish et al. compared AR, VR, and traditional methods for training on an electronic actuator maintenance task. They found that although there were some reductions in errors made, both the AR and VR training methods resulted in longer training times than their respective control groups (Gavish et al., 2015).

Cost Benefits of XR-enabled Training

All of the advantages listed in the previous subsections result in cost savings for training stakeholders. In fact, Hou et al. conducted a detailed cost/benefit analysis of using AR to provide instruction for a manufacturing task. They found that the reduction in overall task time, paired with the reduction in errors and resulting re-work resulted in an overall cost savings of 66% (Hou et al., 2015). However, in order to justify the use of XR training, the cost savings associated with the benefits must outweigh the increased cost of implementing specific XR technologies (Chung et al., 2002). This includes the cost of hardware and software. Luckily, the cost of AR and VR HMDs has been greatly reduced in recent years with the advent of commodity XR equipment. Additionally, development for VR has become more accessible with the release of free and inexpensive 3D development platforms like Unity and Unreal Engine. This has resulted in a lower barrier to adoption of XR technology for training (Checa & Bustillo, 2020).

In summation, XR technology has been shown to improve performance factors in various training applications and industries. In some cases, it can also improve mental workload on trainees. However, improvements can still be made in these areas, and the amount of time required for XR training is often still greater than that of traditional training, especially in the case of VR technology. This is compounded when one considers not only the trainee's time, but that of the expert instructor as well, resulting in higher monetary costs to the organization overall. One method with the potential to reduce both training time and instructor time – thereby reducing costs – is the integration of adaptive systems with XR technology.

Automation

Put simply, automation is the computer execution of tasks that were once the responsibility of a human (Parasuraman et al., 2000). These tasks can range from gathering and providing information to making decisions and acting on them. Often, humans work together with a computer in a human-automation team to complete tasks, resulting in a tradeoff of control between these two parties. Levels of Automation (LOA) are used to describe this tradeoff.

In 1978, Sheridan and Verplank published their seminal work defining the concept of LOA. These ten levels defined how much control the human and the computer have over the system and are used throughout research literature to classify human-automation systems (Sheridan & Verplank, 1978). The ten levels are defined as follows according to Sheridan & Verplank, with one representing no automation, and ten representing complete automation with no human input (Sheridan & Verplank, 1978) :

1. The computer offers no assistance
2. The computer offers a complete set of decision/action alternatives
3. The computer narrows the selection down to a few

4. The computer suggests one alternative
5. The computer executes the one alternative if the human approves
6. The computer allows the human a restricted amount of time to veto before automatic execution
7. The computer executes the action automatically and then immediately informs the human
8. The computer informs the human of its actions only if asked
9. The computer informs the human of its actions online if the computer decides to
10. The computer decides everything, and acts autonomously with no regard for the human

Since then, several other authors have published alternative models for defining the levels of automation. However, Sheridan and Verplank's version remains the most cited in research literature (Vagia et al., 2016). More recently, Parasuraman, Sheridan, and Wickens amended the original LOAs to support a two-dimensional model with human vs computer control on one axis, and stages from Wickens's human information processing model on the other (Parasuraman et al., 2000). Their model differentiates between types of automation based on four possible functions served by the computer: 1) information acquisition, 2) information analysis, 3) decision selection, and 4) action implementation. Automation can occur at any level and any stage of human information processing. In fact, in many cases, automation occurs at multiple stages and levels within the same human-automation system.

Adaptive Automation

While traditional automation occurs at a fixed level over the duration of its use, adaptive automation is characterized by dynamic changes in function allocation between the human and the computer over time (Rouse, 1976). Oppermann further differentiated between *adaptive* and *adaptable* automation (Oppermann, 1994). By his definition, adaptive automation is characterized by computer control of function allocation, while adaptable automation allows a human to be in control. Furthermore, adaptable automation can be divided into categories based on how a human manipulates the system. Direct control describes when a human directly reallocates tasks between themselves and a computer. Alternatively, in supervisory control, a human acts on a system by changing the parameters of a task (Sheridan, 2011).

Adaptation Triggers

Adaptive automation relies on two main components: triggers and adaptation methods (also sometimes referred to as adaptation strategies). The trigger is the criteria that must be met to activate a change in function allocation. The adaptation method determines how, and which, functions are reallocated. Feigh, Dorneich, and Hayes created a taxonomy of triggers and strategies in 2012, which is relevant to this work. Their taxonomy of triggers included five overarching categories with various subcategories. Those five categories were related to subjects that could be measured within a human-automation team: 1) the operator, 2) the system, 3) the environment, 4) the task/mission, and 5) spatio-temporal factors (Feigh et al., 2012). Feigh et al. also proposed a taxonomy of adaptation methods that included modification of function allocation, task scheduling, interaction, and content (Feigh et al., 2012). To understand the effects of different triggers on adaptive automation, Lagu, Landry, and Yoo investigated the effects of physiological- (heart rate) and performance-based triggers on function allocation in adaptive automation systems using the "Multitask" software (Lagu et al., 2013). They found that adaptive automation resulted in increased levels of performance and that physiological triggers led to more variance in function allocation than performance-based triggers alone (Lagu et al., 2013).

Effects of Adaptive Automation

Some argue that situational awareness is decreased, and workload is increased, leading to potential failures when the human is not kept in the loop with automation (Kaber & Endsley, 2004; Lee & See, 2004; Sheridan, 2008). This is in part because an automation system may not be transparent in its decision-making process, leading to miscalculated trust and over-reliance on the automation (Shively et al., 2018; Smith et al., 1997). In their 1997 study, Smith et al. studied the effects of automated flight plan suggestions on the ability of pilots to perform a flight planning task. Participants completed a route planning task using one of three systems, each with varying levels of automation and amount of time spent using this automation versus manual control. They found that higher levels of automation often led to poor solutions, when the model used to inform the automated system was not robust. Since participants put too much trust in the automation, they were less accurate and efficient during training (Smith et al., 1997). Other researchers have sought to quantify and explain the effects of adaptive automation on these factors as well. One such study, by Kaber and Endsley, set out to analyze the relationship between workload, situational awareness, and LOA (Kaber & Endsley, 2004). They conducted a user study of a simulated air traffic control task using varied levels of automation as well as varied automation allocation cycle time. Participants were also directed to perform a secondary gauge-monitoring task as well. They used the Situational Awareness Global Assessment Technique (SAGAT) and NASA Task-Load Index (NASA-TLX) to measure the SA and workload during each condition level. Their research showed that an LOA that maintained human decision making with computer automated implementation paired with longer durations of automation resulted in the best task performance but poor SA and moderate workload. Additionally, performance on the secondary task suffered under these conditions. Contrastingly, performance suffered, and SA and workload improved with intermediate LOAs and shorter automation periods (Kaber & Endsley, 2004). This work demonstrated the tradeoffs between performance, SA, and workload, paving the way for more research into the use of adaptive automation for training.

The transitions between different levels of automation, and their effects on users, is also important. In 2005, Kaber et. al conducted a study on an air traffic control task, where the level of automation was varied between four levels. They tested how a user's performance changed when transitioned between their assigned automation level and fully manual control. The researchers concluded that these transitions led to less performance degradation when the automation was responsible for tasks such as information acquisition and action implementation, which require less analytical thinking and decision making (Kaber et al., 2005). These results suggest that human-automation teams perform best when a human remains responsible for processing information and making decisions, while a computer is responsible for gathering information and executing the decisions of a human. However, unlike computers, humans are unpredictable, and often do not use automation in this way.

Kirlik sought to explain why users did not always employ automation in expected ways during a simulated helicopter scouting task. During the task, teams of two controlled many simulated helicopters to complete certain missions. Users had the option to manually control the helicopters, or issue commands to be executed by the automation. He found that three design features (ease of use of automation vs manual control, autopilot engagement time, and autopilot disengagement time) as well as three task design features (secondary task duration, cost of secondary task delay, and interval between secondary task demands) impacted the way in which pilots chose to employ automation (Kirlik, 1993). Therefore, simply allocating tasks based on the innate skills of humans and computers is not enough to optimize performance.

Another factor that can impact how a human uses automation is trust. Systems in which humans over trust in the computer can have catastrophic effects (Shively et al., 2018). Similarly, under trusting the computer in a human-automation team can render processes inefficient. Nass, Fogg, and Moon studied how humans form teaming relationships with computers (Nass et al., 1996). They found that interdependence was the key factor influencing teaming behavior between a human and a computer. Users who were told their relationship with a computer was interdependent trusted the computer recommendations more and saw the computer as an ally (Nass et al., 1996). Similarly, Yang and Dorneich studied the effects of using affect-aware adaptive audio feedback to tutor students in math. They found that using an adaptive affect-aware etiquette strategy, rather than a random etiquette strategy, improved motivation, confidence, satisfaction, and performance (Yang & Dorneich, 2018). Using this affect aware strategy could increase a human’s trust in their computer counterpart during human-automation teaming.

Design Considerations or Effective Automation

Some considerations when designing automation that have been discussed include workload, situational awareness, and performance. Shively identified four more persistent problems facing human-automation teaming: 1) brittle automation, 2) lack of transparency, 3) lack of shared awareness, and 4) monitoring challenges (Shively et al., 2018). Brittle automation, due to an insufficient computational model underlies all these challenges and can lead to miscalculated trust in the abilities of the automation. This is especially true if the automation is not transparent about its decision-making process and the state of the system. To reduce the impact of these automation issues, Shively also proposed three design solutions. First, bi-directional communication should be employed so that a human can remain aware of the computer's state. Second, transparency should be provided by a computer so that a human can evaluate the performance of an automation. Lastly, a direct interface should be provided so a human may allocate tasks (Shively et al., 2018). Kaber et al. also recommended three research areas to help improve the future design of automation: 1) optimizing the tradeoffs between situational awareness and workload in adaptive automation, 2) designing UIs to better support function allocation and transparency, and 3) studying the impacts of adaptive automation on teams with multiple humans (Kaber et al., 2001). Although these guidelines represent a good starting point for the design of adaptive automation, they are not specific to the additional challenges of integrating adaptive automation into XR training.

Defining Adaptive Virtual Environments

Unlike in the field of automation research, terminology is not as clear cut when it comes to the field of XR, specifically with regards to training. Researchers have yet to conform to a uniform terminology to describe adaptive automation as a part of an XR system, likely because it is not yet a common occurrence in research literature. In one paper, which began to solve this problem, Herbert et al. defined a new term for intelligent tutors paired with AR technology. They named these systems “Augmented Reality Adaptive Tutors” (ARAT) (Herbert et al., 2018). However, this term is exclusive to AR systems, leaving out a large portion of relevant work in VR. Some of the other terms used in existing research include Intelligent Tutoring System (ITS), Computer-Assisted Instruction (CAI), adaptive Virtual Learning Environment (VLE), intelligent VLE, and adaptive tutor/training/simulator. While all these terms are used to describe instances of adaptive automation paired with a VE, some of the terms listed have more specific definitions. For example, an ITS is broadly used to describe systems, which provide a personalized learning experience by adapting to a learner's performance. ITSs are typically characterized by the use of a domain, student, and pedagogical model used to adapt to the educational needs of a learner (Wenger, 1987). However, like a VLE, ITS does not necessitate the use of XR technology. Both can be presented on a traditional computer system with a 2D monitor as well, making a literature search all the more difficult.

Without a suitable all-encompassing term from the literature to use, systems that combine adaptive automation qualities and XR technologies will be referred to as "adaptive XR/AR/VR/MR" (depending on the context). To fit the scope of this dissertation, the following subsections will present the use of adaptive automation within XR applications. These subsections will include training and non-training applications, since both provide relevant insight into the design and implementation of adaptive automation in XR.

Augmented Reality Training Applications

Herbert et al. reviewed existing literature on AR and ITS systems and came up with a set of guidelines for combining these two technologies. To do this, they created a two-dimensional taxonomy to describe the new design space with Milgram's Reality-Virtuality Continuum (Milgram et al., 1994) on one axis, and their own proposed levels of ITS on the other axis. They also proposed a three-part definition for this technology space, which they named Augmented Reality Adaptive Tutors (ARAT). They defined ARAT as having three key features:

1. Uses spatial information from the real-world to detect errors, provide feedback and/or sequence tasks using an ITS.
2. Uses Augmented Reality to enhance learning.

3. Creates context by using a combination of instructional cues and Augmented Reality. (Herbert et al., 2018)

The authors go on to detail considerations when choosing the technology architecture for a new ARAT system including authoring tools, system requirements, and choosing between non-intelligent and intelligent models (Herbert et al., 2018). However, they do not discuss the specific design of adaptation methods and triggers pertaining to AR. Furthermore, the authors identify that pedagogical design and evaluation for ARAT systems are important areas for future work. The following paragraphs will outline some of the existing applications-based research conducted using adaptive AR systems. Note that not all the examples given fit Herbert and his colleagues definition of ARAT.

The development of adaptive AR technology is a relatively new area of research. One of the early examples was published by LaViola et al. in 2015. The authors explored the challenges of integrating intelligent tutoring with augmented reality technology. Their primary concerns had to do with implementing real time occlusion in an AR training simulation. However, they also discussed their integration of an ITS architecture called the Generalized Intelligent Framework for Tutoring (GIFT) into their application to allow for adaptive feedback based on a learner's competency (LaViola et al., 2015). Their work ultimately resulted in a functioning proof of concept adaptive AR application for a military training system named ARWILD.

At about the same time that LaViola et al. were developing ARWILD, Radkowski, Herrema, and Oliver were developing an adaptive augmented reality aide for assembly tasks. Their assembly aide could adapt visual instructional elements based on the difficulty of the current step in the assembly. However, the results of this research were mixed. The study showed that simpler visual elements that used concrete text and animations resulted in higher performance on all tasks. This contradicted their hypothesis that concrete instructions would result in better performance on simpler tasks, and abstract instructions would increase performance on more difficult tasks (Radkowski et al., 2015). Although the hypothesis was not proven, their work represented a paradigm shift toward using adaptive AR to customize information to fit a user's current state.

Another example that used adaptive AR was a medical training tool developed by Almiyad et al. They created an adaptive AR tutor for training radiology students in a needle insertion procedure. The system provided adaptive text-based feedback when certain criteria were reached such as prolonged task completion time, or incorrect needle angle (Almiyad et al., 2017). Unfortunately, only preliminary results were provided, which did not allow for conclusions about the efficacy of the system.

Finally, in 2015, a team from the University of Canterbury studied the use of an ITS paired with adaptive AR on an HMD for the assembly of a motherboard. The application they built was able to provide adaptive feedback in the form of text and audio instructions that changed based on a user's performance. They compared the adaptive version to a version without feedback and found that using an adaptive system for training improved unassisted assembly scores by 25% and improved assembly speeds by 30% on average (Westerfield et al., 2015). This research represents one of the only studies that systematically evaluated the efficacy of adaptive AR for training. The results of this research showed the potential training advantages for combining AR and adaptive technologies.

Mixed Reality Training Applications

This subsection presents examples of Mixed Reality systems that employ adaptive technologies. For the purposes of organization and clarity, this section uses MR to describe systems that do not quite fit the definition of AR or VR but still represent didactic interaction with computer generated content. Specifically, the following examples feature haptic interaction with VEs, which are represented in 2D.

One example is the work by Cameirão et al., who systematically developed an arm rehabilitation tool for stroke patients (Cameirão et al., 2010). The application used an adaptive MR approach that tracked arm movement in real time while patients played a rehabilitation game on a 2D monitor. Their system used arm tracking data as an input for their personalized training module to manipulate the speed, interval, and range of falling objects in a virtual catching task. This research was unique because an adaptive training model was created systematically by conducting a full factorial analysis of four levels of four different game parameters. Ultimately, the researchers tested their training model by comparing treatment of healthy subjects to those recovering from a stroke. They found that the technology was well accepted by the users and that the model was able to differentiate between these two groups (Cameirão et al., 2010).

A similar study, by Duff et al., was the development and analysis of an XR tool for helping to rehabilitate upper arm movement in stroke victims as well (Duff et al., 2010). Their system used real time motion capture and analysis to track and adapt to a user's progress during various VR tasks presented on a 2D monitor. The system used a patient's kinematic performance to suggest feedback and level progression. In a two-week pilot test with three stroke rehabilitation patients, all participants significantly improved their arm movement capability using the system (Duff et al., 2010). Although the sample size for this study was small, it represents a step toward the practical application for adaptive XR technology in the field of physical rehabilitation.

Lécuyer, George, and Marchal studied adaptive MR for different application (Lecuyer et al., 2013). They developed a haptic maze task shown on a 2D monitor that used brain wave data from an EEG to monitor workload. The program evaluated the amount of workload and made the task easier or harder by providing haptic assistance to a user when the workload exceeded a certain threshold. The authors compared a user's performance using the adaptive automation with EEG to a version with no assistance and one with continual assistance. They found that performance was significantly better than the condition with no assistance and only marginally worse than continual assistance. Additionally, mental workload improved over the version with no assistance. Although the performance results were mixed, the adaptive brain-computer interface was able to successfully identify times of high mental workload and apply interventions accordingly. Concepts used in this, and other adaptive MR examples could also be applied toward the development of adaptive VR systems.

Virtual Reality Applications

Unlike AR, there are no prominent papers establishing nomenclature or taxonomy for VR systems combined with adaptive automation or similar technologies. However, many examples of adaptive VR implementation exist. although the evaluation of these systems to date has been limited.

One of the earliest examples of adaptive VR was the development of virtual agents for instruction and assistance during VR training. This concept was proposed by Rickel and Johnson (Rickel & Johnson, 1999), who developed a virtual human called Steve for military team training that could function as an instructor or as a team member in the absence of real teammates. Their concept was unique because Steve could adapt instructions and feedback based on the performance of a trainee. Additionally, Steve could demonstrate tasks and use body language to communicate with a trainee. Steve was also able to perform complex Naval training tasks (12 steps) both by itself and as a member of a heterogeneous team with both virtual agents and up to five human students (Rickel & Johnson, 1999). O'Hare et al. also employed virtual agents to facilitate adaptive training in VR (O'Hare et al., 2000). Their application, ECHOES, was created to support the development of adaptive multi-user VR learning environments. Students communicated with the agents, who then adapted the VE by presenting relevant course work depending on a student's previous progress, scores, and expertise (O'Hare et al., 2000). Lastly, Santos and Osório developed an adaptive virtual environment called AdapTIVE that employed a virtual agent assistant in addition to organizing information spatially using machine learning (Santos & Osório, 2004a). They tested their design in an e-commerce case study simulating a virtual bookstore. The categories of books were represented as rooms in the VE along with shelves designated for different sub-categories. Machine learning was used to update the proximity of relevant books (rooms) to a user based on their previous book selections. They also published work using a similar system to organize information in an education setting (Santos & Osório, 2004b).

Early adaptive VR systems often took the form of virtual agents, but adaptive VR can take on other forms as well. Aotake et al. developed an adaptive system for training nuclear power plant workers to identify and diagnose operational anomalies (Aotake et al., 1999). They used eye-tracking, heart rate, and utterances to predict the confidence and thought process of the trainee. Although the authors suggest that the training could be delivered in 3D using a VR HMD, the system only used 2D renderings. No evaluation of the effectiveness of the system was discussed, but the authors conducted user studies to determine the relationships between the physiological measures and the state of a trainee in the simulation. They found that blink burst rate could be used to gauge a learner’s confidence in their answers, and speech analysis could be used to classify their thought process (Aotake et al., 1999). This work showed the potential for applying physiological- and speech-based triggers to adaptive XR systems.

Some researchers also created systems using performance as a metric to trigger adaptations in VR. Buche et al. proposed a model called "MASCARET" for developing adaptive, multi-agent, VR training simulations. Their methodology employs models representing the domain, learner, errors, interface, and pedagogy (Buche et al., 2004). The authors described this model and demonstrated it by developing a multi-agent firefighting simulation called SecuReVi (Querrec et al., 2003). Their model was used to create simulations where a learner can experience one of several roles on a firefighting team, with the remaining roles performed by virtual agents. The agents could then adapt their own actions based on the actions of a learner. Additionally, the instructor could trigger events in the simulation resulting in a hybrid system using both adaptive and adaptable (human in the loop) methods. Their work showed proof of concept for this method but did not offer an evaluation as to the system’s effectiveness.

Other adaptive VR systems include several from the transportation domain. Specifically, this technology has been used for VR trainers in truck and automobile driving, flight simulators, and fluvial navigation (Fricoteaux et al., 2014; Lopez-Garate et al., 2008; Ludwig et al., 2005; Ropelato et al., 2018). For example, Lopez-Garate et al. proposed an adaptive VR truck driving training system. Their VR CAVE simulator could be customized by an instructor before use by a student. An instructor can give different weights to different types of mistakes and change the intrusiveness (frequency) of the feedback messages (Lopez-Garate et al., 2008). Similarly, Ropelato et al. developed a VR driving simulator combined with an intelligent tutoring system that adapted the subsequent simulation activities based on a user's performance on a previous task (Ropelato et al., 2018). Performance measures included vehicle position, lane deviation, head rotation (checking mirrors), vehicle state (indicator on/off), vehicle velocity, and reaction time (Ropelato et al., 2018). Another example of adaptive VR applied to the transportation industry is GULLIVER, which is an adaptive training tool for fluvial navigation (Fricoteaux et al., 2014). Specifically, the system monitored user heart rate, head position, interactions, and performance metrics to adapt the training situation to the learner's physiological condition and learning needs. For example, the system could add or remove obstacles to make the task more difficult or easy and show or hide HUD elements in the simulation to provide varied levels of instruction. These examples of adaptive VR are particularly interesting because they address the specific challenges of non-procedural tasks.

Another industry with several examples of adaptive VR is medicine. Developments in this field include both procedural and non-procedural applications. One of the non-procedural examples is from Rossol et al., who designed an adaptive training system for the use of electric wheelchairs (Rossol et al., 2011). Their system relied on clinicians to design the layout of the training environments. Then, Bayesian networks were used to predict levels of performance in different rehabilitative areas and suggest a subsequent training task, which was then implemented by a clinician (Rossol et al., 2011). This example represents an adaptable system, keeping a human in the loop during the decision-making process. In another application, Barzilay and Wolf created an adaptive VR trainer for upper limb rehabilitation. Their system used motion tracking and EMG plus artificial neural networks to create an inverse model of a patient and customize exercises to their needs. The system was successful and improved tricep performance by an average of 33% over the course of treatment (Barzilay & Wolf, 2013). Lastly, Siu et al. (2016) employed adaptive training for a surgical task using Virtual Reality. Their model used the learner’s performance on a basic motor skills task to gauge skill degradation and recommend training activities and schedule training. In a small-scale user study, they found that their adaptive VR trainer could accurately identify skill degradation in learners (Siu et al., 2016). The previous examples have shown that adaptive VR has promising applications in medical training as well as medical treatment and physical rehabilitation. The following paragraph will demonstrate its applications in the field of psychology.

One example of adaptive VR in psychology is a study by Ćosić et al. (Ćosić et al., 2011). They developed a VR adaptive training tool for developing stress resistance in soldiers. This system used visual and physiological measures of a soldiers' emotional and cognitive responses to select subsequent training videos that would have the most impact. This was intended to provide a personalized training program for each soldier. One drawback of their method was a baseline of physiological responses had to be established for each learner beforehand for the system to work properly. In another study, Chollet et al. developed an adaptive virtual reality tool to train people in public speaking skills. Their tool provided recommendations to improve performance before training and included virtual audience members who reacted to a user's performance in real time using body language cues like head nods or shakes and looking away or leaning forward. They compared the adaptive training tool to a condition with direct, real-time feedback, and a control with no real-time feedback. The results showed that all conditions improved user performance, but the adaptive condition was found to be more engaging, captivating, and challenging to the participants.

References

Almiyad, M. A., Oakden-Rayner, L., Weerasinghe, A., & Billinghurst, M. (2017). Intelligent Augmented Reality Tutoring for Physical Tasks with Medical Professionals BT - Artificial Intelligence in Education. In E. André, R. Baker, X. Hu, M. M. T. Rodrigo, & B. du Boulay (Eds.), International Conference on Artificial Intelligence in Education (pp. 450–454). Springer International Publishing.

Aotake, Y., Sasai, H., Ozawa, T., Fukushima, S., Shimoda, H., & Yoshikawa, H. (1999). New adaptive CAI system based on bio-informatic sensing: study on real-time method of analyzing ocular information by using eye-sensing HMD and method of adaptive CAI system configuration. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 3, 733–738. https://doi.org/10.1109/ICSMC.1999.823319

Baird, K. M., & Barfield, W. (1999). Evaluating the effectiveness of augmented reality displays for a manual assembly task. Virtual Reality, 4(4), 250–259. https://doi.org/10.1007/BF01421808

Barzilay, O., & Wolf, A. (2013). Adaptive rehabilitation games. Journal of Electromyography and Kinesiology, 23(1), 182–189. https://doi.org/10.1016/j.jelekin.2012.09.004

Botden, S. M. B. I., Buzink, S. N., Schijven, M. P., & Jakimowicz, J. J. (2007). Augmented versus virtual reality laparoscopic simulation: What is the difference? A comparison of the ProMIS augmented reality laparoscopic simulator versus LapSim virtual reality laparoscopic simulator. World Journal of Surgery, 31(4), 764–772. https://doi.org/10.1007/s00268-006-0724-y

Boud, A., Haniff, D., Baber, C., & Steiner, S. (1999). Virtual Reality and Augmented Reality as a Training Tool for Assembly Tasks. IEEE International Conference on Information Visualization, 32–36. https://doi.org/10.1109/IV.1999.781532

Buche, C., Querrec, R., de Loor, P., & Chevaillier, P. (2004). Mascaret: A pedagogical multi-agent system for virtual environments for training. Journal of Distance Education Technologies, 2(4), 41–61. https://www.igi-global.com/article/mascaret-pedagogical-multi-agent-system/1639

Buttussi, F., & Chittaro, L. (2018). Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. IEEE Transactions on Visualization and Computer Graphics, 24(2), 1063–1076. https://doi.org/10.1109/TVCG.2017.2653117

Cameirão, M. S., Badia, S. B. I., Oller, E. D., & Verschure, P. F. (2010). Neurorehabilitation using the virtual reality based Rehabilitation Gaming System: Methodology, design, psychometrics, usability and validation. Journal of NeuroEngineering and Rehabilitation, 7(1), 48. https://doi.org/10.1186/1743-0003-7-48

Checa, D., & Bustillo, A. (2020). A review of immersive virtual reality serious games to enhance learning and training. Multimedia Tools and Applications, 79(9–10), 5501–5527. https://doi.org/10.1007/s11042-019-08348-9

Chung, K. H., Shewchuk, J. P., & Williges, R. C. (2002). An analysis framework for applying virtual environment technology to manufacturing tasks. Human Factors and Ergonomics in Manufacturing & Service Industries, 12(4), 335–348. https://doi.org/10.1002/hfm.10023

Ćosić, K., Popović, S., Horvat, M., Kukolja, D., & Dropuljić, B. (2011). Virtual reality adaptive stimulation in stress resistance training. Proceedings of 2011 Mental Health and Well-Being across the Military Spectrum.

De Crescenzio, F., Fantini, M., Persiani, F., Di Stefano, L., Azzari, P., & Salti, S. (2011). Augmented reality for aircraft maintenance training and operations support. IEEE Computer Graphics and Applications, 31(1), 96–101. https://doi.org/10.1109/MCG.2011.4

Doshi, A., Smith, R. T., Thomas, B. H., & Bouras, C. (2017). Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. The International Journal of Advanced Manufacturing Technology, 89(5–8), 1279–1293. https://doi.org/10.1007/s00170-016-9164-5

Duff, M., Chen, Y., Attygalle, S., Herman, J., Sundaram, H., Qian, G., He, J., & Rikakis, T. (2010). An adaptive mixed reality training system for stroke rehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 18(5), 531–541. https://doi.org/10.1109/TNSRE.2010.2055061

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. Human Factors, 54(6), 1008–1024. https://doi.org/10.1177/0018720812443983

Fricoteaux, L., Thouvenin, I., & Mestre, D. (2014). GULLIVER: A decision-making system based on user observation for an adaptive training in informed virtual environments. Engineering Applications of Artificial Intelligence, 33, 47–57. https://doi.org/10.1016/j.engappai.2014.03.005

Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., & Tecchia, F. (2015). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. Interactive Learning Environments, 23(6), 778–798. https://doi.org/10.1080/10494820.2013.815221

Hall, C. R., Stiles, R. J., & Horwitz, C. D. (1998). Virtual reality for training: Evaluating knowledge retention. Proceedings - Virtual Reality Annual International Symposium, November, 184–189. https://doi.org/10.1109/vrais.1998.658488

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index). Advances in Psychology, 52, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Henderson, S., & Feiner, S. (2009). Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. Mixed and Augmented Reality, 2009.

Herbert, B., Barrett, E., Weerasinghe, A., Billinghurst, M., & Wigley, G. (2018). Design considerations for combining augmented reality with intelligent tutors. Computers & Graphics, 77, 166–182. https://doi.org/10.1016/j.cag.2018.09.017

Hoover, M., Miller, J., Gilbert, S., & Winer, E. (2020). Measuring the performance impact of using the microsoft HoloLens 1 to provide guided assembly work instructions. Journal of Computing and Information Science in Engineering, 20(6). https://doi.org/10.1115/1.4046006

Hořejší, P. (2015). Augmented reality system for virtual training of parts assembly. Procedia Engineering, 100(January), 699–706. https://doi.org/10.1016/j.proeng.2015.01.422

Hou, L., Wang, X., & Truijens, M. (2015). Using Augmented Reality to Facilitate Piping Assembly: An Experiment-Based Evaluation. Journal of Computing in Civil Engineering, 29(1), 05014007. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000344

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theoretical Issues in Ergonomics Science, 5(2), 113–153. https://doi.org/10.1080/1463922021000054335

Kaber, D. B., Riley, J. M., Tan, K.-W., & Endsley, M. R. (2001). On the Design of Adaptive Automation for Complex Systems. International Journal of Cognitive Ergonomics, 5(1), 37–57. https://doi.org/10.1207/s15327566ijce0501_3

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. Human Factors, 47(4), 730–741. https://doi.org/10.1518/001872005775570989

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2020). The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. Human Factors. https://doi.org/10.1177/0018720820904229

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an "aid" can (and should) go unused. Human Factors, 35(2), 221–242. https://doi.org/10.1177/001872089303500203

Lagu, A. V., Landry, S. J., & Yoo, H. sang. (2013). Adaptive function allocation stabilization and a comparison of trigger types and adaptation strategies. International Journal of Industrial Ergonomics, 43(5), 439–449. https://doi.org/10.1016/j.ergon.2013.02.006

LaViola, J. J., Williamson, B. M., Brooks, C., Veazanchin, S., Sottilare, R. A., & Garrity, P. (2015). Using Augmented Reality to Tutor Military Tasks in the Wild. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), December, 1–10.

Lecuyer, A., George, L., & Marchal, M. (2013). Toward adaptive VR simulators combining visual, haptic, and brain-computer interfaces. IEEE Computer Graphics and Applications, 33(5), 18–23. https://doi.org/10.1109/MCG.2013.80

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. In Human Factors (Vol. 46, Issue 1, pp. 50–80). Human Factors and Ergonomics Society. https://doi.org/10.1518/hfes.46.1.50_30392

Lele, A. (2013). Virtual reality and its military utility. Journal of Ambient Intelligence and Humanized Computing, 4(1), 17–26. https://doi.org/10.1007/s12652-011-0052-4

Link, E. A. (1931). Patent: Combination training device for student aviators and entertainment apparatus (Patent No. 1,825,462). In Google Patents (1,825,462). https://patents.google.com/patent/US1825462A/en

Loch, F., Quint, F., & Brishtel, I. (2016). Comparing video and augmented reality assistance in manual assembly. Proceedings - 12th International Conference on Intelligent Environments, IE 2016, 147–150. https://doi.org/10.1109/IE.2016.31

Lopez-Garate, M., Lozano-Rodero, A., & Matey, L. (2008). An Adaptive and Customizable Feedback System for VR- Based Training Simulators. Proceedings of 7th International Conference on Autonomous Agents and Multiagent Systems, 1635–1638. https://dl.acm.org/citation.cfm?id=1402944

Ludwig, J., Ramachandran, S., & Howse, W. R. (2005). Developing an adaptive intelligent flight trainer. ARMY RESEARCH INST FIELD UNIT FORT RUCKER AL.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1994). Augmented Reality: A class of displays on the reality-virtuality continuum. Telemanipulator and Telepresence Technologies, 2351, 282–292. http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=981543

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? International Journal of Human Computer Studies, 45(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

O'Hare, G., Sewell, K., Murphy, A., & Delahunty, T. (2000). ECHOES: An Immersive Training Experience. International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, 179–188. https://doi.org/10.1007/3-540-44595-1_17

Oppermann, R. (1994). Adaptive user support: ergonomic design of manually and automatically adaptable software. CRC Press.

Page, R. L. (2000). Brief History of Flight Simulation. SimTecT. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.5428&rep=rep1&type=pdf

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans., 30(3), 286–297. https://doi.org/10.1109/3468.844354

Peniche, A., Diaz, C., Trefftz, H., & Paramo, G. (2012). Combining virtual and augmented reality to improve the mechanical assembly training process in manufacturing. In Proceedings of the 6th WSEAS International Conference on Computer Engineering and Applications, and Proceedings of the 2012 American Conference on Applied Mathematics, 292-297. World Scientific and Engineering Academy.

Querrec, R., Buche, C., Maffre, E., & Chevaillier, P. (2003). SecuReVi : Virtual environments for fire fighting training. 5th Virtual Reality International Conference, 169–175. http://www.vectorcommand.com

Radkowski, R., Herrema, J., & Oliver, J. (2015). Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty. International Journal of Human-Computer Interaction, 31(5), 337–349. https://doi.org/10.1080/10447318.2014.994194

Rickel, J., & Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: perception, cognition, and motor control. Applied Artificial Intelligence, 13(4–5), 343–382. https://doi.org/10.1080/088395199117315

Ropelato, S., Zünd, F., Magnenat, S., Menozzi, M., & Sumner, R. W. (2018). Adaptive tutoring on a virtual reality driving simulator. International Series on Information Systems and Management in Creative EMedia (CreMedia), 2017(2), 12–17. https://doi.org/10.3929/ethz-b-000195951

Rossol, N., Cheng, I., Bischof, W. F., & Basu, A. (2011). A framework for adaptive training and games in virtual reality rehabilitation environments. Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry - VRCAI '11, 343. https://doi.org/10.1145/2087756.2087810

Rouse, W. B. (1976). Adaptive Allocation of Decision Making Responsibility between Supervisor and Computer. In Monitoring Behavior and Supervisory Control (pp. 295–306). Springer US. https://doi.org/10.1007/978-1-4684-2523-9_24

Santos, C. T. dos, & Osório, F. S. (2004a). AdapTIVE: An intelligent virtual environment and its application in e-commerce. Proceedings - International Computer Software and Applications Conference, 1, 468–473. https://doi.org/10.1109/cmpsac.2004.1342880

Santos, C. T. dos, & Osório, F. S. (2004b). An intelligent and adaptive virtual environment and its application in distance learning. Proceedings of the Working Conference on Advanced Visual Interfaces - AVI '04, 362–365. https://doi.org/10.1145/989863.989925

Sheridan, T. B. (2008). Risk, human error, and system resilience: Fundamental ideas. In Human Factors (Vol. 50, Issue 3, pp. 418–426). SAGE PublicationsSage CA: Los Angeles, CA. https://doi.org/10.1518/001872008X250773

Sheridan, T. B. (2011). Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: Distinctions and modes of adaptation. IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans, 41(4), 662–667. https://doi.org/10.1109/TSMCA.2010.2093888

Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators. https://apps.dtic.mil/sti/citations/ADA057655

Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V., & Johnson, W. W. (2018). Why human-autonomy teaming? Advances in Intelligent Systems and Computing, 586, 3–11. https://doi.org/10.1007/978-3-319-60642-2_1

Siu, K.-C., Best, B. J., Kim, J. W., Oleynikov, D., & Ritter, F. E. (2016). Adaptive Virtual Reality Training to Optimize Military Medical Skills Acquisition and Retention. Military Medicine, 181(5S), 214–220. https://doi.org/10.7205/milmed-d-15-00164

Smith, P. J., McCoy, C. E., & Layton, C. (1997). Brittleness in the design of cooperative problem-solving systems: The effects on user performance. IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans., 27(3), 360–371. https://doi.org/10.1109/3468.568744

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? Applied Ergonomics, 53, 190–202. https://doi.org/https://doi.org/10.1016/j.apergo.2015.09.013

Vora, J., Nair, S., Gramopadhye, A. K., Duchowski, A. T., Melloy, B. J., & Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training: Presence and comparison studies. Applied Ergonomics, 33(6), 559–570. https://doi.org/10.1016/S0003-6870(02)00039-X

Waller, D., & Miller, J. (1998). A desktop virtual environment trainer provides superior retention of a spatial assembly skill. CHI 98 Conference Summary on Human Factors in Computing Systems., April, 339–340. https://doi.org/10.1145/286498.286807

Weaver, K. A., Baumann, H., Starner, T., Iben, H., & Lawo, M. (2010). An empirical task analysis of warehouse order picking using head-mounted displays. Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10, 1695. https://doi.org/10.1145/1753326.1753580

Wenger, E. (1987). Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge. http://hal.in2p3.fr/TICE/hal-00703016v1

Westerfield, G., Mitrovic, A., & Billinghurst, M. (2015). Intelligent augmented reality training for motherboard assembly. International Journal of Artificial Intelligence in Education, 25(1), 157–172. https://doi.org/10.1007/s40593-014-0032-x

Wickens, C., Gordon, S., Liu, Y., & Lee, J. (1998). An introduction to human factors engineering. Longman New York. http://www.academia.edu/download/44059426/WickensEtAl-HFE-Ch4.PDF

Wiedenmaier, S., Oehme, O., Schmidt, L., & Luczak, H. (2003). Augmented reality (AR) for assembly processes design and experimental evaluation. International Journal of Human-Computer Interaction, 16(3), 497–514. https://doi.org/10.1207/S15327590IJHC1603

Yang, E., & Dorneich, M. C. (2018). Affect-Aware Adaptive Tutoring Based on Human–Automation Etiquette Strategies. Human Factors, 60(4), 510–526. https://doi.org/10.1177/0018720818765266

CHAPTER 3. RESEARCH QUESTIONS

In summation, researchers have proven the advantages of XR training and adaptive automation independently. More recently, researchers have endeavored to combine these two technologies to compound their benefits. Although many proof-of-concept examples exist, few studies have been conducted to prove their efficacy. Furthermore, design principles specific to adaptive XR systems do not currently exist in the literature, leading to confusion over best practices as well as nomenclature. XR researchers also often fail to explain their choices when designing adaptive automation components. This makes their results less replicable for future XR training applications. By employing inconsistent automation techniques, XR training applications can be rendered ineffective or even be detrimental to a learning experience. These problems are the motivation for the research questions posed in this dissertation. The first research question (RQ1) to be investigated is:

RQ1: When are adaptations needed by a leaner in an XR training task?

To answer this question, guidelines for the use of adaptive automation in combination with XR simulations and training will be developed. These guidelines will be derived from semi-structured interviews with XR training and simulation experts. To answer the research question, the author crafted interview questions to help elicit understanding of how expert instructors deal with confusion during XR training including:

RQ 1.1. What causes learners to become confused during VR/AR training?

RQ 1.2. How do you know when a learner is confused during VR/AR training?

RQ 1.3. When is intervention necessary, and what does it look like?

Once the first research question is answered and guidelines for the development of adaptive XR training systems set forth, the preliminary recommendations resulting from the first phase of research will be implemented and tested. An experiment comparing the resulting adaptive VR simulation to a traditional VR simulation with a human instructor will be conducted. The study will show whether the implementation of the adaptive XR guidelines were successful and a final list of guidelines will be put forth based on these results. In addition to the publication of these guidelines, the experiment in phase 2 will explore the following research questions:

RQ2: Can an automated adaptive feedback model supplant the need for a live instructor in an XR training system?

The true potential of XR training with adaptive feedback is the ability of a learner to reach a specified level of expertise without the constant supervision of an expert trainer. The data collected from the second study of this research will provide insights into specific metrics such as the success rate of task completion, task accuracy, and completion time to determine if automated adaptive feedback provided sufficient instruction for immediate (i.e., short term) learning.

RQ3: What specific automated adaptive feedback mechanisms/practices are required for XR training tasks?

XR training is unique in many aspects such as visual realism, presentation of information, and a user's sense of presence in a VE to perform the training. Automated adaptations can simply be applied with positive outcomes expected. Due to the uniqueness of XR technology, specific practices and mechanisms must be determined as to maximize the benefits of learners using the system. The helpfulness, intrusiveness, and preferred frequency of feedback implemented in the second phase of the research will be examined to gain these insights. Finally, the recommendations presented in the first phase of research will be adjusted based on these results.

The creation and validation of guidelines will help inform the future development of adaptive XR simulations and training. Ultimately, this research will help optimize the design of successful adaptive XR systems and, in turn, alleviate the burden placed on XR instructors to constantly monitor and offer feedback to learners. The benefits of this research will have the added effect of helping in the design of XR studies that must be conducted remotely due to the COVID-19 pandemic.

CHAPTER 4. SITUATIONAL AWARENESS METHODS IN VIRTUAL REALITY TRAINING: A SCOPING REVIEW

Kaitlyn Ouverson, Melynda Hoover, Stephen Gilbert, Michael Dorneich & Eliot Winer

Iowa State University, Ames, Iowa, USA

Modified from a manuscript published in the 2020 *Interservice/Industry Training, Simulation, and Education Conference*

Melynda Hoover's role in this paper included the background research on Virtual Reality (VR) simulation and training, and communication between instructors and learners who are immersed in virtual environments. She can also be credited with the creation of the matrix of natural and computer-mediated communication channels in various co-located training configurations (Table 1). Melynda was not responsible for the methodology used in the scoping review, however, she did collect and analyze the papers pertaining to adaptive VR solutions for situational awareness and the analysis and discussion of this category.

This paper was important to this research because it established the importance of situational awareness to the role of the XR trainer. It also set the groundwork for showing that the current processes used for XR training inhibits the situational awareness of the trainer, potentially degrading the quality of feedback received by the learner. Ultimately, this paper helped to justify the need for adaptive systems to augment the feedback provided by human instructors in XR training scenarios.

Abstract

In 2020, the US Military budget for Air Operations Training increased by $197.7m to accommodate additional virtual training, instructor pilots, and air support. These virtual trainers are essential for preparing warfighters for scenarios that are rare, dangerous, and complex. While virtual training has historically been conducted in costly and immobile “big box” simulators, they can now be deployed using consumer-grade immersive virtual reality (VR) head-mounted displays (HMDs). For example, Air Force maintenance airmen use VR HMDs to train on the C-130 due to savings of time and money over live training, without loss of training effectiveness. However, one challenge when using an HMD for training is giving the instructor complete awareness of what the learner is doing both in the virtual environment. Typically, instructors observe a learner’s progress in a simulation from a monitor that provides a window into the virtual environment. This window is missing affordances for interaction that make communicating with the learner difficult. The challenge of the instructor and learner’s different access to the virtual environment, and the resulting lack of situational awareness, can cause a disruption in communication and degrade learning outcomes. The authors propose that this could be mitigated using a number of techniques from existing research. This paper provides a scoping literature review to explore five potential solutions: asymmetric, symmetric, asynchronous, substitutional, and adaptive VR training systems. The authors evaluated each of these innovations in VR collaboration for its impact on 1) instructor-learner workspace awareness and 2) communication in VR simulations to guide industry and interservice training professionals. Results show that each of the current VR collaboration techniques has strengths and weaknesses, and understanding these trade-offs is crucial to derive the maximum benefit for a specific training task.

Introduction

Virtual Reality (VR) technology is seeing an increase in adoption for training applications since the release of consumer-level Head-Mounted Displays (HMDs) like the Oculus Rift and HTC Vive. These systems make high-quality VR simulations portable and affordable for the first time. The US military was quick to take notice and has invested heavily in VR HMDs and other simulation technology to train their warfighters. In 2020, the US Military is increasing spending on Air Operations Training by $197.7 million to accommodate additional virtual training, instructor pilots, and air support (Office of the Under Secretary of Defense, 2019). More and more frequently, this includes VR solutions such as HMDs for training C-130 airmen (Stancy Correll, 2020). Research and practical applications have shown that this technology can be extremely effective for training purposes while reducing the resources needed to conduct it (Boud et al., 1999; Kaplan et al., 2020; Lele, 2013). One way that VR HMD training can be improved is by enhancing situational awareness between the instructor and learner, e.g., a warfighter.

Current VR HMDs can impede natural communication that is essential to situational awareness and the learning process (Seymour et al., 2002). For example, the learner may not be able to hear important feedback from the instructor over sounds in the Virtual Environment (VE). Similarly, the instructor cannot see the VE in the same way as the learner, because they are often not using an HMD to view the simulation. Augmented Reality (AR) HMDs can yield similar situational awareness problems if the instructor is not also wearing a headset. All of these examples contribute to a decrease in situational awareness and loss of communication that can result in decreased learning outcomes. As a result, a learner may need to repeat training, perform more training, or go into the field unprepared. All of these consequences result in potential danger to a warfighter and/or additional time and resources to correct. This paper analyzes potential solutions to improve situational awareness of the instructor and learner during VR training applications and optimize the learning process using a scoping review. Scoping reviews are important for identifying and clarifying key characteristics and concepts, as well as gaps in the understanding of both (Munn et al., 2018). By categorizing and characterizing potential solutions from existing VR training literature, this research can recommend key areas of future study into bridging the situational awareness gap. The current scoping review is a necessary step to ensure that future research on the topic of virtual reality training focuses on viable solutions to situational awareness and yields positive training outcomes.

The following section presents a brief review of the history of simulator and VR training, followed by an inspection of the communication needs of the training group and their delivery channels. Next, the methodology of the review, and an analysis of the corpus will be delineated. Then the authors will describe the interaction themes uncovered during the review and relate these themes to training needs. Finally, other considerations that were revealed, as well as those areas that require future investigation will be addressed.

Background

In this paper, the authors will examine the instructor-learner communication paradigm in the context of using a VR HMD and make recommendations on how to facilitate effective communication between an instructor and learner during VR training based on examples from existing peer-reviewed literature. The following subsections will discuss relevant background information on VR simulation and training and situational awareness. This information will show how VR HMDs can inhibit key communication channels between the instructor and the learner and negatively affect learning.

VR Simulation and Training

VR is often thought of as one point of a continuum between reality and virtuality (Milgram et al., 1994). VR lands nearer to virtuality on this spectrum, meaning that the majority of the environment experienced by the user is computer-generated. This ability to create virtual environments (VEs) for training has revolutionized the world of training for scenarios that pose a high-risk or are expensive to recreate in real life.

The development of the Link Flight Trainer in 1931 (Link, 1931) marks the beginning of VR training. This novel machine allowed pilots to learn to "fly by instruments," providing accurate physical and instrument responses to the pilot's input, all while remaining safely on the ground. Many consider this the first example of virtual reality hardware. Since then, new hardware has been developed to accommodate a wide variety of VR applications.

Some VR training systems have used traditional computer displays for their training environments. However, these systems typically do not allow the user to see the environment in 3D, nor do they promote embodiment, or sensori-motoric engagement with training content, which has been identified as important to understanding and learning new content (Johnson-Glenberg et al., 2014). To allow embodied interaction with VR content, training systems can use CAVE Automatic Virtual Environments (CAVEs), which project stereoscopic images on large wall-like screens (Cruz-Neira et al., 1993). Yet another method is to use HMDs to present VR content. Each of these methods has its trade-offs such as cost, portability, and image quality that make them suitable for different VR training applications. However, in 2016, a new generation of VR HMDs was released by Oculus and HTC. These devices greatly improved upon the resolution, form factor, field of view, and price of earlier HMDs and made them a popular choice for VR training.

VR training has been used in a variety of contexts from air traffic control (Zachary et al., 1999) to training team skills (Ostrander et al., 2019). Some early HMD research also pointed to the ability of VR to reduce training time for tasks such as manual assembly and maintenance when compared to training using engineering drawings (Boud et al., 1999). Similarly, Seymour et al. found that VR training for a surgical task led to faster and more accurate procedures when compared to a control group who did not practice the procedure in VR (Seymour et al., 2002). VR HMDs showed promise as a tool for military training, not only because of their effectiveness (Kaplan et al., 2020) but also because of their unique ability to simulate dangerous wartime scenarios in a controlled way (Lele, 2013). For these reasons, the U.S. Military has put its trust behind simulator and VR training by investing millions of dollars toward their implementation (Office of the Under Secretary of Defense, 2019). Therefore, it is imperative that VR training is optimized to increase readiness and the efficacy of the warfighter.

Situational Awareness in VR Training

Situational awareness is an important component of decision making across a variety of contexts, including aircraft control and piloting (Endsley, 1995). In such contexts, as the novice learner is exposed to the virtual environment, the expert instructor needs to maintain situational awareness of the shared environment to offer targeted feedback and assistance (Gutwin & Greenberg, 2000). Such situational awareness is referred to as *workspace awareness*, or "the up-to-the-moment understanding of another person's interaction with a shared workspace… [which] involves knowledge about where others are working, what they are doing, and what they are going to do next" (Gutwin & Greenberg, 2002). The shared workspace in a training environment supports the cooperative task that composes the joint process of teaching (for the instructor) and learning (for the learner) the task being trained.

Proper mediation of communication is important for forming workspace awareness (Dourish & Bellotti, 1992; Gutwin & Greenberg, 2002). Mediation of communication is essential for coordinating remote use of technology, but it can go overlooked in co-located use of technology. When working face-to-face, a team can typically communicate and orient themselves to observe one another's actions in order to maintain workspace awareness (Gutwin & Greenberg, 2000). When using an HMD, however, the learner's senses are immersed in the VE, thus rendering computer mediation necessary – even when both users occupy the same physical space (Kraus & Kibsgaard, 2015). In HMD-based training, workspace awareness cues, such as gaze-direction indicators, must be incorporated into the VE. Such cues support the gathering of workspace awareness information. The instructor needs to be able to monitor the learner and be able to make moves to offer assistance when necessary, while the learner needs to be able to monitor the instructor, in turn (Gutwin & Greenberg, 2002).

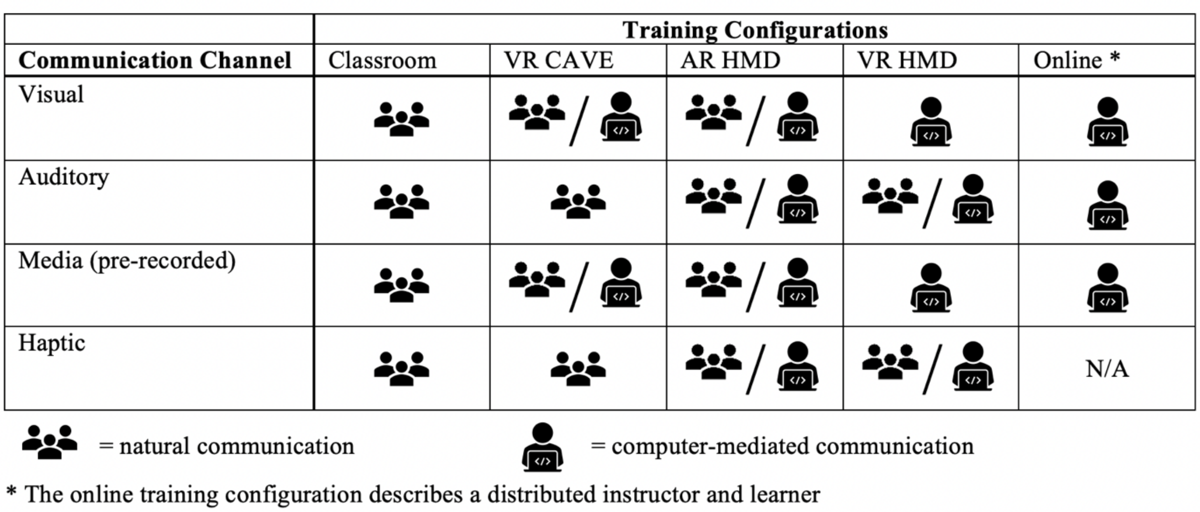
In many cases, this feedback from instructors is essential to the learning process. Kruglikova et al. (2010) found that surgical students who received instructor feedback in addition to VR simulation training achieved proficiency faster than those without instructor feedback. However, the use of HMD hardware can often inhibit communication between the learner and instructor, not only because the learner cannot see the instructor but also because they are experiencing two completely different realities.

Previous research into computer-mediated communication has organized the interaction methods along several axes, most popularly along the axes of time (*synchronous* vs. *asynchronous*) and place (*co-located* vs. *remote*) (Johansen, 1988). Additionally, Benford and colleagues (1998) use a multidimensional view of the virtuality-reality continuum (Milgram & Kishino, 1994) to describe "mixed-realities," or blended reality. In blended reality, two spaces (such as a VE presented in an HMD and the external world) are merged by a window (such as a computer monitor) at their boundaries. In their paper, they identify that communication is important to maintaining awareness in these asymmetric environments.

A typical VR HMD training configuration consists of a learner using an HMD and tracked controllers accompanied by a co-located instructor viewing the VE via a computer monitor, which serves as a window into the VE. This asymmetric experience can inhibit natural communication between the learner and the instructor because the HMD obscures many of the typical human-human communication channels (visual, auditory, and using media). Additionally, Kraus and Kibsgaard (Kraus & Kibsgaard, 2015) noted that computer-mediated communication methods are often of lower quality than unmediated methods. For example, computer-mediated auditory communication often suffers from low sound quality and latency. In contrast, computer-mediated visual communication can result in higher latency, reduced field-of-view, as well as limited stereo depth and color range. Therefore, natural communication is ideal for training, because it allows for the most situational awareness and highest fidelity. Training methods that do not support natural communication (alone or in addition to computer-mediate communication) risk barring critical feedback from the instructor and limiting the associated training benefits.

Table 4.1 shows the four communication channels typically used in a VE and how they can be transmitted (natural, computer-mediated, or a combination of both) in various co-located training configurations. To create this table, the authors evaluated the communication capabilities of off-the-shelf VR and AR hardware available before June 2020. Each training configuration was evaluated for its capabilities across four key communication channels: visual, auditory, haptic, and using pre-recorded media (Kraus & Kibsgaard, 2015). The VR and AR training configurations were then juxtaposed with traditional co-located classroom training on the left of Table 4.1, and distributed online training on the right. These configurations represent fully natural communication and fully computer-mediated communication scenarios, respectively. Additionally, online communication inhibits the use of haptic communication completely under most circumstances, and is, therefore, represented by "N/A" in Table 4.1. Table 4.1 shows that AR HMDs and VR CAVEs both facilitate natural communication in all four channels in addition to computer-mediated communication in some channels (as noted by the presence of both icons); these hardware configurations do not inhibit communication at all. AR HMDs and VR CAVEs can augment communication by providing new ways to communicate visually, using media, and, in the case of an AR HMD with tactile controllers, through haptics.

Table 4.1. Matrix of natural and computer-mediated communication channels in various co-located training configurations.



The ability to combine both natural and computer-mediated communication makes AR HMDs and VR CAVE systems powerful tools for training at the expense of presence, price, and mobility. Contrastingly, VR HMDs facilitate computer-mediated communication in all four channels, but they cannot provide natural communication visually or when using media. Such a disconnect can be detrimental to the overall communication experience when using VR HMDs because natural human-to-human communication requires great detail and typically takes place using multiple channels simultaneously (Kraus & Kibsgaard, 2015). Indeed, it is important to mediate channels that would otherwise be occluded for both learners and instructors during VR training using an HMD.

Methods

For the scoping review, peer-reviewed papers exemplifying methods of interaction between immersed and external HMD users relevant to VR HMD training were gathered following a systematic method Figure 4.1. To curate an overview of the topic area, the final corpus (a collection of papers) was sorted into themes, as described in the text below. This section will describe the research process in detail.

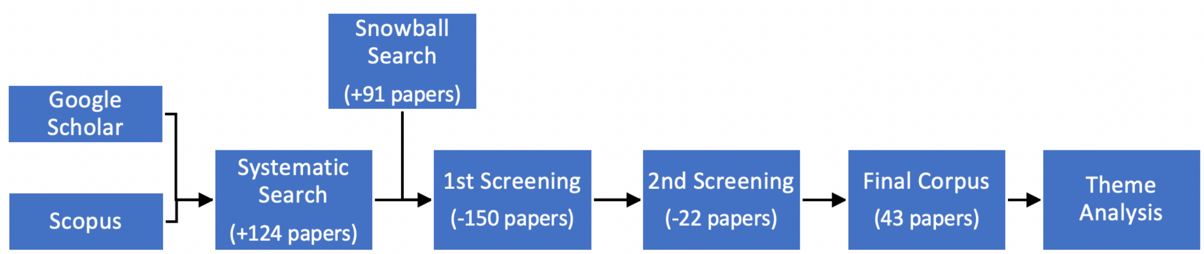


Figure 4.1. Flowchart describing the combined approach to the formation of the corpus of papers.

Using Scopus to retrieve published works, searches were limited to articles on adult humans published in the English language. The database was searched using a VR term (‘Virtual Reality’ or ‘VR’ or ‘Mixed Reality’ or ‘Head-Mounted Display’ or ‘HMD’) AND a collaboration term (‘collaborati*’ or ‘asymmetric technology’ or ‘cooperati*’ or ‘bystander’ or ‘multiple head-mounted displays’ or ‘multiple HMDs’ or ‘multi-user’ or ‘multiple user’) OR a location term (‘co-located’ or ‘collocated’ or ‘shared space’ or ‘same space’). A second search used Google Scholar to identify papers with adaptive systems terms (‘adaptive’ or ‘adaptive system’ or ‘intelligent tutor’ or ‘intelligent tutoring system’ or ‘computer assisted instruction’ or ‘informed virtual environment’) and VR training terms (‘Virtual Reality’ or ‘VR’ or ‘Mixed Reality’ or ‘Head-Mounted Display’ or ‘HMD’ or ‘simulation’ or ‘serious game’ or ‘virtual learning environment’).

To be included in the corpus, the articles were required to meet the following inclusion criteria: (i) include the use of an HMD, (ii) discuss communication or awareness and/or include a facilitator role in conjunction with HMD use, and (iii) discuss or test co-located use of said HMD or collect data at the level of the individual or group (i.e., not be a review, extended abstract, or workshop proposal). It should be noted that because simulator training and education are traditionally delivered face-to-face, the focus of this review is on co-located use of VR HMD training. Additionally, while HMDs prior to 2014 (when the Oculus Dev Kit 2 was released) were lower quality in terms of refresh rate and field of view and were less affordable than HMDs developed more recently, the year was not an exclusionary criterion in this survey. Insights from relevant articles over six years old are carefully extracted, taking care to caveat performance concerns while maintaining the findings relevant to group function in the co-located use of VR HMDs.

Using the keywords, the researchers identified 124 articles between December 2019 and April 2020. An additional 91 articles were identified and were added to the corpus through follow-up snowball searches based on the reference lists of relevant articles. Next, abstracts and titles were used to determine whether each article met the inclusion criteria. After this first screening, the corpus included 65 articles that were eligible for continued review. After applying the relevant exclusion criteria each paper's body in the second round of screening, the corpus totaled 43 papers spanning just over 20 years.

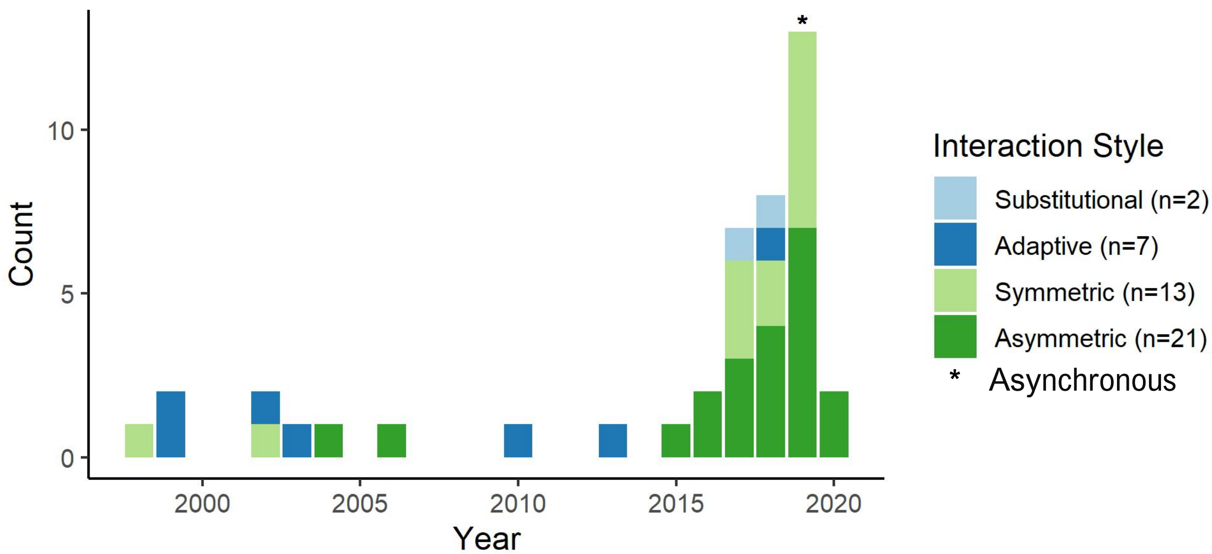


Figure 4.2. The distribution of identified themes by publication year. Asynchronous is double-counted, representing an asymmetric, asynchronous VR interaction style.

A valuable output of a scoping review is the list of themes that emerge from the analysis of the corpus. These themes aid the analyst in understanding the history of research in the field and often point to research gaps that need to be addressed in the future. Analysis of the corpus of papers resulted in five themes. Three themes were grounded in prior literature (introduced in the Background section, and discussed below) relating to computer-mediated communication and training: *asymmetric VR*, *symmetric VR*, and *adaptive VR*. Two were emergent themes:

asynchronous VR (after-action review) and *substitutional VR*. Figure 4.2, shows the count of studies categorized under each interaction style (identified themes) over the years.

Based on the literature surrounding VR use in training and computer-mediated communication, the authors began the review with three thematic categories, or themes, pertaining to interaction styles that can be used to expand instructor situational awareness in VR. *Symmetric VR*, or the use of the same technology by all parties accessing a single virtual environment, can be considered opposite *asymmetric VR*, which incorporates different devices into the collective experience. These categories are related to the work of Benford et al. (1998). They discussed ways to join mixed-realities at their boundaries and noted the impacts of improperly mediated asymmetric experiences on awareness. *Adaptive VR* systems use a feedback loop to change the system state (such as the VE) based on a stimulus (such as the user input) (Kelley, 1969). This automatic adaptation can reduce the need for workspace awareness on the part of the instructor while still providing customized feedback for the learner (Vaughan et al., 2016). Intelligent Tutoring Systems (ITSs) are a subset of adaptive systems that are used to customize learning feedback (Rickel & Johnson, 1997). ITS can be paired with VR training to create an adaptive Virtual Learning Environment and limit the need for instructor-learner communication, freeing up instructor time for other tasks, or potentially enabling them to instruct multiple learners simultaneously.

The remaining two themes were either new or unexpected, based on previous literature. *Substitutional VR* is a way to provide the details of the VE to non-immersed users (such as instructors) by aligning the physical space with the virtual workspace both by carefully matching virtual objects to their physical counterparts and projecting the VE onto the tracked space (Zenner et al., 2018). Substitutional VR is, therefore, a form of asymmetric VR, but it is separated because of its attention to passive haptics by the incorporation of the physical environment in the VE. The final method of VR HMD use which emerged during the review is known as *asynchronous VR*. While Johansen's (1988) framework included time (synchronous vs. asynchronous) as one way in which computer-mediated communication may be used, the recording of VR experiences to be consumed later (e.g., asynchronously) in VR is not yet a ubiquitous interaction technique and was not originally included as a theme. However, after identifying one such experience, the theme was added. The following sub-sections will summarize how each of these interaction styles in VR training attempts to solve the instructor-learner communication gap and increase situational awareness when using VR HMDs.

Results

This section will summarize the research on co-located HMD interaction and how these innovations tie into VR training. Specifically, the authors define the five themes of interaction styles in co-located VR (asynchronous VR, symmetric VR, asymmetric VR, substitutional VR, and adaptive VR) and offer insights relevant to conducting HMD-based training using each specific interaction style. Each of the first four themes describes methods that are meant to replace the original method of using a computer monitor as a window into the VE. The final theme, adaptive VR, is a method that can be paired with either the existing window-into-the-VE method or any of the other interaction styles to ameliorate the instructor's awareness.

Within this section, citations are representative of the literature belonging to any specific theme. The full corpus of papers is located in a public Mendeley library at https://www.mendeley.com/community/trainer-situational-awareness-methods-in-virtual-reality-a-scoping-review/. After presenting each interaction style in turn, the final section includes an interpretation of the state of the field for VR training and offers future directions for research in this area based on the findings of this review.

Asynchronous VR

One method of VR HMD use, which emerged during the review, is known as *asynchronous VR*. Johansen's (1988) framework, which is famous in group-based computer interaction circles, included place (remote vs. co-located) as well as time (synchronous vs. asynchronous). However, the recording of VR experiences to be consumed later (e.g., asynchronously) in VR is not yet a ubiquitous interaction technique. Chow et al. (2019) introduced a method for viewing previously recorded virtual interactions, wherein a learner could watch an expert's action in first-person view using an HMD of their own. Asynchronous VR is reminiscent of the after-action review from training literature (Zachary et al., 1999), as it could allow the instructor to join the learner in VR after the training has ended so that they may better illuminate the actions which require feedback. Another study that focused on the simultaneous review of 360-degree VR film used techniques such as view-locking to aid the group in workspace awareness (Nguyen et al., 2017). While that study focuses on the synchronous use of headsets, thus excluding it from the thematic area *asynchronous VR*, the interaction could be a useful method for after-action review, where the instructor guides the learner to observe and learn from their actions. Naturally, there are time costs to the described methods within this interaction style. The remaining themes describe the synchronous use of VR HMDs.

Symmetric VR

Introducing a second HMD comes with new challenges. First, safety should be a priority (Lacoche et al., 2017). In some training scenarios, the learner may be completely stationary, allowing the instructor to move around without worry; however, tracking errors may cause misjudgments of distances that could result in tripping or other accidents. Second, workspace awareness can be difficult to maintain. Visual cues are difficult to display accurately in VR (Nguyen et al., 2017), and without external motion-tracking technology, current algorithms can only approximate arm and leg positions. Relatedly, if there is any real-world aspect to the training, such as a mocked-up controls interface used as a prop to afford passive haptics, the instructor loses awareness for learner interactions with the prop when they enter an HMD. There are also limits to note-taking in VR, which need to be overcome for immersion to be truly useful in this situation (Nguyen et al., 2017). Lastly, verbal communication may be impacted when both training users are wearing VR HMDs, especially if they are required to use microphones rather than natural audio communication since this can introduce disruptive sound feedback-loops and echoes.

Asymmetric VR

The third theme is that of Asymmetric VR. Asymmetry can be used to encourage interdependence and to cater to the diversity of interaction preferences or needs (Gugenheimer et al., 2017). If two users of a shared VE have different goals, asymmetric VR can be used to reduce the cognitive resources needed to accomplish each user's goal by offering an interface that is tailored to those specific goals. For example, when used in therapeutic settings, the therapist and client have different requirements for effective interactions with the VE that can be supported by the therapist using peripheral interfaces, such as an interactive tablet, while the client is immersed in an HMD (Elvezio et al., 2018). Other authors use different interactions such as showing third-person views and giving external users touchscreen portals to the VE to aid in film design (Henrikson et al., 2016) and exploratory data analysis (Cavallo et al., 2019). By introducing a small screen tied (via code) to a tracked VR controller, Gugenheimer et al. (2017) create a series of VR games that provide engaging, high-presence experiences to non-HMD users of the environment. At the same time, these experiences provide position information which increases HMD-user awareness of outsiders, such as the instructor during HMD-based training.

Asymmetric VR is a clear contender for VR-based simulation training environments, due to its demonstrated ability to increase awareness of the virtual world (Chan & Minamizawa, 2017) and its alignment with the naturally dominant position of the instructor (Gugenheimer et al., 2018). Additionally, these innovations can aid in the spread of VR in the education context, offering options for the simultaneous awareness of multiple users at once, without necessarily immersing more participants into a shared VE (Albæk Thomsen et al., 2019; Chan & Minamizawa, 2017). With proper implementation in training, innovations in asymmetric VR could facilitate the embodied learning needs of the learner while also providing the instructor with the tools they need to deliver effective feedback.

Substitutional VR

Like other forms of asymmetric VR, substitutional VR is a way to provide the details of the VE to non-immersed users (such as instructors). What sets this form of asymmetric VR apart from the others is that it also aligns the physical space with the virtual workspace by carefully matching virtual objects to their physical counterparts (Zenner et al., 2018), thereby bringing the physical world and the virtual world closer together. Substitutional VR, therefore, forms its own emergent category in the analysis of the corpus. In substitutional VR, objects in the physical space that may have previously needed to be pushed to the edge of the room can be incorporated as parts of the VE (i.e., as a prop for passive haptics on a virtual table). To increase the engagement for HMD-outsiders (such as the instructor), a projection of the VE on the real space can make the tie between the two worlds more visible. In training, projection-based VR can be combined with immersive VR (as in Ishii et al., 2017) to allow an outside view of the world within the HMD without sacrificing the instructor's awareness of the physical world.

Adaptive VR

Contrasting with many of the other methods presented here, Adaptive VR does not require additional hardware and can be used to take on some of the workload of the instructor. The distribution of awareness responsibility can lessen the need for instructor-learner communication. Additionally, instructors may have different methods of intervening during training, based on their experience and preferences (George et al., 2019). Adaptive systems can help eliminate biases and effects caused by these individual differences by interrupting training more consistently.

The adaptive VR literature exemplifies two main methods that ITS can be used to communicate personalized training information to the learner in VR. The first method is using adaptive virtual agents. These virtual agents can serve multiple roles in the simulation, including an instructor, but also members of a virtual team working alongside the learner (Querrec et al., 2003). The system senses the student's learning state and adapts the feedback given by the virtual agent to their needs (Rickel & Johnson, 1999). The second and more prevalent method is adapting the training content itself. This method involves adapting the difficulty of the subsequent task based on the user's performance during previous tasks (Barzilay & Wolf, 2013;

Matsubara & Yamasaki, 2002). Self-adaptive systems are those which make these adaptations completely based on a learning model with no input from the instructor (Ropelato et al., 2018). Other systems keep the instructor in the loop by suggesting adaptations but keeping the impetus to intervene with the instructor (Ćosić et al., 2010). Though previous research in adaptive VR technology has shown that these systems are viable, few have evaluated their effectiveness when compared to traditional VR training. More research should be done to compare training transfer when ITSs partially or fully replace human instructors.

Furthermore, adaptive systems have the potential to be used to mitigate comfort, safety, and usability problems that are outside of the typical scope of ITS. In a VR training scenario, the instructor also serves roles that indirectly affect learning, such as showing the student how to use the VR hardware and helping them avoid collisions with environmental obstacles. These roles are not fulfilled by ITS, but other adaptive systems could be implemented to help solve these problems in VR**.** For example, an affect-aware adaptive system could detect and adjust to physiological symptoms of negative emotions and stress in VR (Saha et al., 2017). More research should be done to investigate other applications of adaptive technology on VR training.

Discussion and Conclusions

This paper has presented five co-located VR interaction themes that have been identified from a comprehensive literature search as potential methods to mitigate communication and shared workspace awareness problems in VEs. These interaction styles are useful for the transition to HMD-based VR training. By providing training in VR, and especially with responsive avatars that allow embodied interaction, more effective learning gains can be realized (Johnson-Glenberg et al., 2014). As VR HMDs have become cheaper and more robust, they grow more ubiquitous as tools for facilitating embodied training within virtual facsimiles of the real contexts. The instructor, however, has seen fewer advances in how they interact with the training material and the learner, typically observing a mirrored version of the learner's view of the scene from a nearby computer monitor. While studies on VR training do not traditionally consider the experience of the instructor with the simulation, research on collaboration in VR, reviewed here, offers insights that are useful to facilitating instructor workspace awareness in immersive training.

Each of the presented methods of instructor-learner interaction has different benefits and drawbacks. Within the VR HMD training configuration, there are various ways to deliver communication (computer-mediated compared to natural) between the instructor and the learner. For instance, while adding a second HMD for an instructor seems like an easy solution for providing full awareness access to the VE, research has shown that view congruency is still challenging to provide (Wong & Gutwin, 2014). It also cuts-off an instructor's perception of the physical space, which can be critical to completely understanding a learner's actions. Additionally, simultaneously accessing the VE with two HMDs is often not cost-effective. While asynchronous interaction allows the use of a single headset, it still means that one member of the group would be required to use a stationary 2D screen to access a 3D world and to engage in spatial reorientation to understand the other's actions. Alternatively, the instructor and learner may revisit the recorded experience to glean or give information, which means that while costs associated with procuring an additional HMD are cut, the time cost is potentially doubled. Projectors can be added to the physical set-up to enhance the outsider's understanding of the VE; however, space and money may not be sufficient to support these technologies. Instead, asymmetric VR which incorporates more typical technology such as computer monitors and keyboard/mouse set-ups or smartphones may be useful in mediating the interactions between the

HMD-user and external users. By allowing the instructor to interact with the VE using such devices, HMD-users may be guided, aided, or tested during their embodied learning experience. Lastly, adaptive training is an expanding area of research in VR that can be used to provide custom training material and feedback, lessening the need for communication and workspace awareness on the part of the instructor.

Based on this review of the literature, the authors believe the best solution to the workspace awareness problem in HMD-based training is often a combination of existing methods. For example, by combining adaptive VR's automation of performance feedback with asymmetric VR's unique interfaces, instructors may be able to supplement feedback in real-time due to their simultaneous awareness of the learner's actions in both the virtual and physical worlds. Another potential solution may combine asynchronous VR use with asymmetric and adaptive VR as a way to review the actions of the learner in a robust way that is not currently possible. More work is needed to evaluate such hybrid approaches.

In addition to future research into combined categories of interactions, more research is needed in five specific directions: (1) Interaction tasks should be categorized and analyzed to better understand which are best supported by different interfaces. This work has begun in areas such as joint-manipulation of virtual objects (Grandi et al., 2019). However, more attention to observation and interaction tasks that move beyond artifact-interaction is needed. (2) The training transfer should be evaluated for training in adaptive systems, which either include or exclude a human instructor in-the-loop. By better establishing the impacts of instructor interaction on training outcomes, development can be directed appropriately. (3) The overall effectiveness of both asymmetric and adaptive systems for training should be evaluated. Relatedly, (4) there is an additional need for research on the behavioral impacts of asymmetric

VR, especially due to its place in social VR interactions (Gugenheimer et al., 2017). Lastly, (5) future work should examine what aspects of the VR training experience should be automated to promote the highest efficiency of the HMD-based simulator training. As mentioned above, a combination of VR interaction approaches will likely be necessary to create the highest-quality VR training solutions.

This review aimed to survey the existing scope of research on VR use that is relevant to simulation training in immersive HMDs. Five themes of interaction styles were identified in the research to date and the literature's applicability to VR training in HMDs was discussed. This work will serve to embolden interservice, industry, and academic professionals interested in employing immersive VR and AR training to seek solutions to the challenges posed by this training-delivery method.

> This paper explored the different ways in which trainers and learners experience XR scenarios together, namely synchronously or asynchronously and symmetrically or asymmetrically. The paper explored the pros and cons of each of these methods and addressed how frequently each method is referenced in the literature. Ultimately, the research showed that the majority of trainer/learner teams experience the VE synchronously and asymmetrically. Lastly, the paper established the merits of using substitutional and adaptive methods to solve the issue of reduced situational awareness on the part of the trainer. The rest of this dissertation will focus on the practical implementation of adaptive automation to provide effective feedback when the trainer cannot.

References

Albæk Thomsen, L., Nilsson, N. C., Nordahl, R., & Lohmann, B. (2019). Asymmetric collaboration in virtual reality. Tidsskriftet Læring Og Medier (LOM), 12(20), 1–28. https://doi.org/10.7146/lom.v12i20.109391

Barzilay, O., & Wolf, A. (2013). Adaptive rehabilitation games. Journal of Electromyography and Kinesiology, 23(1), 182–189. https://doi.org/10.1016/j.jelekin.2012.09.004

Benford, S., Greenhalgh, C., Reynard, G., Brown, C., & Koleva, B. (1998). Understanding and Constructing Shared Spaces with Mixed-Reality Boundaries. ACM Transactions on Computer-Human Interaction, 5(3), 185–223. https://doi.org/10.1145/292834.292836

Cavallo, M., Dholakia, M., Havlena, M., Ocheltree, K., & Podlaseck, M. (2019, November). Immersive insights: A hybrid analytics system for collaborative exploratory data analysis. Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST. https://doi.org/10.1145/3359996.3364242

Chan, L., & Minamizawa, K. (2017). FrontFace: Facilitating Communication Between HMD Users and Outsiders Using Front-Facing-Screen HMDs. Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17, 1–5. https://doi.org/10.1145/3098279.3098548

Chow, K., Coyiuto, C., Nguyen, C., & Yoon, D. (2019). Challenges and design considerations for multimodal asynchronous collaboration in VR. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW). https://doi.org/10.1145/3359142

Ćosić, K., Popović, S., Kukolja, D., Horvat, M., & Dropuljić, B. (2010). Physiology-Driven Adaptive Virtual Reality Stimulation for Prevention and Treatment of Stress Related Disorders. Cyberpsychology, Behavior, and Social Networking, 13(1), 73–78. https://doi.org/10.1089/cyber.2009.0260

Elvezio, C., Ling, F., Liu, J.-S., & Feiner, S. (2018). Collaborative Virtual Reality for Low-Latency Interaction. The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings - UIST '18 Adjunct, 179–181. https://doi.org/10.1145/3266037.3271643

George, C., Janssen, P., Heuss, D., & Alt, F. (2019). Should I interrupt or not? Understanding interruptions in head-mounted display settings. DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference, 497–510. https://doi.org/10.1145/3322276.3322363

Grandi, J. G., Debarba, H. G., & Maciel, A. (2019). Characterizing Asymmetric Collaborative Interactions in Virtual and Augmented Realities. IEEE Conference on Virtual Reality and 3D User Interfaces (IEEEVR 2019), 2(March), 127–135. https://doi.org/10.1109/vr.2019.8798080

Gugenheimer, J., Stemasov, E., Frommel, J., & Rukzio, E. (2017). ShareVR: Enabling co-located experiences for virtual reality between HMD and Non-HMD users. Conference on Human Factors in Computing Systems - Proceedings, 2017-May(May), 4021–4033. https://doi.org/10.1145/3025453.3025683

Gugenheimer, J., Stemasov, E., Sareen, H., & Rukzio, E. (2018). FaceDisplay: Towards Asymmetric Multi-User Interaction for Nomadic Virtual Reality. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18. https://doi.org/10.1145/3173574.3173628

Henrikson, R., De Araujo, B., Chevalier, F., Singh, K., & Balakrishnan, R. (2016). Multi-device storyboards for cinematic narratives in VR. UIST 2016 - Proceedings of the 29th Annual Symposium on User Interface Software and Technology, 787–796. https://doi.org/10.1145/2984511.2984539

Ishii, A., Tsuruta, M., Suzuki, I., Nakamae, S., Minagawa, T., Suzuki, J., Ochiai, Y., Tsuruta, M., Minagawa, T., Ochiai, Y., Suzuki, I., & Suzuki, J. (2017). ReverseCAVE experience: Providing reverse perspectives for sharing VR experience. SIGGRAPH Asia 2017 VR Showcase, SA 2017, 2016–2017. https://doi.org/10.1145/3139468.3139482

Johansen, R. (1988). Groupware computer support for business teams. In Series in communication technology and society.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. Journal of Educational Psychology, 106(1), 86–104. https://doi.org/10.1037/a0034008

Kelley, C. R. (1969). What is Adaptive Training? Human Factors: The Journal of the Human Factors and Ergonomics Society, 11(6), 547–556. https://doi.org/10.1177/001872086901100602

Lacoche, J., Pallamin, N., Boggini, T., & Royan, J. (2017). Collaborators awareness for user cohabitation in co-located collaborative virtual environments. Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17, Part F1319, 1–9. https://doi.org/10.1145/3139131.3139142

Matsubara, Y., & Yamasaki, T. (2002). VR-based interactive learning environment for power plant operator. Proceedings - International Conference on Computers in Education, ICCE 2002, 922–923. https://doi.org/10.1109/CIE.2002.1186114

Nguyen, C., DiVerdi, S., Hertzmann, A., & Liu, F. (2017). CollaVR: Collaborative in-headset review for VR video. UIST 2017 - Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, 267–277. https://doi.org/10.1145/3126594.3126659

Querrec, R., Buche, C., Maffre, E., & Chevaillier, P. (2003). SecuReVi : Virtual environments for fire fighting training. 5th Virtual Reality International Conference, 169–175. http://www.vectorcommand.com

Rickel, J., & Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: perception, cognition, and motor control. Applied Artificial Intelligence, 13(4–5), 343–382. https://doi.org/10.1080/088395199117315

Rickel, J., & Johnson, W. L. (1997). Intelligent Tutoring in Virtual Reality: A Preliminary Report. Proceedings of the Eighth World Conference on AI in Education, 294–301. http:::www.isi.eduuisddVETvet.html

Ropelato, S., Zünd, F., Magnenat, S., Menozzi, M., & Sumner, R. W. (2018). Adaptive tutoring on a virtual reality driving simulator. International Series on Information Systems and Management in Creative EMedia (CreMedia), 2017(2), 12–17. https://doi.org/10.3929/ethz-b-000195951

Saha, D. P., Knapp, R. B., & Martin, T. L. (2017). Affective feedback in a virtual reality based intelligent supermarket. UbiComp/ISWC 2017 - Adjunct Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, 646–653. https://doi.org/10.1145/3123024.3124426

Vaughan, N., Gabrys, B., & Dubey, V. N. (2016). An overview of self-adaptive technologies within virtual reality training. In Computer Science Review (Vol. 22, pp. 65–87). Elsevier. https://doi.org/10.1016/j.cosrev.2016.09.001

Wong, N., & Gutwin, C. (2014). Support for deictic pointing in cues: Still fragmented after all these years? Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, 1377–1387. https://doi.org/10.1145/2531602.2531691

Zachary, W., Cannon-Bowers, J., Bilazarian, P., Krecker, D., Lardieri, P., & Burns, J. (1999). The Advanced Embedded Training System (AETS): An Intelligent Embedded Tutoring System for Tactical Team Training. International Journal of Artificial Intelligence in Education (IJAIED), 10, 257–277.

Zenner, A., Kosmalla, F., Speicher, M., Daiber, F., & Krüger, A. (2018). A Projection-Based Interface to Involve Semi-Immersed Users in Substitutional Realities. 2018 IEEE 4th Workshop on Everyday Virtual Reality (WEVR).

CHAPTER 5. DESIGNING ADAPTIVE EXTENDED REALITY TRAINING SYSTEMS BASED ON EXPERT INSTRUCTOR BEHAVIORS

Melynda Hoover & Eliot Winer

Iowa State University, Ames, Iowa, USA

Modified from a manuscript accepted to *IEEE Access* on September 25, 2021.

Abstract

Advances in training technologies using Extended Reality (XR) offer dramatic improvements in terms of time and associated costs over traditional training methods. Industries that are reliant on human labor are starting to embrace XR devices such as Virtual and Augmented Reality (VR/AR) Head-Mounted Displays (HMDs) to capitalize on these advantages. Additionally, developers now have the capacity to create customized user experiences and reduce instructor workload by employing adaptive automation in XR training simulations. Adaptive systems analyze the behavior of a learner for a trigger and then deploy an adaptation to alter the system state. While examples of adaptive XR training systems in current literature show that they are feasible to develop, further research must be done to quantify their efficacy. Furthermore, recommendations for how to design triggers and adaptations for XR training simulations are currently non-existent, resulting in negative impacts to training if inappropriate adaptations are implemented. This paper provides novel, evidence-based recommendations for the design of future adaptive XR training systems by learning from expert XR instructors. To this end, semi-structured interviews were conducted with 11 XR trainers. Participants were asked to discuss their experiences dealing with learners who exhibited confusion during XR training. Interviews were analyzed for existing and emerging themes. Finally, these themes were applied to existing trigger and adaptation models and synthesized into

design recommendations for XR training developers. The outcomes of this work will inform the future development of adaptive XR training platforms.

Keywords: Adaptive automation, augmented reality, extended reality, training, virtual reality

Introduction

Virtual Reality (VR) and Augmented Reality (AR) technologies use computer graphics to generate 3D content and simulated environments. VR does this by replacing the user's view of the real world with a completely artificial one. AR, on the other hand, supplements the user's view of the real world with additional computer-generated content, creating an environment that blends the two. Extended Reality (XR) encompasses both AR, VR, and more. It also includes spatial computing technology, Head Mounted Displays, projections systems, and non-invasive brain interfaces. These XR technologies are being applied to a myriad of industries, including training and education.

Many researchers have studied the impact XR training has on learners, also known as trainees. Many of these studies have found that the use of XR technology significantly increases learner performance metrics during and after training (Kaplan et al., 2020; Waller & Miller, 1998). XR systems were also found to decrease mental workload by providing relevant information in the proper training context (Loch et al., 2016). Furthermore, the time savings associated with increased performance and training speed, has been shown to decrease associated costs by up to 50% (Hou et al., 2015). However, one place where XR training is still lacking is in reducing the effort required of an instructor, also sometimes referred to as the trainer. XR instructors must maintain a high level of situational awareness and attention in order to give appropriate feedback during learning. However, this is inhibited by XR devices, such as an

HMD, because the instructor often cannot see the virtual environment as the learner does (Seymour et al., 2002). One way to remedy this problem is by supplanting or augmenting the role of the instructor with adaptive automation.

Adaptive automation is a system that actively adjusts the allocation of tasks between a human and a computer based on a trigger event (Rouse, 1976). Adaptive automation has shown promise as a tool for increasing learning outcomes in training (Kelley, 1969; Yang & Dorneich, 2018). By combining adaptive automation and XR to create an adaptive training system, greater performance gains can be made while reducing the workload placed on XR instructors (Vaughan et al., 2016). Several examples employing adaptive XR training systems have been published in fields ranging from military training to physical rehabilitation (Barzilay & Wolf, 2013; Rickel & Johnson, 1999). However, little research currently exists on the design of adaptive triggers, and how to choose the appropriate type of adaptation to best suit the needs of a learner in an XR training system. Additionally, if employed improperly, adaptive automation can lead to overreliance on the system, meaning that errors are likely to occur if the computer's logic fails, or automation is removed. Additionally, improperly designed automation can increase mental workload, making research in this area even more necessary (Smith et al., 1997).

This paper begins to fill the void of research in the design of triggers and adaptations for XR training by drawing from existing experts. For this work, 11 XR training experts were interviewed about their experiences providing feedback to learners who were immersed in virtual environments. Seeing as the population of VR training experts is relatively low, the sample size was determined to be sufficient. By applying their experience to existing adaptive system models, the authors identify common triggers and adaptation methods used by experts. Finally, the authors discuss how this information can be implemented in the development of future adaptive XR training systems by setting forth design recommendations.

Background

This section reviews previous research inherent to understanding adaptive XR training systems. The first is XR training, which has been thoroughly researched and shown to provide many benefits in terms of training and procedural task performance. The second is adaptive automation, the tradeoffs of which will be discussed and exemplified in the text. Lastly, the scarce amount of research combining these two technologies will be discussed along with the need for the current research.

XR Simulations for Training

XR technologies have rapidly evolved in the twenty-first century and are being embraced as a training tool in various industries. One reason is because XR simulations can be used to train for situations that are potentially hazardous or difficult to simulate in real life (Lele, 2013). Therefore, by using a controlled virtual simulation, learners can be better prepared for dangerous or uncommon circumstances when they arise. Additionally, applications such as manufacturing, assembly, and maintenance are frequently cited applications of XR training in literature and have been found to have dramatically positive effects on training outcomes (Baird & Barfield, 1999; Ong et al., 2008; Peniche et al., 2012). XR training has also been shown to improve performance and reduce training time for medical students learning to conduct various types of laparoscopic surgery (Ahlberg et al., 2007; Nagendran et al., 2013; Wang et al., 2014). Additionally, a recent review of studies that used VR HMDs for training and education found that existing research supports the use of this technology for training in psychomotor tasks such as surgery or search tasks and affective skills such as coping with fears and phobias (Jensen & Konradsen, 2018). These studies demonstrate the varied applications of XR training, but each industry has a different reason for choosing XR technology over traditional training mediums.

Many industries have adopted XR technologies because they offer unique advantages over traditional training methods. For example, the use of XR technology in training has been shown to improve a learner's ability to perform spatial assembly tasks and knowledge retention over time (Billinghurst & Dünser, 2012; Waller & Miller, 1998). Miller and Waller studied subjects who were trained to complete spatial circuit board assembly tasks using one of three training methods: desktop VR, paper, or video tutorials. They found that the VR training condition took more time to complete, but led to better performance retention over the course of two-weeks than the traditional training methods (Waller & Miller, 1998). Another study by Ganier, Hoareau, and Tisseau found that VR training could be successfully transferred to real-world tasks (2014). Their study tested 42 participants who were trained for a maintenance procedure on a control panel using one of three methods 1) VR training, 2) traditional training, and 3) no training. After training, all participants were asked to complete the task on a real control panel. They found that those trained in VR performed just as well as those trained using traditional methods (Ganier et al., 2014).

The most frequently proven advantage of XR training is performance. In one example, researchers studied the use of VR and AR for training in the assembly of an electronic actuator. They found that VR training was equally effective as traditional video training and AR training provided performance improvement over video training for the application in question (Gavish et al., 2015). Another study by Hou et al. compared traditional paper schematics to AR instructions in the context of a pipe fitting task (2015). Their study showed that the use of AR instructions reduced completion time and the number of errors by 50%. These reductions were then used to compute a 66% cost savings for stakeholders (Hou et al., 2015). In a more recent study, Doshi et al. found that using AR technology could increase industrial welding accuracy by as much as 52% (Doshi et al., 2017). Lastly, XR technologies like wireless HMDs have recently become cheaper and more portable. This allows training to occur where learners are located, instead of being limited to a specialized training facility.

Despite all of these potential benefits of XR training, there is still room for improvement when it comes to providing crucial feedback to learners who are immersed in virtual training simulations. In a study by Kruglikova et al., it was found that medical students using VR training equipment learned to perform an endoscopic colonoscopy more quickly when they received feedback from an instructor. Additionally, the number of failures due to perforations were reduced from seven in the non-feedback group to zero in the group that received feedback from instructors (2010). However, it can still be difficult for instructors to provide effective and timely feedback because observation of, and communication with, the learner is often hindered by an XR device, such as an HMD (Kraus & Kibsgaard, 2015). One solution to this problem is to use adaptive automation to monitor and provide customized feedback to the user. This will ultimately reduce the burden that is placed on the instructor while maintaining the benefits that feedback provides during XR training.

Adaptive Automation

In simplest terms, automation is the allocation of a task, once performed by a human, to a computer (Parasuraman et al., 2000). Adaptive automation goes one step further by sensing the state of a user and reallocating tasks between the human and the computer in real time with the goal of increasing overall performance of the task (Rouse, 1976).

While many different models of adaptive automation exist today, most are not specific to training. However, Kelley defined several key characteristics that all adaptive training systems share. He argued that they are all feedback loops including four basic elements: a stimulus, a human learner, a performance measurement, and adaptive logic (1969). Figure 5.1, which is adapted from a similar flow diagram by Feigh, Dorneich and Hayes (2012), shows how this process works in practice when combined with XR technology. In the figure, the stimulus is delivered to a user in the form of the virtual environment (VE) and aids in performance measurement by processing user inputs to the system. Next, adaptive logic is used to select a response. During this stage, the adaptive system identifies the need for automation using a trigger, which is activated when a performance measurement meets a certain threshold. Then, an adaptation method (also sometimes called an adaptation strategy) is employed based on the trigger and other system characteristics (Lagu et al., 2013). Finally, the system applies the

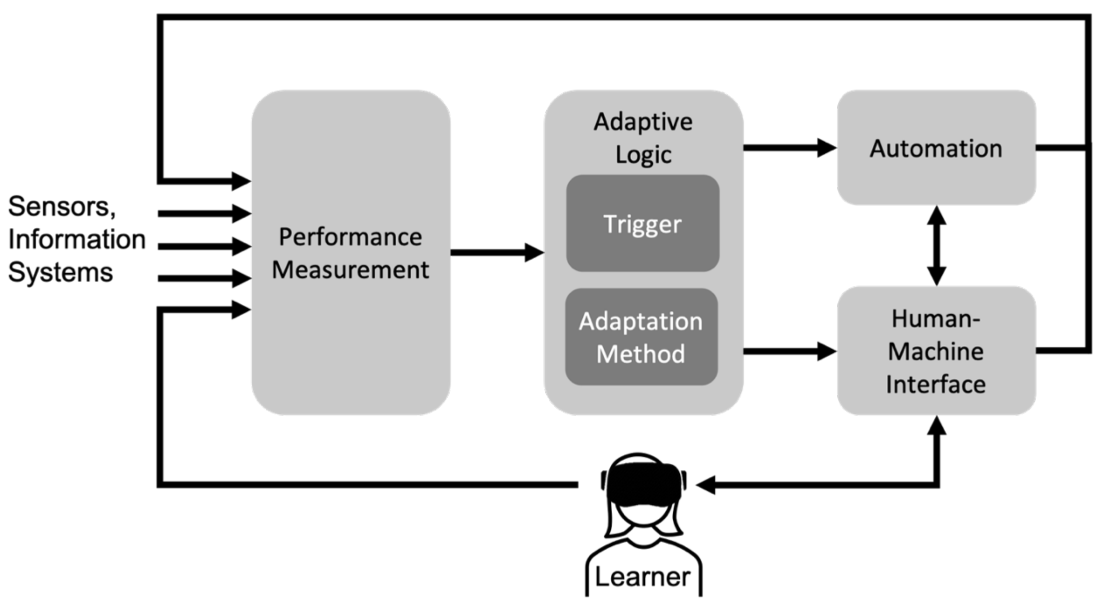


Figure 5.1. Self-adaptive system model acting on an XR learner. *Adapted from "Toward a characterization of adaptive systems: A framework for researchers and system designers," by K. M. Feigh, M.C. Dorneich and C. C. Hayes, 2012, Human Factors, 54(6), p.*

automation and repeats the process until the user's training goals are met or the session has ended. The adaptation made to produce this figure, from the original, is the inclusion of the learner wearing an XR device.

Many researchers have studied, and defined a concept called Levels of Automation (LOA). LOA defines how much control the computer has versus the human (Vagia et al., 2016) in an automated adaptive system. One of the most commonly used models of LOA has ten levels ranging from complete manual control at level 1 to complete computer control at level 10 (Sheridan & Verplank, 1978). This LOA model was later augmented by Parasuraman, Sheridan, and Wickens to include a second axis corresponding to the stages of human information processing: information acquisition, information analysis, decision selection, and action implementation (Parasuraman et al., 2000). Kaber et al. tested various levels of automation using an air traffic control task. They found that the human-automation team performed best when the human remained responsible for analytical processes like information analysis and decision making, while the computer acquired the information and executed the action dictated by the human (Kaber et al., 2005). Other research has found that the use of automation can lead to over-reliance and miscalculated trust, especially when the human believes the computer is faultless (Smith et al., 1997). This finding is especially important in the context of adaptive training, because learners may not have the help of automation after training is complete. Therefore, overreliance on automation could lead to failure in the field, making it even more important to carefully consider the design of triggers and adaptation methods for adaptive XR training.

Adaptive XR Training Systems

While adaptive automation is not completely new to the field of XR training, its efficacy for XR application is not well studied. Furthermore, a consistent naming convention for this combination of technology does not yet exist, making it difficult to find relevant information. Several researchers have developed VR training systems that incorporate adaptative elements, but few have studied its effects on training transfer or discussed the process of designing an adaptive XR training system. One of the earliest examples of adaptive training in XR was a system developed by Rickel and Johnson that used a virtual agent acting as either an instructor or a member of a training team for Naval operations (1999). The virtual agent, “Steve”, could provide verbal adaptive feedback and instructions based on the performance of the trainee and could give physical cues when it was the learner’s turn to act. Although no formal evaluation of the efficacy of this type of training was presented, learners were able to successfully complete an aircraft carrier maintenance task with the help of the adaptive virtual agent (Rickel & Johnson, 1999). Similarly, in 2003, Querrec et al. employed virtual agents for firefighter training in virtual reality (Querrec et al., 2003). In their system, the virtual agents filled other roles on the firefighting team while live instructor-observers could manipulate events in the VE. This system was unique at the time, because it allowed for both automated and human-in-the-loop adaptations (Querrec et al., 2003). Another promising study was conducted by Fricteaux, Thouvenin, and Mestre in 2014. Their work evaluated a 3D power wall simulator system for fluvial navigation that could manipulate task difficulty and user interface elements based on user input and computed performance metrics. They found that the adaptive system helped learners better predict and control the movement of the boat during the simulated training by 10% over non-adaptive training (Fricoteaux et al., 2014). Physical rehabilitation is another application where adaptive XR training has been shown to be effective (Barzilay & Wolf, 2013). In their experiment, Barzilay and Wolf used EMG, motion tracking and machine learning techniques to adapt upper arm rehabilitation to a patients’ needs. They found that their system resulted in a

33% increase in tricep performance (Barzilay & Wolf, 2013). However, they did not compare their system to one without adaptation. While many examples of adaptive XR training systems have been published, most are descriptive in nature, and lack rigorous evaluation into their effectiveness for training when compared to non-adaptive methods. Furthermore, none of these works describe the process used to design adaptations for XR training systems, leaving the rest of the community in the dark when attempting to design new and effective adaptive XR training systems. This gap in research can lead to the design of ineffective, and even detrimental, XR training systems due to overreliance and mistrust in the system. The research presented in this paper begins to fill this void in the current body of work by interviewing XR training experts and synthesizing adaptation design recommendations for the XR training community. These recommendations will allow training designers to more effectively combine these two powerful training technologies.

Methodology

In this research, the authors posit that drawing from the experiences of real XR training personnel can aid in the effective design of triggers and the selection of adaptation methods for adaptive XR training systems. In the following subsection, justification for the "instructor as an adaptive system” model will be provided. Then the data collection procedures and analysis methods will be discussed.

Instructor Behaviors as Models for Adaptive Automation

One way to explore the design considerations for adaptive automation within XR training is to examine the behavior of XR training instructors. In fact, Kelley describes adaptive training systems as “merely the automation of a function performed by a skilled instructor” (Kelley,

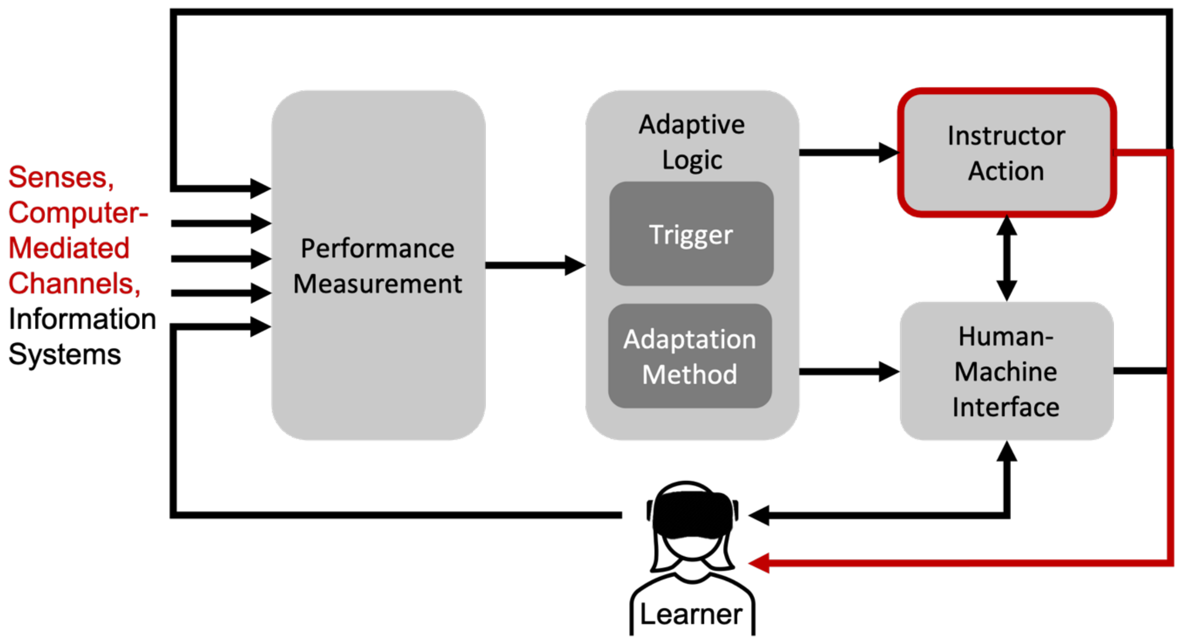


Figure 5.2. Model of instructor process for mitigating learner confusion in XR.

1969, p. 547). The process an instructor uses to provide feedback to learners (Figure 5.1) is very similar to the process shown in the adaptive automation model that was shown in Figure 5.2. The instructor perceives the current state of the learner and assesses their performance. The instructor then uses their knowledge of the system to identify triggers and select the proper adaptation. Lastly, they implement the adaptation either by communicating with the learner or by making a change to the system. However, differences arise when you look at how information is received by an instructor and translated into an adaptation for the learner. These differences are highlighted in red in Figure 5.2. The instructor observes the learner through their own senses, computer-mediated channels such as screen mirroring, and sometimes by viewing information from sensors within the system. The instructor can also choose to act directly on the user, on the simulation, or both, depending on the software features.

The similarities in the two models presented serve to illustrate how the themes gathered from the present interviews of XR training experts can be applied to an adaptive XR training framework. In the adaptive automation model, shown in Figure 5.1, triggers are used to determine when an adaptation is needed. This is analogous to a human instructor identifying signs of confusion in a learner. Furthermore, the adaptive automation model uses pre-programmed rules or artificial intelligence to select an adaptation method. Similarly, a human instructor uses the information at hand, and previous experience, to choose a mitigation strategy appropriate for the circumstances. Therefore, the methods used by an instructor may also be used to develop XR adaptive training systems. Using semi-structured interviews, the authors identified common triggers and adaptation methods used by adaptive XR training instructors. The data gathered was analyzed and used to make recommendations for the design of adaptive automation within the XR training domain.

Study Procedure

All procedures described in this section were approved by the Institutional Review Board at Iowa State University. The approval form for this research can be found in Appendix B: IRB Approval Form. Participants were recruited using word of mouth and emails to the authors' connections within the XR community. Additional candidates were recruited via referrals from participants. Before participating in the study, potential participants were asked to complete a screener survey to ensure they qualified to participate. Participants qualified if they indicated having previously facilitated training using AR or VR technology. In total, 17 people completed the screener survey. Because of the targeted recruitment tactics, all of the individuals who completed the screener qualified to be interviewed. Of the 17 instructors who qualified, 11 agreed to participate in an interview via Zoom video conferencing software. The sample size of 11 interviews was determined to be sufficient for this research for three reasons:

1. The authors were satisfied that they had reached a point at which no new themes were emerging, also known as a saturation point.
2. The size of the XR training instructor population is quite small since the technology is relatively new.
3. The goal of this research is to inform the development of future adaptive XR training research and not to draw inferential conclusions.

All interviews were conducted in April and May of 2020. The durations of the interviews ranged from 24 to 61 minutes with an average of 37 minutes per interview. The interviews had a semi-structured format. Participants were asked to describe their previous experience with XR training, and then prompted to answer specific questions about learners who experienced confusion during XR training. A list of interview questions was composed and used to address three research questions:

RQ1. What are common sources of confusion during XR training?
RQ2. How do instructors know that a trainee is confused during training?
RQ3. What do instructors do when they realized a trainee is confused?

A complete list of questions asked during the interviews can be found in Appendix A: Interview Questions. Additionally, follow up questions were posed when appropriate to gain a more detailed understanding of the participant's experience as it pertained to this research. Informed consent was collected from each interview participant. All the interviews were recorded and transcribed either manually or using the Rev.com transcription services.

Analysis

The first step in analyzing the transcripts was to create and implement a coding system. The R Qualitative Data Analysis (RQDA) package was used to apply codes to the data. General code categories that aligned with the research questions were created at the onset of the interviews with additional, more specific codes, developed as the coding progressed. The final code book consisted of 100 unique codes grouped into 15 categories.

The majority of coding was conducted by a single coder, while a second coder analyzed a subsample of the data (n=3) in order to calculate intercoder reliability. Transcripts were divided into paragraphs based on each new line of questioning during the interview. Relevant codes were then applied to each paragraph. To calculate reliability, each code was considered dichotomous (present or not present in the paragraph) and independent of the other codes. The reliability for each code was calculated using Cohen's Kappa for two raters with unweighted values (Cohen, 1960). Using this method, the agreement between the two coders across all codes was almost perfect ($\kappa=0.917$, $p<.0005$). This showed that the coding system was robust. In the following section the variable n is used to describe the number of interview participants who gave a response that fit the theme or had the characteristic described.

Results

Interview participants consisted of two women and nine men. They had varying amounts of experience with XR training ranging from 1 to 10 years with an average of 3 years. Participants also had experience with different forms of XR technology. Four of those interviewed had experience facilitating training with VR technology, one had experience with only AR technology and 6 had experience with both AR and VR training technology. The models of hardware used by the instructors varied as well. VR equipment used included the HTC

Vive, HTC Vive Pro, Oculus Rift, Oculus Go, Oculus Quest, Valve Index, HTC Cosmos, PlayStation VR, and Windows Mixed Reality devices. AR equipment used by the participants included the Microsoft HoloLens, HoloLens 2, Google Glass, and RealWear HMT-1. Additionally, one participant reported using a flight simulator system with a projection screen for training.

Table 5.1. Summary of interviewees.

ID*	Primary Training Group	Learner/Instructor Location	XR Hardware Used	Years of XR Training Experience
P1	Machine operators	Co-located	HTC Vive Pro	1
P2	Army aviators	Co-located	black-hawk flight simulator	4
P3	University students	Co-located	HTC Vive Pro, Oculus Quest, Oculus Go	4
P4	General public, Teachers	Distributed	PCVR, Windows Mixed Reality HMD	10
P5	Business executives & leadership	Co-located	Dell Toughbook, Google Glass, Microsoft Hololens	1
P6	HVAC technicians	Co-located	Oculus Rift S, HTC Vive	2
P7	Manufacturing engineers	Co-located	HTC Vive	3
P8	College professors	Distributed	Oculus Quest, HTC Vive, HTC Cosmos	2
P9	Military maintainers	Co-located	Valve Index	4
P10	Control room system operators	Co-located	Oculus Quest	2
P11	Assembly operators	Co-located	Google Glass Enterprise Edition 2, RealWear HMT-1, Microsoft HoloLens, Microsoft HoloLens 2	1

The participants had experience conducting XR training for various industries which fell into three main categories: manufacturing/maintenance (n=6), military (n=2), education (n=2), and demonstrations/research (n=1). Table 5.1 further summarizes the characteristics of the interview participants.

The following analysis is divided into three categories that describe the participants' experiences with managing confusion during XR training, the first step when determining the potential need for an adaptation. The first section will describe the causes of confusion, the second will list the ways that instructors identify confusion in learners, and the third section will show how instructors intervene to mitigate this confusion.

Sources of Confusion

When asked to recall common causes of confusion for learners during XR training, instructors had a wide range of answers that were often specific to the particular type of training they were facilitating. However, trends emerged that allowed the authors to categorize these sources of

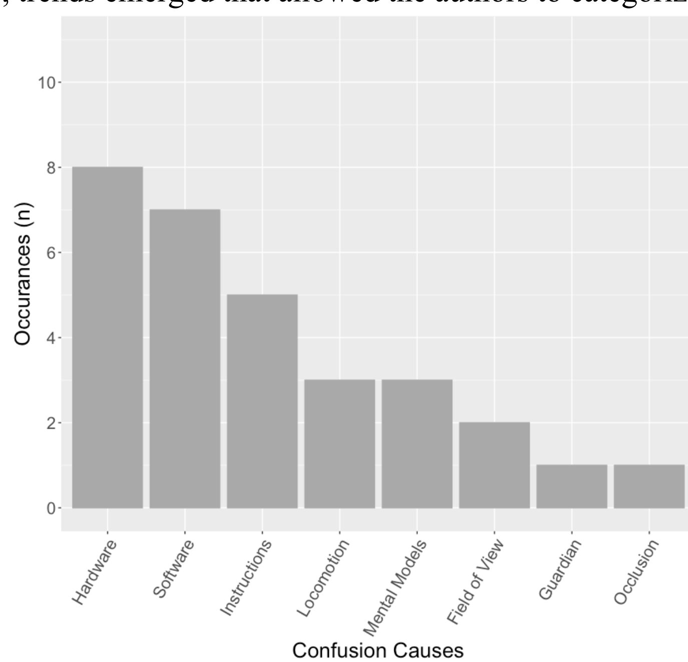


Figure 5.3. Bar chart of mentions of various confusion causes across all expert interviews.

confusion. The graph shown in Figure 5.3 depicts the frequency of different causes of confusion as reported by the participants. This chart, as well as subsequent charts, depict the number of distinct experts who cited each cause of confusion, not the total number of mentions summed across all interviews. This method was chosen in order to show the pervasiveness of each cause throughout the XR training community. From Figure 5.3, it is clear that hardware, software, and instructions were the largest drivers of learner confusion.

Hardware

The most commonly cited source of confusion was the hardware itself (n=8). Instructors often responded that learners became confused because the hardware did not work as intended or they did not know how to perform the required action using it. One expert, quoted below, noted how learners would become confused because they did not know how to use the VR controllers:

> "And we really started to notice that, especially for new people too, anything that we asked them to do beyond just that trigger, the difficulty they had just went up exponentially." (P9)

In another example, Participant 8 reported that a learner experienced confusion when the batteries in the controllers died and the buttons became unresponsive. These examples showed how learners became confused when they did not understand how to use their XR hardware or it did not work properly. Similarly, learners became confused or frustrated when their immersion was interrupted by the hardware. For example, immersion could be lost when the virtual guardian, which is used to render the boundary of the VR play area, was activated, when the learner came in contact with cords, or (in the case of AR) when the visual fidelity was compromised by light.

Software

Another frequently reported source of confusion was software not behaving as the developers of the simulation had intended (n=7). These instances occur for a variety of reasons including software bugs or user error. In her interview, Participant 1 described this phenomenon in the context of a battery manufacturing simulation:

> "If the system does something [that] we're not expecting it to do. For instance, if the instructions say that a battery is supposed to come down the line, and the battery doesn't come down the line." (P1)

These software "bugs" or "glitches" often caused a learner to miss information that was crucial to the completion of their training task. Some software issues also interfered with the learner's ability to interact with the environment. Instructors reported software issues that affected interactions with buttons in menus, natural user interfaces (such as picking up parts), and locomotion.

Instructions

A third notable theme that caused learner confusion was instructions (n=5). Confusion caused by instructions came in two forms: 1) misunderstanding the instructions and 2) not following them. The former often resulted when language used in the instructions was either too technical or not specific enough. For example, Participant 11 reported having to change relative directions like "forward" to objective ones like "clockwise" to ensure learner understanding.

Learners who read, but did not follow, the instructions did so for several reasons. First, they may not have seen or heard the instructions (e.g., if they were temporal in nature). Second, they may have executed an incorrect action that caused them to skip an instruction. Lastly, some instructors reported that learners purposefully attempted the task without reading the instructions.

In these cases, the learners felt that they could execute the task without the help of the instruction, such as in this example given by Participant 1:

> "They skip a step, which really shows me that they may not be reading. It could be something as simple as not going to get the cart to pull it near the machine… That tells me right away they didn't read the instructions, they're just going according to what they think is best." (P1)

Typically, learners proceeded without reading the instructions when they were already somewhat experienced with the task. It was in these situations that a learner would become confused if the simulation did not behave as they expected. This was often a result of the simulation behavior not matching the mental model of the learner.

Other Confusion Causes and Factors

Two other causes of confusion that came up during the interviews are not expressed as themes because of their low number of mentions but are worth noting. The first was visual obstructions such as limited field of view and occlusion of objects in the simulation (n=3). Visual obstructions resulted in confusion because learners could not see the part of the environment with which they needed to interact. The second had to do with learners' mental models of objects in the environment (n=3). In these cases, learners either didn't know they were supposed to interact with an object or didn't know how to properly interact with an object in the simulation. Finally, when asked about causes of confusion, several of the instructors mentioned other factors, which were not direct causes of confusion but could augment the feeling of confusion in the learner. These factors included discomfort (n=6) and inexperience with XR technology (n=4).

Identifying Confusion

The second goal of this research was to understand how instructors identify confusion in learners while they are immersed in an XR training environment. Based on the interview responses, all of the instructors (n=11) recalled noticing confusion in learners during training. Additionally, two of the instructors reported that they identified confusion following a training session through verbal feedback from the learners. The following subcategories will describe how instructors identified confusion during training rather than after.

Verbal

Methods of identifying confusion during XR training were varied and often included the observation of more than one identifying characteristic. The most frequently reported method of identifying confusion were verbal cues made by the learner (n=10). Learners used several different kinds of verbal cues to indicate their confusion as shown in the comments below:

> “In a face-to-face training or orientation, they're wearing a headset and I'm standing nearby, and they'll usually ask the question, "How do I grab this?"” (P8)

> “They'll just say, ‘What do I do next?’ Or, ‘What am I supposed to do now?’ or something along those lines.” (P3)

Sometimes direct statements were used to indicate the state of confusion as well as the source of confusion, such as in an example given by participant 8 (above). Other times, indirect statements, or vocalizations which indicated the state of confusion but not the source, were employed by learners such as in the example from Participant 3 (above). This placed the responsibility of finding the cause of the confusion on the instructor.

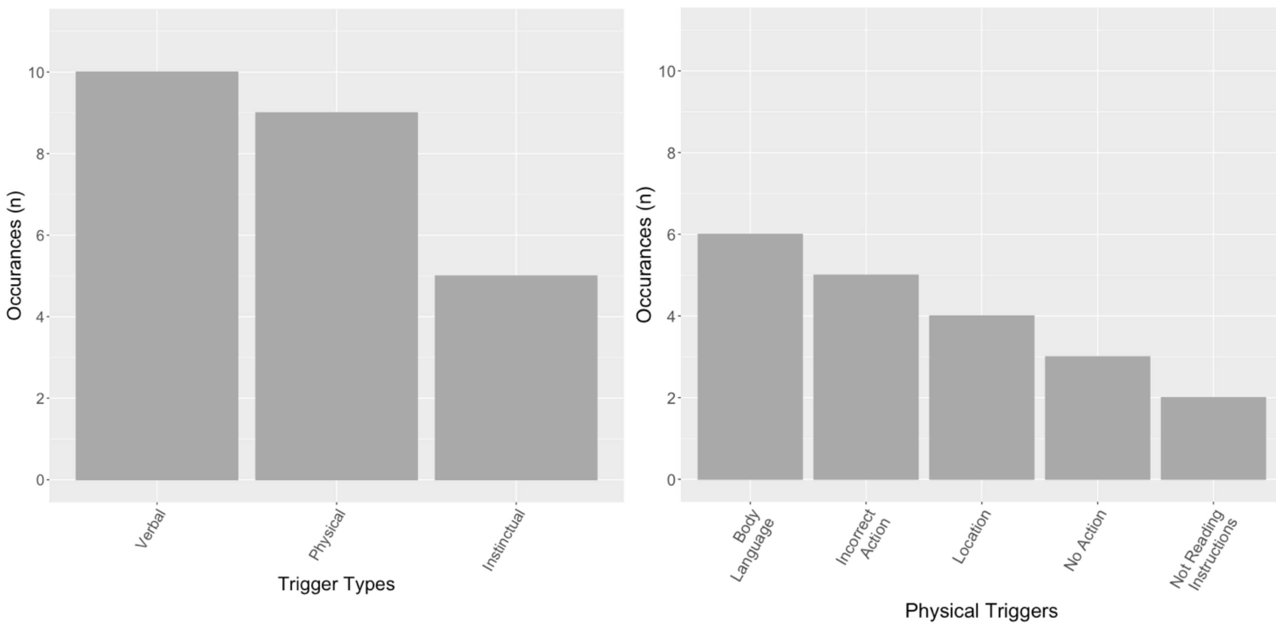


Figure 5.4. Bar chart of frequency of trigger types (left) and a breakdown of physical triggers (right) used by expert instructors to identify confusion in learners.

Verbal signs of confusion were conveyed either naturally or through computer-mediated channels. Natural communication occurred when instructors and learners were co-located. During distributed training, microphones and speakers were used to mediate verbal communication. If any part of this computer-mediated communication channel was interrupted, such as by muting of a learner's microphone, signs of confusion could go unnoticed by the instructor. Computer-mediated communication allowed learners to respond to verbal prompts from the instructor in new ways as well. One instructor (P4), who used an online virtual meeting space to facilitate training, reported that she could gauge understanding based on a learner's use of emojis in VR. She referred to these emojis as a type of virtual "body language" that she used to quickly gauge understanding and engagement among large groups of XR learners.

Physical

The data collected during interviews with XR training instructors showed that physical behavior of the learners could also indicate confusion (n=9). Some instructors cited specific body language characteristics that could indicate confusion such as increased head movements, or lack of body movements. However, the majority of physical indicators described by the participants were more specific to the training task.

Instructors often identified confusion when participants were exhibiting physical behaviors that did not aid in their progress toward the training goals. According to the instructors interviewed, behaviors that indicated the learner was not making progress toward the training goal included performing body language (n=6), incorrect actions (n=5), standing in the wrong location (n=4), not making an action (n=3), and not reading the instructions (n=2). Figure 5.4 illustrates these trends for physical indicators of confusion.

In order to notice these types of physical indicators of confusion, it was necessary for the instructor to have familiarity with the task and to monitor how the learner was progressing through the training. During VR training, instructors described observing the learner's actions in three ways: 1) by mirroring the learner's view on a 2D monitor, 2) by watching them in real life, and 3) by immersing themselves in the virtual environment with the learner. Instructors who were co-located with the learner typically used the first two methods, while distributed instructors and learners used the third method. AR instructors observed the learner's actions in the real world and did not report using mirroring or immersing themselves in the VE. In cases where computer-mediated observation was not used, the instructor had to be extremely familiar with the task and remain vigilant to accurately assess the learner's progress and identify confusion. Five of the instructors interviewed for this research described their ability to identify confusion in learners as a product of their previous experience with the training materials. The following quote from P7 illustrates the importance of having knowledge and experience of the simulation in order to identify confusion proactively:

> "The team of us created those tasks, and so in my head I know how they're working through and the steps that they're working through, so I can walk along with them as they do it." (P7)

This quote exemplifies the current requirement that XR training instructors must be experts in both the subject of the training and in operating XR hardware. This level of expertise takes significant time and effort to gain. By using adaptive automation, stakeholders can reduce the number of expert instructors needed and increase the number of learners who can be trained at any given time.

Adapting to Confusion

Once the instructors identified learner confusion, the next step was typically to employ a corrective action, also known as an adaptation, to help the learner meet their training goals. A variety of adaptation techniques were reported by the instructors and could be used alone or in combination to yield the desired training results. Figure 5.5 shows the most common adaptations ascertained from the interviews and their occurrences.

Verbal

The most frequently used adaptation method was to use a verbal intervention (n=11). Using this method, the instructor could confirm the source of confusion, and then provide a verbal adaptation to allow the learner to perform a corrective action. This allowed the learner to remedy their confusion autonomously and form mental models to prevent future confusion.

When instructors used verbal adaptation, they often used different strategies to help learners overcome confusion. Some instructors first asked clarifying questions to further understand the root of the problem before giving further instruction, such as in this example from P11:

> “They would say, "I don't understand what this means," and we'd say, "Okay, what step are you on?" I would validate what step I'm on from my documentation and we would discuss what they thought was confusing.” (P11)

Others directed the learner’s attention to key components of the simulation such as written instructions or user interface elements to address their confusion. In these cases, the

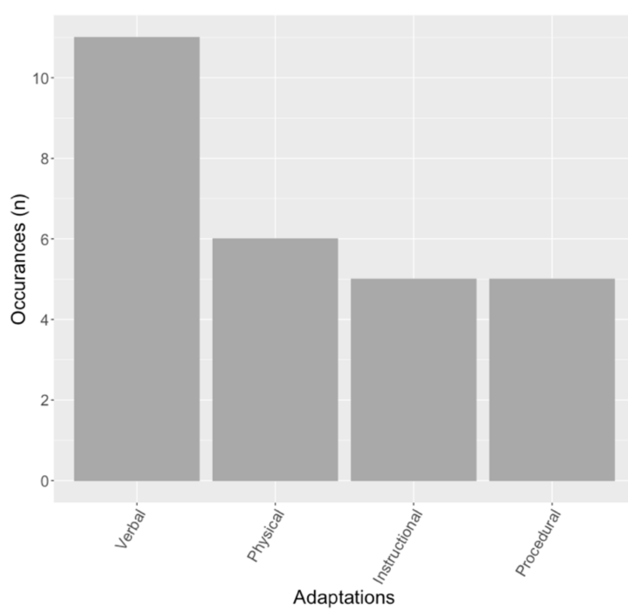


Figure 5.5. Instructor adaptation types used during XR training to mitigate learner confusion.

instructor did not give the learner any new information, but rather highlighted a part of the simulation that could help them reach the training goal. Lastly, many of the instructors supplemented the information given in the simulation with more detailed instructions.

Physical

The second most common adaptation method was to physically intervene with a learner (n=6). In these cases, the instructor used their body to show a learner where to go or how to interact with the training system. This type of adaptation was used in two different ways. The first, and most common, was for the instructor to guide the learner through touch. For example, instructors described touching a learner's hands to show them how to use XR controllers or leading them by the shoulders to stand in the correct position to execute a task. The second way instructors physically intervened was by making gestures. Two of the instructors reported pointing to areas of interest or waving to get a learner's attention. One instructor (P8) also described taking the controller from a learner and demonstrating the correct action for them. However, it should be noted that these methods only work when the learner can see the instructor (such as in AR) or a representative avatar (such as in symmetric VR).

Instructional

The third type of adaptation was to change the instructional material itself (n=5). This meant changing the simulation in an effort to eliminate the source of confusion rather than changing the behavior of the learner. Changes to the simulation took several forms including changing the written instructions, the interaction technique, or the entire task. Typically, changes to the task or interaction technique were made to create a simulation that more closely matched the learner's existing mental model.

Procedural

The final adaptation method used by expert instructors was procedural in nature (n=5). In this category, instructors changed the order, or duration, of the tasks presented in the training. This included pausing, restarting, or repeating all or part of the training simulation. For example, Participant 1 recalled having to restart a training simulation after experiencing a software issue:

> "That can be really confusing for a new person, because we might have to restart the whole program and have them start from the beginning, which means that we may have spent a lot of time already in there…" (P1)

In the experiences recalled by the experts, procedural adaptations did not affect the material presented in the simulation. The same material was presented again or resumed after a break. It is also notable that the instructor's goal when applying procedural adaptations was not to reinforce training goals, but rather to combat software malfunctions or learner fatigue and frustration.

Incremental Adaptations

Early in the interview process it became clear that instructors often use more than one adaptation method to mitigate confusion, depending on the circumstances. Analyses showed that instructors did not always apply the intervention that would solve the problem the fastest, but rather increased the amount of intervention incrementally (n=6). Multiple instructors first used a verbal intervention before escalating to a physical intervention such as in the following example given by Participant 3:

> “So, we have one of the headphones propped open, and then we'll just try to verbally explain it, and if not just let them know, "okay I'm going to touch your hands," and then place their finger on the button they should be using so that they understand where it is.” (P3)

This allowed instructors to apply the least intrusive adaptation first, observe the outcome, and apply a more aggressive adaptation, if necessary. This also had the advantage of allowing a learner to potentially remedy their confusion somewhat independently.

Discussion

Determining when and how instructors adapt to mitigate confusion is the first step in an informed design process for adaptive training systems. By looking at the way instructors deal with learner confusion in XR through the lens of adaptive systems, instructors can increase the usability of their simulations and decrease the need for observation during XR training. This discussion section is divided into three sub-sections. The first will discuss treating disparate causes of confusion in XR training, along with the feasibility of mitigating them using adaptive automation. The second section will describe how the information from the interviews can be used to design triggers in adaptive XR training simulations. Lastly, how adaptive methods used by human instructors can be applied to adaptive methods within closed-loop XR training simulation will be discussed.

Treating Disparate Causes of Confusion

Causes of confusion were shown to vary widely based on the interview data because of different training goals and conditions. The present work generalized the sources of confusion and summarized them based on factors that were common among all, or most, of the XR training scenarios: hardware, software, and instructions. Another way to explain confusion is through the gulf of evaluation and gulfs of execution (Norman, 2013). Gulfs of evaluation occur when a learner does not know what task they are meant to execute. Gulfs of execution occur when a learner does not know how to execute a task. Each of the causes of confusion reported by instructors can be categorized into these two groups.

Gulfs of Evaluation

A gulf of evaluation occurs when a user's goal is unclear. During the interviews, this was best exemplified by confusion caused by the instructions in the simulation. Gulfs of evaluation also occurred when software bugs prevented instructions, or other pertinent information, from being conveyed to the learner. In all of these cases, the learner did not have enough information to know what task to complete, let alone how to complete it. Therefore, when treating confusion caused by instructions (or lack thereof) the gulf of evaluation can be bridged by providing the missing information. In many cases, the problem can be solved in the development stage by increasing specificity, changing the wording of instructions, or presenting the information through a more salient channel. Adaptive automation can be helpful for solving the gulf of evaluation when the XR simulation is meant to provide varying levels of difficulty to the learner. For example, adaptive automation can be used to provide increasingly detailed instructions as the system detects inaction or incorrect actions performed by the learner.

Gulfs of Execution

A gulf of execution occurs when the goal is clear, but the user does not possess sufficient information to make progress toward that goal. During this research, it became clear that confusion caused by system hardware often resulted in gulfs of execution. In these cases, learners knew their intended goal, but the hardware prevented them from executing the requisite actions. For example, not understanding how to use the controllers, dead batteries, or loss of tracking all interfered with the execution of key actions. Similarly, confusion derived from software bugs resulted in gulfs of execution when learners were prevented from interacting with the environment. While some of these causes of confusion could be solved by providing more information regarding system interaction, others require more disruptive interventions such as system restarts or recalibration. This can make confusion resulting from gulfs of execution difficult for an adaptive system to diagnose and remedy. However, it is still possible to use adaptive automation to remedy confusion caused by gulfs of execution as long as ample signals are still reaching the system. For example, a learner who does not know how to use the controllers to execute a task may be identified using a trigger such as incorrect button presses. The system could adapt by providing the learner with more in-context information about how to use the controllers. The following subsections discuss recommendations for the design of triggers and adaptations that can help overcome both gulfs of evaluation and execution.

Implications for Triggers

Many of the methods that XR instructors used to identify learner confusion can be replaced by adaptive automation in the form of triggers. However, the system must also be able to receive the necessary signals in order to recognize these triggers. The following paragraphs discuss how the data collected during the expert interviews can be used to develop more effective triggers for adaptive XR training simulations.

Verbal

Since most modern HMDs include built-in microphones, verbal triggers are easily extensible to adaptive XR systems. The microphones can be used to capture verbal signs of confusion in the form of utterances or using certain words. One disadvantage of using XR technology to identify verbal triggers is the wide variety of languages that learners may use. Therefore, coding triggers to specific words or phrases can result in false positives. Artificial Intelligence or natural language processing tools may be used to design triggers that more adequately identify confusion using verbal signals. Furthermore, without a human instructor present, it is unknown if a learner will be as likely to use verbal signals to indicate confusion. Simply informing the user to the presence of verbal cues or asking them to use an activation phrase similar to when using voice assistants (such as Apple's Siri, or Amazon's Alexa) may mitigate this effect and make verbal triggers more effective.

Physical

Physical signs of confusion can be used to create adaptive triggers in XR as well. A detailed picture of the physical environment can be obtained using head and hand position data, button presses, and tracking interactions with objects in the simulation. This data can be used to determine if the learner is performing an incorrect action, not making an action, or standing in the wrong location. For example, a trigger could be designed so that an adaptation is applied if the learner is not standing in proximity to the object required for the task for a specified period of time. Similarly, a trigger for not reading the instructions could be created by checking to see if the learner's head position was oriented towards the text for an adequate length of time. Eye tracking data from HMDs such as the HTC Vive Pro Eye could also be used to increase the precision of the measurements and make the trigger more accurate. AR HMDs like the Microsoft HoloLens and Magic Leap are at a disadvantage for identifying physical triggers because they do not continuously track hand positions. Therefore, adaptive systems that obtain input using these devices would have trouble identifying triggers that rely on the handling of objects in the simulation.

Implications for Adaptation Methods

The adaptation methods used by human instructors during XR training can be used to inform the design of adaptive methods within XR training systems. As long as the XR device has the ability to output stimuli in ways that are similar to a human instructor (e.g., provide audible feedback or text to a learner), it is possible to emulate these behaviors in an adaptive system.

Verbal

The most common adaptive method used by expert instructors when combatting confusion in XR was verbal intervention. If the cause of the confusion can be reliably identified by the system using the measurements from XR input devices, the system can apply verbal adaptations using pre-recorded messages or text-to-speech software. Similarly, a voice assistant interface could be implemented to answer common questions asked during training. The same information could also be presented visually using text instructions to supplant or supplement audio cues. By using multiple channels, the information can be reinforced to the learner for more effective retention.

Physical

Physical interventions present a unique set of challenges when designing for adaptive systems. Most notably, current HMDs have limited abilities to apply physical stimuli on a learner besides applying vibration to the controllers. Therefore, adaptations designed to imitate physical interventions used by instructors must be supplemented by other sensory channels. For example, instead of physically touching a learner’s hand to communicate how to use a controller, an adaptive XR system could use a visual substitution to highlight the buttons on a model of the controller in a simulation.

Instructional

Instructional interventions are the most easily extensible to current XR technology because they involve an instructor acting on a human-machine interface. Since the human-machine interface (in this case the HMD) is already a part of the system flow, the adaptive system can act on the interface as the human instructor would. For example, instructors reported changing the wording of instructions as a confusion mitigation technique. Similarly, when triggered, an adaptive system could alter the specificity of the text, using more details for novices and fewer for expert learners to test their skills. Adaptive automation has the added advantage of being able to implement changes to the VE more quickly, saving time and effort normally exerted by instructors and developers. Additionally, adaptative automation can replace the need for constant observation by an expert instructor, saving time and increasing training throughput.

Procedural

Instructors use procedural adaptations by altering when and how learners experience a simulation. Typically, this is done by restarting the simulation, which can be replicated in an adaptive system by reloading the scene to pre-selected checkpoints. However, the interviews showed that this type of intervention is often used by instructors when the learner is experiencing confusion from hardware malfunctions. In these cases, the software or hardware has failed, and must be restarted by the instructor, rendering adaptations ineffective. However, if the failure is not catastrophic, or the failure is isolated to a certain feature of the simulation, adaptive methods could be integrated to help the learner troubleshoot the problem. For example, the simulation could use long periods of inaction coupled with input from the learner to identify a malfunction and direct a simulation restart. Additionally, features that allow the software to save the user's progress should be implemented to avoid repeating training.

Conclusions

In the present paper, the authors have summarized the need for adaptive XR training as well as the need for guidance in the design of adaptive XR training software. By using data from interviews with XR training experts from various fields, developers can make more informed decisions when designing triggers and choosing adaptation types. Furthermore, research should be conducted to fully evaluate the efficacy of each of the XR trigger and adaptation methods concluded from this research.

In summation, the authors recommend the following guidelines as a starting point for the design of adaptive triggers in XR training:

1. Verbal triggers should be prioritized when designing adaptive XR training systems. However, measures should be taken to inform and promote the use of vocalization from the learner, especially when a human instructor is not present.
2. Physical triggers should also be of high priority, especially in XR systems where appropriate stimuli can be applied to a learner. These include monitoring for signals of inaction, incorrect action, and body position.

When designing adaptation methods for XR training, the authors make the following recommendations:

1. Verbal and textual adaptations should be prioritized to get relevant information to the learner and allow them to correct their own behavior.
2. Physical adaptations can be replicated using models and animations in the VE. They should be prioritized to indicate points of interest in the VE, visualize hardware interactions, and even demonstrate the correct performance of the task.
3. Incremental adaptations can be used to allow the learner to think through a problem on their own.

4. Instructional changes can be used to increase or decrease the amount of instruction given to the learner. This is useful for adapting to different learners' experiences and creating varying levels of difficulty.
5. Procedural adaptations are of lower priority and involve the restart of the application or repetition of some, or all, tasks in the XR simulation. They can be used to reinforce a training concept, or to help the learner troubleshoot a problem with the system.

It should be noted that the recommendations in this work have not yet been tested in live adaptive XR training systems, nor does this constitute an exhaustive list of adaptation types and triggers. Therefore, designers and developers using these recommendations to inform the design of their XR training systems should conduct their own testing to ensure that the adaptations are effective for their system. Furthermore, continued research should be done to evaluate the accuracy of various triggers in adaptive XR training, as well as the efficacy of different adaptation types so that a standard can be created and dispersed to the XR training community.

References

Ahlberg, G., Enochsson, L., Gallagher, A. G., Hedman, L., Hogman, C., McClusky, D. A., Ramel, S., Smith, C. D., & Arvidsson, D. (2007). Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. American Journal of Surgery, 193(6), 797–804. https://doi.org/10.1016/j.amjsurg.2006.06.050

Baird, K. M., & Barfield, W. (1999). Evaluating the effectiveness of augmented reality displays for a manual assembly task. Virtual Reality, 4(4), 250–259. https://doi.org/10.1007/BF01421808

Barzilay, O., & Wolf, A. (2013). Adaptive rehabilitation games. Journal of Electromyography and Kinesiology, 23(1), 182–189. https://doi.org/10.1016/j.jelekin.2012.09.004

Billinghurst, M., & Dünser, A. (2012). Augmented reality in the classroom. Computer, 45(7), 56–63. https://doi.org/10.1109/MC.2012.111

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

Doshi, A., Smith, R. T., Thomas, B. H., & Bouras, C. (2017). Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. The International Journal of Advanced Manufacturing Technology, 89(5–8), 1279–1293. https://doi.org/10.1007/s00170-016-9164-5

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. Human Factors, 54(6), 1008–1024. https://doi.org/10.1177/0018720812443983

Fricoteaux, L., Thouvenin, I., & Mestre, D. (2014). GULLIVER: A decision-making system based on user observation for an adaptive training in informed virtual environments. Engineering Applications of Artificial Intelligence, 33, 47–57. https://doi.org/10.1016/j.engappai.2014.03.005

Ganier, F., Hoareau, C., & Tisseau, J. (2014). Evaluation of procedural learning transfer from a virtual environment to a real situation: A case study on tank maintenance training. Ergonomics, 57(6), 828–843. https://doi.org/10.1080/00140139.2014.899628

Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., & Tecchia, F. (2015). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. Interactive Learning Environments, 23(6), 778–798. https://doi.org/10.1080/10494820.2013.815221

Hou, L., Wang, X., & Truijens, M. (2015). Using Augmented Reality to Facilitate Piping Assembly: An Experiment-Based Evaluation. Journal of Computing in Civil Engineering, 29(1), 05014007. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000344

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. Education and Information Technologies, 23(4), 1515–1529. https://doi.org/10.1007/s10639-017-9676-0

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. Human Factors, 47(4), 730–741. https://doi.org/10.1518/001872005775570989

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2020). The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. Human Factors. https://doi.org/10.1177/0018720820904229

Kelley, C. R. (1969). What is Adaptive Training? Human Factors: The Journal of the Human Factors and Ergonomics Society, 11(6), 547–556. https://doi.org/10.1177/001872086901100602

Kraus, M., & Kibsgaard, M. (2015). A Classification of Human-to-Human Communication during the Use of Immersive Teleoperation Interfaces. Proceedings of the 2015 Virtual Reality International Conference - VRIC '15, 08-10-Apri, 1–8. https://doi.org/10.1145/2806173.2806198

Kruglikova, I., Grantcharov, T. P., Drewes, A. M., & Funch-Jensen, P. (2010). The impact of constructive feedback on training in gastrointestinal endoscopy using high-fidelity virtual-reality simulation: A randomised controlled trial. Gut, 59(2), 181–185. https://doi.org/10.1136/gut.2009.191825

Lagu, A. V., Landry, S. J., & Yoo, H. sang. (2013). Adaptive function allocation stabilization and a comparison of trigger types and adaptation strategies. International Journal of Industrial Ergonomics, 43(5), 439–449. https://doi.org/10.1016/j.ergon.2013.02.006

Lele, A. (2013). Virtual reality and its military utility. Journal of Ambient Intelligence and Humanized Computing, 4(1), 17–26. https://doi.org/10.1007/s12652-011-0052-4

Loch, F., Quint, F., & Brishtel, I. (2016). Comparing video and augmented reality assistance in manual assembly. Proceedings - 12th International Conference on Intelligent Environments, IE 2016, 147–150. https://doi.org/10.1109/IE.2016.31

Nagendran, M., Gurusamy, K. S., Aggarwal, R., Loizidou, M., & Davidson, B. R. (2013). Virtual reality training for surgical trainees in laparoscopic surgery. In Cochrane Database of Systematic Reviews (Vol. 2013, Issue 8). John Wiley and Sons Ltd. https://doi.org/10.1002/14651858.CD006575.pub3

Norman, D. (2013). The Design of Everyday Things: Revised and Expanded Edition. Basic Books.

Ong, S. K., Yuan, M. L., & Nee, A. Y. C. (2008). Augmented reality applications in manufacturing: a survey. International Journal of Production Research, 46(10), 2707–2742. https://doi.org/10.1080/00207540601064773

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans., 30(3), 286–297. https://doi.org/10.1109/3468.844354

Peniche, A., Diaz, C., Trefftz, H., & Paramo, G. (2012). Combining virtual and augmented reality to improve the mechanical assembly training process in manufacturing. In Proceedings of the 6th WSEAS International Conference on Computer Engineering and Applications, and Proceedings of the 2012 American Conference on Applied Mathematics, 292-297. World Scientific and Engineering Academy.

Querrec, R., Buche, C., Maffre, E., & Chevaillier, P. (2003). SecuReVi : Virtual environments for fire fighting training. 5th Virtual Reality International Conference, 169–175. http://www.vectorcommand.com

Rickel, J., & Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: perception, cognition, and motor control. Applied Artificial Intelligence, 13(4–5), 343–382. https://doi.org/10.1080/088395199117315

Rouse, W. B. (1976). Adaptive Allocation of Decision Making Responsibility between Supervisor and Computer. In Monitoring Behavior and Supervisory Control (pp. 295–306). Springer US. https://doi.org/10.1007/978-1-4684-2523-9_24

Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., Satava, R. M., Pellegrini, C. A., Sachdeva, A. K., Meakins, J. L., & Blumgart, L. H. (2002). Virtual reality training improves operating room performance results of a randomized, double-blinded study. Annals of Surgery, 236(4), 458–464. https://doi.org/10.1097/00000658-200210000-00008

Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators. https://apps.dtic.mil/sti/citations/ADA057655

Smith, P. J., McCoy, C. E., & Layton, C. (1997). Brittleness in the design of cooperative problem-solving systems: The effects on user performance. IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans., 27(3), 360–371. https://doi.org/10.1109/3468.568744

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? Applied Ergonomics, 53, 190–202. https://doi.org/https://doi.org/10.1016/j.apergo.2015.09.013

Vaughan, N., Gabrys, B., & Dubey, V. N. (2016). An overview of self-adaptive technologies within virtual reality training. In Computer Science Review (Vol. 22, pp. 65–87). Elsevier. https://doi.org/10.1016/j.cosrev.2016.09.001

Waller, D., & Miller, J. (1998). A desktop virtual environment trainer provides superior retention of a spatial assembly skill. CHI 98 Conference Summary on Human Factors in Computing Systems., April, 339–340. https://doi.org/10.1145/286498.286807

Wang, Z., Ni, Y., Zhang, Y., Jin, X., Xia, Q., & Wang, H. (2014). Laparoscopic varicocelectomy: Virtual reality training and learning curve. Journal of the Society of Laparoendoscopic Surgeons, 18(3). https://doi.org/10.4293/JSLS.2014.00258

Yang, E., & Dorneich, M. C. (2018). Affect-Aware Adaptive Tutoring Based on Human–Automation Etiquette Strategies. Human Factors, 60(4), 510–526. https://doi.org/10.1177/0018720818765266

Appendix A: Interview Questions

1. Background

    1.1. What kinds of Virtual or Augmented Reality training have you facilitated?

    1.2. What was the task?

    1.3. What was the learning objective?

    1.4. How many people were trained?

    1.5. Did the trainees have any previous experience with VR/AR?

    1.6. What were the demographics of the trainees?

2. Setting and Methodology

    2.1. Tell me about the last time you facilitated a VR/AR training session with a trainee.

    2.2. What hardware was used?

    2.3. Where did the training take place?

    2.4. Was there any introductory training to use the hardware?

    2.5. How long did the training take?

    2.6. How did you monitor the progress of the trainee?

    2.7. How were the tasks conveyed to the trainee during the simulation?

    2.8. How did you communicate with the trainee while they were immersed in the Virtual Environment?

    2.9. What were the pain points for the trainee during the simulation?

3. RQ1 - Causes of Confusion

    3.1. Tell me about a time when you had to help a trainee who was confused in VR/AR.

    3.2. What was the source of their confusion?

3.3. How often does this scenario result in confusion?

3.4. What other examples of confusion during VR/AR training can you tell me about.

4. RQ2 - Recognizing Confusion

4.1. How did you know that the trainee was confused during the training?

4.2. What do/don't they do?

4.3. What do they say?

4.4. What other cues tell you that there is a problem?

4.5. What did you do when you realized the trainee was confused?

5. RQ3 - Intervention Methods

5.1. If you intervened, how did you intervene?

5.2. How did you determine when to intervene?

5.3. What did you say or do?

5.4. Why did you choose to intervene this way?

5.5. How did your method of intervention help the trainee overcome their confusion?

5.6. How did you know the trainee had overcome their confusion?

5.7. How often do you have to intervene to clarify or resolve confusion during VR training?

5.8. If you didn't intervene, did you use any other technique to mitigate the trainee's confusion in the future?

5.9. If you didn't intervene, did you use any other technique to mitigate the trainee's confusion in the future?

Appendix B: IRB Approval Form

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
2420 Lincoln Way, Suite 202
Ames, Iowa 50014
515 294-4566

Date: 02/28/2020

To: Eliot H Winer

From: Office for Responsible Research

Title: **Study on Trainer Intervention in Virtual and Augmented Reality Training Simulations**

IRB ID: **20-081**

Submission Type: Initial Submission **Exemption Date:** 02/28/2020

The project referenced above has been declared exempt from most requirements of the human subject protections regulations as described in 45 CFR 46.104 or 21 CFR 56.104 because it meets the following federal requirements for exemption:

2018 - 2 (ii): Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) when any disclosure of the human subjects' responses outside the research would not reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, educational advancement, or reputation.

The determination of exemption means that:

- **You do not need to submit an application for continuing review. Instead, you will receive a request for a brief status update every three years. The status update is intended to verify that the study is still ongoing.**
- **You must carry out the research as described in the IRB application.** Review by IRB staff is required prior to implementing modifications that may change the exempt status of the research. In general, review is required for any *modifications to the research procedures* (e.g., method of data collection, nature or scope of information to be collected, nature or duration of behavioral interventions, use of deception, etc.), any change in *privacy or confidentiality protections*, modifications that result in the *inclusion of participants from vulnerable populations*, removing plans for informing participants about the study, any *change that may increase the risk or discomfort to participants, and/or* any change such that the revised procedures do not fall into one or more of the regulatory exemption categories. The purpose of review is to determine if the project still meets the federal criteria for exemption.
- All ***changes to key personnel*** must receive prior approval.
- **Promptly inform the IRB of any addition of or change in federal funding for this study.** Approval of the protocol referenced above applies only to funding sources that are specifically identified in the corresponding IRB application.

CHAPTER 6. INTEGRATING ADAPTIVE FEEDBACK INTO AN ONLINE VR RESEARCH TASK

Melynda Hoover, Jonathan W. Kelly, Stephen B. Gilbert, and Eliot Winer

Iowa State University, Ames, Iowa, USA

Modified from a manuscript submitted to the IEEE VR journal track for publication in *TVCG* on September 6, 2021

Abstract

The rising popularity of commodity Head Mounted Displays means participants who own VR devices can be recruited to complete unsupervised studies online, on their own time. This method of performing research has the potential to save VR researchers significant time and effort normally spent conducting lab studies. However, for unsupervised studies to be beneficial, researchers must ensure that participants understand the task and complete the study as intended. The authors propose using adaptive automated feedback to supplant the need for researcher supervision in remote VR studies. A study application was designed with five different kinds of adaptive feedback triggered by common errors made during a specified task by monitoring operator, spatial, and temporal inputs. Participants included 47 VR HMD owners with a successful completion rate of 89%. Results showed that the adaptive feedback had the intended effect of intervening when participants needed more instruction (at the beginning of the study) and teaching the participants to complete the task in a manner that was not overtly intrusive. Future work includes determining what type of feedback is most helpful during remote VR studies and examining other triggers and feedback channels such as audio or haptics.

Key Words: Virtual Reality, Remote Research, Unsupervised Research, Adaptive Automation, Feedback

Introduction

When the COVID-19 pandemic began in 2020, researchers all over the world had to engineer new ways to conduct research and collect data. In many cases, this meant turning to the internet. In some research fields, like human-computer interaction (HCI), the practice of completing entire user studies online has been commonplace for many years (Andrews et al., 2003; Goodman & Paolacci, 2017; Kittur et al., 2008). Online user studies in HCI have been made possible by the prevalence of computers, laptops, and smartphones, which made finding participants relatively easy. Virtual reality (VR) does not yet benefit from the same pervasiveness and therefore necessitates more training in order to use it effectively. However, commodity VR head-mounted displays (HMDs) are becoming increasingly popular, making large-scale, remote VR research feasible for the first time (Kelly et al., 2021; Muñoz-Saavedra et al., 2020).

Before the pandemic began, scant research existed regarding how to conduct VR research online. Out of necessity, researchers have begun to devise and share their methods for collecting VR data online from participants who have access to their own HMDs (Ratcliffe et al., 2021). Although the world is beginning to re-open, and face-to-face studies are once again possible, some researchers will continue to conduct VR studies online because in many cases, online research is faster and less time consuming than conducting a study in the lab (Goodman & Paolacci, 2017). Online VR research has other benefits as well, such as reaching more geographically diverse populations, and allowing many participants to complete a study simultaneously (Moelson & Hornbæk, 2017). However, online VR research is not without its challenges. One of the biggest challenges of online data collection is ensuring data validity. When the researcher is not in the room with the VR user, it can be difficult or impossible to troubleshoot problems, or even to recognize when they occur. This can result in errors and noise in the data set, compromising the power of the results (Moelson & Hornbæk, 2017). To solve this problem, the authors propose introducing adaptive automation to replace or supplement the role of a researcher during online VR studies.

Adaptive automation is a process that uses information about a human-computer system to automatically diagnose the system state and reallocate tasks accordingly (Rouse, 1976). This technique has been employed in training applications to customize feedback for learners and improve training outcomes (Kelley, 1969; LaViola et al., 2015). The authors propose using adaptive automation to provide responsive, real-time instructions during remote VR research to replace feedback typically provided by a co-located instructor in the lab as shown in Figure 6.1. Learning to complete a study task in VR is not very different from training applications studied in existing adaptive training literature (Zahabi & Abdul Razak, 2020). Therefore, the authors hypothesize that adaptive automation applied to task instructions will help participants complete remote studies successfully.

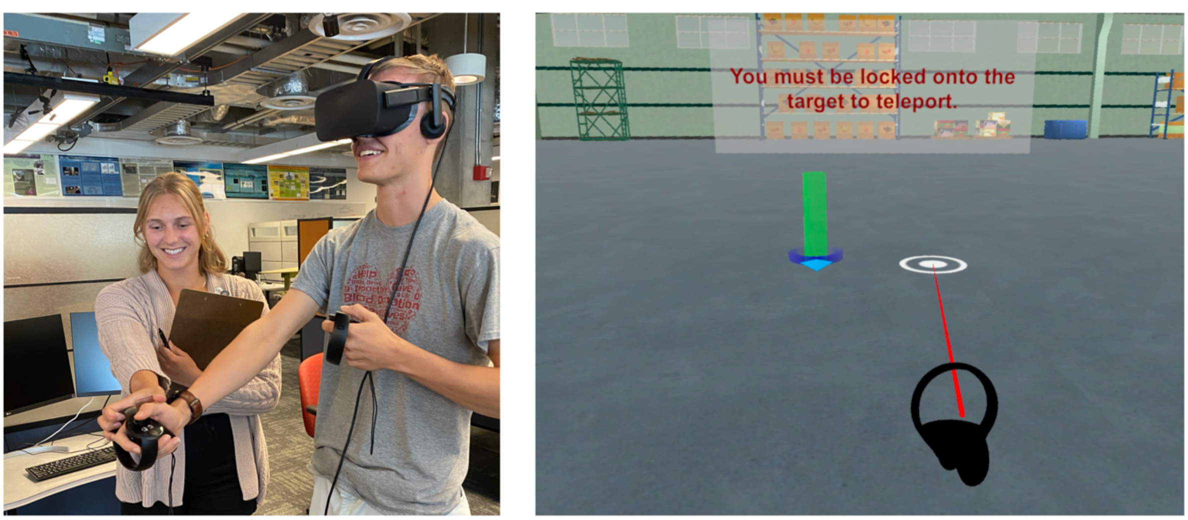


Figure 6.1. **Left:** A researcher training a participant on how to teleport during a VR lab study. **Right:** An example of adaptive feedback that is triggered when a participant attempts to teleport to the wrong location during an unsupervised online study.

To test this hypothesis, an experiment was devised that used adaptive feedback to help participants complete a navigation task in VR. Five different types of automated adaptive feedback were developed for the application based on feedback from pilot testing. Data was collected from remote participants with no supervision from a researcher. The data was then investigated to see if the adaptive feedback impacted navigation error and task completion time. Additionally, more detailed analysis showed relationships between task conditions and the types of adaptive feedback triggered. Finally, subjective data from the participants was analyzed.

Related Work

The following section reviews the existing literature on remote VR research. However, the amount of research on the subject is somewhat limited at this time, so the authors draw from literature regarding remote research in other similar fields as well. The subsequent section details literature pertaining to adaptive automation and its use in training applications.

Remote VR Research

During the COVID-19 pandemic, conducting VR research in person became inherently unsafe due to the risk of spreading disease. To remedy the situation, researchers began to conduct more studies online. In computing, having users participate in product testing on their own devices, and in their own spaces is often referred to as research "in the wild" or "crowdsourcing" (Crabtree et al., 2013; Goodman & Paolacci, 2017). This type of research can elicit richer and more contextual data than lab tests because the product is being used in an environment that is more natural to the user. However, researchers must find a qualified pool of participants for their online study. HCI researchers have demonstrated that online work sites, such as Amazon Mechanical Turk (MTurk) and Prolific can be used to find large sample populations (Buhrmester et al., 2011; Mason & Suri, 2012). However, finding those who own very specific hardware, such as a VR HMD, lends an additional challenge. Crabtree et al. also suggested that preparing for "in the wild" research can be trickier. For example, developers have to spend additional time automating processes and data collection (Crabtree et al., 2013). Additionally, recruiting VR participants from online work sites can lead to the recruitment of more male than female participants (Kelly et al., 2021).

One practical study, which was published before the pandemic began, showed that VR research could be conducted successfully online. In that paper, Steed et al. used participants' own smart phones to display VR content and conduct research on embodiment and presence in VR (Steed et al., 2016). They collected survey data, hardware data, and head tracking information, which was uploaded directly to their server from participants' devices. This was innovative because it required no synchronous contact (face-to-face or virtual) between the researcher and the participant, allowing for more efficient data collection than traditional lab-based research methods. However, there were some additional challenges with their online VR research method. Specifically, the process required more development time and time spent checking data to ensure participants performed tasks correctly without supervision (Steed et al., 2016).

Since the pandemic began, more researchers have been conducting unsupervised VR experiments online (Ratcliffe et al., 2021). However, detailed descriptions of their work have only recently entered into publication. In 2021, Mottelson et al. conducted two unsupervised, online studies using participants' own VR devices. The first studied learning in virtual environments (VEs) by having participants respond to a knowledge questionnaire before and after a VR learning session. This study was conducted both online, and in a lab for comparison purposes. They found that the online sample had a higher percentage of negative or identical preto post-test scores, and therefore a lower rate of attentive participation than the lab sample. In the second experiment, the researchers asked participants to complete a hand gesturing task, and tracked their finger movement using an Oculus Quest 1 or 2. They found that gesture tracking was not always reliable because of environmental factors such as lighting, and had to implement additional measures, such as a screening task, to avoid recording useless data (Mottelson et al., 2021).

The studies by Mottelson et al. demonstrated that VR research conducted online is prone to increased errors and attention lapses that can render data unusable (Mottelson et al., 2021). In order to mitigate the collection of erroneous data, the authors propose incorporating adaptive automation into a VR experiment. This feedback could serve a role similar to that of a researcher in a lab environment to guide a participant and reduce errors during data collection. By automating feedback to a participant, the research can remain unsupervised, resulting in benefits to researchers, such as savings of time and effort.

Automating the Researcher Role

A researcher typically plays a large role in the success of the participant during VR research. For example, in a online VR study by Saffo et al., a VR chat room application called VRChat was used to conduct research (Saffo et al., 2020). Using VRChat, researchers could meet with participants distributed across the globe. In their case study, they report on a procedure that includes researchers meeting "face-to-face" with participants in VR to explain the study objective and instruct them how to complete the task (Saffo et al., 2020). Although it adds a personal element to the study, remote, supervised research requires significant time and effort on the part of the researcher, similar to that required by in-person research. Not to mention, many studies have multiple supervising researchers, which can lead to less internal validity if protocols are not followed very closely (MacKenzie, 2012). By automating the researcher's role in the study, data can be collected much more quickly, efficiently, and consistently online.

In general, automation means replacing a manual task with a task performed by a computer system (Parasuraman et al., 2000). In the case of this research, the manual task being replaced is that of a researcher identifying incorrect actions performed by a participant and providing information to supplement the participant's understanding of the research task. However, it is not sufficient to simply create a user interface (UI) with text instructions in a VE, because it is common for trainers to adapt their instructions according to the performance level of the learner (or participant) (Kelley, 1969). Therefore, by using adaptive automation to provide instructions to a participant, the experience can be similar to that of a supervised study performed in a lab. Kelley et al. described the process of adaptive training as a closed-loop system where a system monitors inputs from a trainee, such as moving through a Virtual Environment (VE). Their input is then evaluated, and adaptive logic intervenes and changes the stimuli experienced by the trainee (Kelley, 1969). In Figure 6.2, this feedback loop has been modified to reflect the factors specific to training in a VR experiment.

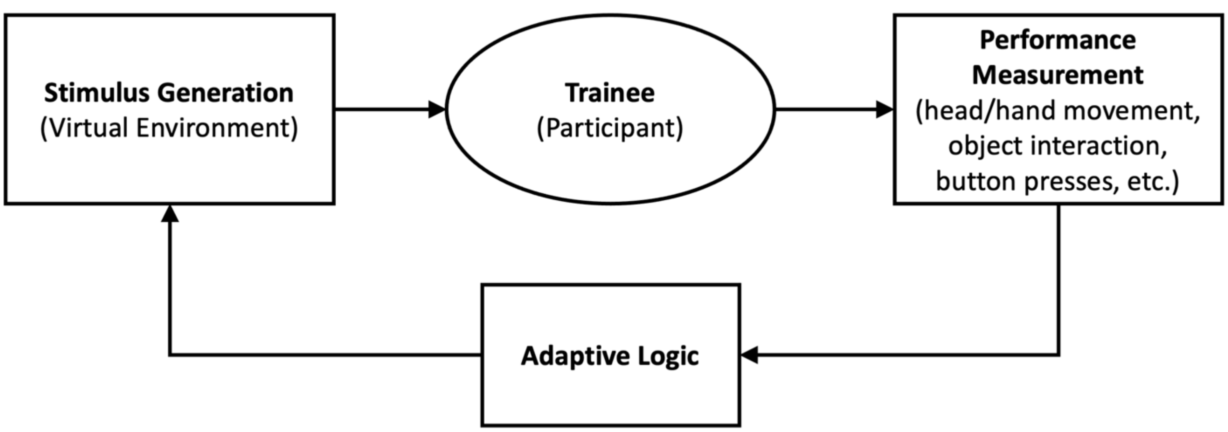


Figure 6.2. Flowchart demonstrating adaptive automated feedback during VR study training. Adapted from Kelley et al., 1969.

It is worth noting the difference between true adaptive automation and the use of a wizard (Van Welie et al., 2001). A wizard is a UI design pattern in which users answer a series of questions while clicking Next after each one. It can be used to guide users through complex decisions, but the process is not adaptive to different contexts. In order to design adaptive automation for VR applications, app designers must select triggers that activate predefined changes to a VE (Lagu et al., 2013). Defining these triggers allows a system to identify certain user behaviors automatically and implement a prescribed adaptation matching the trigger. Furthermore, adaptive automation can be used to continuously update a system and apply increasingly aggressive interventions, often described using different Levels of Automation (LOAs) (Sheridan & Verplank, 1978).

In human subjects research, it is often not possible to allocate the performance of tasks to a computer because it could compromise the resulting data. However, adaptive automation can be used to manipulate how much information a participant receives. For example, remembering which button to press and which part of the task to work on can be easily monitored and displayed by the computer so the participant doesn't have to remember them.

Designing Adaptive Automation for VR

When designing adaptations for computer systems, there are two main factors to consider: the inputs that trigger an adaptation, and the way the adaptation modifies the system (Feigh et al., 2012). The paper by Feigh et al. demonstrated these factors by creating taxonomies. They categorized inputs into five categories: Operator, System, Environment, Task/Mission, Spatio-Temporal. Then, they grouped adaptations into four types of modifications to the system: function allocation, task scheduling, interaction, and content (Feigh et al., 2012). These generalized classifications explain the options available when designing adaptive systems.

However, it remains up to the designer or developer to determine how best to apply these methods to VR applications.

Researchers have found some success in implementing adaptive automation into their VR and augmented reality (AR) applications, however, few provide empirical evidence or case studies in real-world settings (Zahabi & Abdul Razak, 2020). In the absence of extensive literature on how to design systems that combine VR and adaptive automation, these studies are useful tools for choosing how to combine these two technologies. One quite literal example of adaptive automation replacing human supervision during training is the virtual agent called "Steve" (Rickel & Johnson, 1999). This early example used a human-like avatar to instruct and assist trainees in conducting naval operations in VR. Steve used gestures and verbal instructions when guidance was requested by a trainee. Although using an avatar is the most analogous representation of a researcher in the virtual space, providing feedback via an avatar is not always feasible or necessary.

Other examples of adaptive VR/AR systems have used automation to adapt the difficulty of a VR task based on real-time user performance (Barzilay & Wolf, 2013; Fricoteaux et al., 2014). In one study, authors used signals such as heart rate, head position, interactions, and performance metrics to change the interface elements and task difficulty in a fluvial navigation task (Fricoteaux et al., 2014). Another study used electromyographic (EMG) sensors and motion tracking data to adapt the difficulty of physical therapy tasks to the needs of users (Barzilay & Wolf, 2013). In both of these cases, the adaptations were effective, but changing the difficulty of the task at hand is not typically feasible in a research experiment, since the task needs to be very controlled for consistency across participants. Additionally, if the research is conducted online, using the participant's own hardware, it is not possible to include measures like heart rate and

EMG in the design of triggers. However, motion tracking data, such as head and hand positions, are easily monitored using the innate inertial measurement units (IMUs) in most VR HMDs and controllers, making them viable data to use in the design of triggers. Interactions with the environment can also be easily monitored through event systems in game design software like Unity, making them feasible adaptation triggers as well.

Other research in the AR training domain has pointed to the helpfulness of text and graphics in the VE to provide guidance to the trainee. In a study by Westerfield et al., the authors developed an adaptive AR application for assembling a computer motherboard (Westerfield et al., 2013). Their application focused on adapting the text and 3D models seen by the trainees to help them self-correct their errors. They found that the adaptive interface helped improve practical test scores by 25% over those who did not receive adaptive feedback. Radkowski et al. supported the use of text and 3D graphics to present training instructions as well by creating an AR application that varied UI elements such as text and graphics depending on the difficulty of each task in a pump assembly (Radkowski et al., 2015). Their results showed that using text and animated 3D models of parts to present instructions resulted in better performance than using text and arrows alone. Based on these results, it is clear that text and 3D models can be useful targets for adaptations in AR. It stands to reason that these may also be helpful outputs for adaptations in VR systems.

Online Experiment with Adaptive Feedback

Rotational self-motion cues are essential to accurate navigation in VR (L. Cherep et al., 2021; L. A. Cherep et al., 2020; Kelly et al., 2020; Lim et al., 2020). The importance of self-motion cues is especially evident in navigation performance when using teleporting interfaces, which necessarily lack some self-motion cues. When using a partially concordant teleporting interface, participants maintain their heading when teleporting between points and physically rotate their body to turn. When using a discordant teleporting interface, participants manipulate their desired heading using a joystick or capacitive touchpad before teleporting, effectively removing rotational self-motion cues. The terminology, partially concordant and discordant, refers to the extent that movement through a VE is consistent with movement of the body. The navigational benefit of rotational self-motion cues (i.e., the benefit of partially concordant teleporting compared to discordant teleporting) has been demonstrated repeatedly (L. Cherep et al., 2021; L. A. Cherep et al., 2020; Kelly et al., 2020; Lim et al., 2020). One experiment used a triangle completion task to evaluate the effects of three variables (teleporting interface, path size, and VE size) on triangle completion performance (Kelly et al., 2020).

This original study was used as the foundation for the current experiment because it involved a simple task with many similar trials, which meant that learning and training predominantly occurred early in the study. Additionally, the authors conducted informal interviews with researchers who conducted the original triangle completion task study in a lab. The researchers reported that participants often struggled with initially learning how to use the controllers to navigate through the environment. This made the task perfect for implementing adaptive feedback to replace the need for customized instructions from a supervising researcher.

In the current study, rather than bringing participants to a lab, the same experiment was conducted online with participants using their own VR equipment. The authors expected that a remote, unsupervised study could result in more variability in measurements (compared to a supervised, lab-based study) due to differing equipment and environments. However, conducting the study remotely was unavoidable in this case due to the COVID-19 pandemic.

Rather than analyzing the factors that affect navigation performance (those data are presented in a separate paper (Kelly et al., n.d.)), this paper focuses on how adaptive feedback impacts a user's performance and task completion time during an unsupervised, remote study. The relationships between the treatments and the quantity and type of adaptive feedback are also analyzed. Additionally, the authors address how users felt about the quantity, helpfulness, and intrusiveness of the adaptive feedback provided.

Overview

At this time, no published works have addressed the application of adaptive feedback and automation in an online, unsupervised VR study. This work will fill the gap in the current research by applying adaptive feedback to train participants to complete a VR research task. The task used in this experiment is a triangle completion task, in which participants travel two legs of an outbound path through the VE before attempting to point to the path origin (i.e., to complete the triangle). Three independent variables are manipulated in a 2x2x2 within subjects design: teleporting interface, triangle path length, and VE size.

Hypotheses

First, the authors wish to investigate whether the adaptive feedback worked as intended, and if it can be improved upon in future study designs. This will be done by testing the following hypotheses.

The first hypothesis (H1) was that the magnitude of errors produced by participants when completing the triangle completion task would decrease over time, as they improve at interacting with the VE. Similarly, the next hypothesis (H2) was that the time needed to complete the task would decrease toward the end of the study as participants spent less time interpreting feedback and acclimated to the task. These predictions were made because adaptive feedback paired with detailed instructions has been shown to be effective at training users to perform a task in VR and AR in (Fricoteaux et al., 2014; Westerfield et al., 2013). However, the feedback was designed to remind participants about instructions they had already been given and help them complete the task, not to help them perform the triangle completion task at a higher level of proficiency.

The third hypothesis (H3) was that participants with more video game and VR experience would trigger less adaptive feedback overall than those with less experience. This was because those with more VR and video gaming experience were likely have more familiarity with system controls, resulting in fewer mistakes.

The fourth hypothesis (H4) was that adaptive feedback would be triggered by participants more frequently at the beginning of the experiment than at the end. This prediction was made because participants would likely make more feedback-triggering mistakes at the beginning, when the task and interface were less familiar. If the information provided by the adaptive feedback was effective, a participant would learn to correct their mistakes as the experiment goes on, causing feedback to diminish.

The fifth hypothesis had to do with the participants' perceptions of the quality and quantity of the adaptations received. It was predicted that participants' perceptions of the intrusiveness, helpfulness, and quantity of the adaptive feedback would not differ, despite variations in the actual quantity of feedback they received (H5). This outcome would show that the adaptive feedback worked as intended because participants who made more mistakes, would receive more feedback, and participants who made fewer mistakes would receive less feedback. If H5 were true, then both of these groups would report that the amount of feedback they received was appropriate and helpful.

Finally, in order to show that the research study with adaptive feedback was effective at replacing the need for a research supervisor in remote VR studies, two things would need to happen: Participants would complete the study at a similar rate as participants in a lab environment (H6), and the participants' performance on the task would be similar to that of lab-based participants (H7). (H7 is addressed in the paper by Kelly et al., n.d.)

Methodology

A remote VR study was designed to answer the proposed research questions and to mitigate the risk of spreading the COVID-19 virus. For this study, participants were asked to use their own VR HMD at home, without the help of a supervising researcher.

Table 6.1. Summary of HMD features.

HMD	**Resolution per eye**	**Binocular FOV***	**Max refresh rate (Hz)**
Oculus Rift	1080 x 1200	~100°	90
Oculus Rift S	1280 x 1440	~100°	80
Oculus Quest	1600 x 1440	~100°	72
HTC Vive	1080 x 1200	~110°	90
HTC Vive Pro	1440 x 1600	~110°	90
Valve Index	1440 x 1600	~130°	144

* Binocular field of view are estimates based on manufacturer’s reported specifications for HTC Vive and Valve Index. And on a combination of manufacturer reported values and comparisons for Oculus products.

Hardware/Software

The VE could be delivered on the Oculus Rift, Oculus Rift S, Oculus Quest, HTC Vive, HTC Vive Pro, or Valve Index. The specific hardware depended on what was available to each participant. Because of this asynchronous, distributed method, graphics were rendered on the user’s VR-capable personal computer. Therefore, the HMD had to be connected to the computer, even if it was capable of wireless functionality (e.g., the Oculus Quest). Since system specifications could not be controlled, graphics card, processor types, and amount of random-access memory were recorded and considered during data analysis. Similarly, the resolution, field of view, and refresh rate depended on the HMD used by each participant. Table 6.1 contains a list of the qualifying HMDs and their specifications.

The HMD used by each participant also impacted the type of controllers used during the experiment, and therefore altered the interaction method. Specifically, this impacted the method used to teleport in the VE. Users of the HTC Vive and Vive Pro used the capacitive trackpad on the Vive motion controllers, while Oculus Rift, Rift S, Quest, and Valve Index owners used the mechanical joystick on their controllers. All of the qualifying controllers have an index finger-activated trigger which was used to point and click when selecting buttons or submitting their response during the triangle completion task.

Application Design

Although performance on the triangle completion task was used as a measure of the success in the adaptive, remote VR experiment, it was not the main focus of this research (those data are reported in detail elsewhere (Kelly et al., n.d.)). The scope of the current research was to evaluate the effectiveness and qualities of adaptive feedback as a method for replacing the need for a human supervisor/trainer in an online VR research task. Therefore, this section will describe the triggers and adaptations employed during the remote VR task.

The VR software application for this research was developed using Unity 3D. During development, it was important that the new VR application provided the same, or nearly the same experience as the original lab study. To achieve this goal, basic instructions were added at the beginning of the application to mimic the ways researchers explained the task during the in-person user study. In between trial blocks, a menu was added to allow participants to take a break and report if they had experienced technical difficulties during the previous block of trials. Since there was no researcher supervising the session, self-report measures were important to determine if the application worked as expected during the experiment. Additionally, data linking features including an interface for entering the participant's ID number, and the ability to upload data automatically to a server via WiFi were implemented. This allowed researchers to receive data from the VR application and match it up with consent forms and other survey data gathered outside the VR app using Qualtrics.

Table 6.2. Design of Triggers and Adaptations.

Trigger	**Task**	**Adaptive Feedback**
Incorrect button press	Teleporting	Use your thumb pad/stick to teleport.
Correct button press & Incorrect teleport position	Teleporting	You must be locked onto the target to teleport.
15 seconds without activating teleportation interface	Teleporting	1. Teleport to the colored marker to continue.
		2. Make sure you are pointing the controller at the ground to teleport.
		3. Press down with your thumb, aim the controller at the colored marker, and then release to teleport.
15 seconds without activating response interface	Answering	1. Point to where the green marker was located to continue.
		2. Make sure you are pointing the controller at the ground to submit your response.
		3. Press down on the trigger with your index finger, aim the controller at the ground, and release the trigger to submit your response.
Incorrect button press	Answering	Use the trigger instead of your thumb to submit your response.

Triggers

Triggers are signal events that are used by the system to choose and apply an adaptation. For this application, the triggers were designed to recognize common task errors during pilot study observations. Specifically, five different triggers were designed to recognize when a participant had missed the teleporting target, pressed the wrong teleporting button, pressed the wrong response button, failed to teleport within 15 seconds, or failed to give a response in the triangle complete task within 15 second. Triggers were created using combinations of several different input types from Feigh et al., 2012, including operator, temporal, and spatial inputs. For the temporal triggers, 15 seconds was determined during pilot tests to be an appropriate amount of time to signal inaction based on pilot testing.

Adaptations

For this research, adaptations appeared as text instructions in the VE. In the experiment, the adaptive feedback was presented using red text presented on a semi-opaque background. The "billboarding" technique was used to keep the feedback in the center of the user's field of view. Feedback was presented for five seconds before disappearing. This gave participants enough time to read the text without interfering with their ability to complete the task during pilot testing. Additionally, only one type of feedback could be displayed at a time so that the user would not be overwhelmed by the information. The adaptive feedback text was customized to the needs of the user based on the triggering event. A complete list of the types of feedback given and their corresponding triggers is shown in Table 6.2. Note that two of the triggers had three different levels of adaptive feedback denoted by numbered lists. The feedback was authored to increase in specificity with each level, starting with a reminder of the general goal, then advice for reaching the goal, and finally step-by-step instructions to accomplish the goal. This method of feedback was modeled off of that used by Corbett et al. in their research on intelligent tutoring systems (Corbett et al., 2001). In these cases, the feedback level increased by one each time the trigger was activated until it reached level three. Level three feedback was repeated until the associated action was completed successfully.

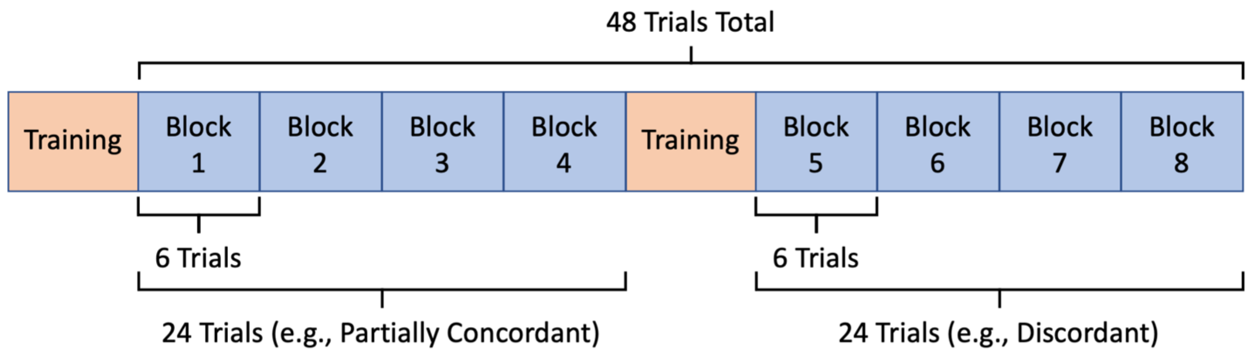


Figure 6.3. Visualization of trial blocks. The study had eight trial blocks, each containing 6 trials. Blocks 1-4 used one teleporting method, while blocks 5-8 used the other teleporting method. The order of the teleporting methods was counterbalanced during the study. Before a participant was exposed to a new teleporting interface, they experienced a training module to teach them the new technique.

Procedure

All procedures described in this section were approved by the Institutional Review Board at Iowa State University. The approval form for this research can be found in Appendix: IRB Approval Form. After a participant completed the screener and informed consent document via Qualtrics, they received a unique six-digit identification number and were directed to a website

with instructions to download and install the VR application. After launching the application and donning their HMD, they entered their identification number and gender. Then, their identification number was used to determine the order in which they would experience the counterbalanced condition levels.

Before the experimental trials, participants were presented with a training module to teach them how to use the first teleportation interface. Once they felt comfortable, the participant could advance to the experiment trials, where they would complete four trial blocks, each comprised of 6 triangle completion task trials. Next, they were trained on how to use the other teleporting interface, and then completed a second set of four trial blocks. Figure 6.3 shows how all 48 trials were broken into blocks with different treatments. The eight treatments were the results of a 2x2x2 full-factorial design with the following conditions and levels: teleporting interface (partially concordant/discordant), triangle path length (10 feet/40 feet), and VE size (small/large). During the trials, adaptive feedback could be triggered based on the participant's inputs. Next, they would repeat the training and experimental trials with the second teleporting condition, resulting in a total of 48 trials. Conversely, the previously conducted lab condition of the study required the participant to complete 96 trials of the triangle completion task (12 of each combination of conditions.) The number of trials in each condition block was reduced by half in order to keep the remote study under 60 minutes in duration and to decrease the rate of study discontinuation, since online users would not have the additional motivation and social pressure resulting from a researcher's physical presence. During the remote study, participants were offered breaks and asked to report on simulator sickness and technical difficulties every 6 trials.

After the VR task, participants were instructed to remove their headset and were directed to a Qualtrics questionnaire via a short link to give feedback on their experience and answer demographics questions. The survey included questions about VR and video game experience, a simulator sickness questionnaire, and subjective questions about the adaptive feedback they received during the experiment. Lastly, participants were compensated through Amazon Mechanical Turk, Prolific or with an e-gift card via email.

Participants

Participants for this research were recruited using Amazon Mechanical Turk, Prolific, Facebook, and Reddit. Participants qualified for the study if they had access to an Oculus Rift, Oculus Rift S, Oculus Quest, HTC Vive, HTC Vive Pro, or Valve Index, which they could connect to a VR enabled PC using Steam. They also had to be 18 years of age or older and reside in the United States. In total, 47 people participated in the online study, with 42 completing the entire study successfully. Of the 47 participants, data from four were excluded because they did not complete the entire study, or some of their data was missing. Another four were excluded because they used an HMD that was not on the approved list to complete the study. One participant was excluded because they reported a technical difficulty that impacted their ability to complete the study as it was designed. Therefore, a total of nine samples were excluded, and the final online study sample included data from 38 unique participants.

Within the group that completed the study online, 33 identified as male, four female, and one who preferred not to specify. Their ages ranged from 18 to 55 with a mean of 28. They also used a variety of approved HMDs with the HTC Vive being the most popular. The distribution of HMDs used can be found in Table 6.3.

Table 6.3. HMDs used in the online study.

HMD	Number of Participants
HTC Vive	8
Oculus Rift	7
Oculus Rift S	7
Oculus Quest 2	6
Valve Index	5
Oculus Quest	4
HTC Vive Pro	1

Results

The results section will first address how the presence of adaptive feedback in the remote study affected the participant's performance on the triangle completion task in terms of error magnitude. In the second section, analysis will be conducted to show how the adaptive feedback affected task completion time. Finally, the authors explore the types and quantity of feedback received by the participants, and subjective measures regarding their opinions of the adaptive feedback.

Task Error

During the triangle completion task, participants were asked to teleport along two legs of a triangle, and then point back at the ground to indicated where the path started (completing the triangle). The amount of error in their response was then calculated to be the shortest distance between the given answer and the correct path origin location along the ground plane measured in meters. Errors were averaged across all trials in the same treatment group for each participant to mitigate the effect of outliers. A natural log transformation was used to remove skewness from the data (Osborne & Costello, 2008). The resulting data was more normally distributed and had more similar variances than the untransformed data. Although there were small departures from normality, the authors determined that the data was within reasonable limits to continue with the statistical test. A linear mixed-effects model (LMM) was found to best describe the relationship between the response and predictor variables because it could account for multiple main effects and incorporate a random effect to account for having multiple data points from each participant (repeated measures).

The main effects incorporated into the model were the number of adaptations triggered and block number (1 to 8), along with the following covariates: teleporting interface (discordant or partially concordant), triangle path length (10 ft or 40 ft), and VE size (small or big). The LMM was fit using the Restricted Maximum Likelihood (REML) and t-tests using Satterthwaite's method. The results of the LMM showed that the number of adaptations triggered, β=0.099, $t(273)$=3.266, p=.001, teleporting interface, β=-0.298, $t(266)$=-5.669, p<.001, path length, β=1.024, $t(252)$=23.865, p<.001, and VE size, β=-0.170, $t(252)$=-3.956, p<.001, had statistically significant impacts on absolute distance error, while block number, *ß*=0.031, *t*(265)=1.699, *p*=.091, had a nearly significant impact and a positive beta value. Lastly, there was a significant interaction between the number of adaptations received and block number when predicting task error magnitude, *ß*=-0.014, t(269)=-2.472, *p*=.014. Details about all the predictors used in this analysis can be found in Table 6.4.

Table 6.4. Results of the LMM to test the effects of adaptive feedback quantity and trial block number on performance and task completion time with covariates of teleporting interface, triangle path length, and VE size.

Independent Var.	Predictor (Reference Point)	Estimate (ß)	Std. Error	t-value	p-value	
Task Error	Intercept	0.404	0.132	3.048	.003	**
	Adaptation Triggers	0.099	0.030	3.266	.001	**
	Block Number	0.031	0.018	1.699	.091	·
	Teleporting Interface (part. Concord.)	-0.298	0.053	-5.669	< .001	***
	Path Length (40ft)	1.024	0.043	23.865	< .001	***
	VE Size (small)	-0.170	0.043	-3.956	< .001	***
	Adaptation Triggers * Block Number	-0.014	0.006	-2.472	.014	*
Completion Time	Intercept	2.537	0.087	29.063	< .001	***
	Adaptation Triggers	0.140	0.019	7.295	< .001	***
	Block Number	-0.005	0.012	-0.443	.658	
	Teleporting Interface (part. Concord.)	-0.274	0.033	-8.199	< .001	***
	Path Length (40ft)	0.206	0.027	7.599	< .001	***
	VE Size (small)	-0.021	0.027	-0.776	.438	
	Adaptation Triggers * Block Number	-0.019	0.004	-5.150	< .001	***

Significance codes: < .001 = ***, .01 = **, .05 = *, .1 = ·

Based on the negative beta coefficient, there is an inverse relationship between the number of adaptations triggered by the participant, and the magnitude of their errors in each block of trials. This means that participants had smaller errors during trial blocks where they triggered less adaptive feedback, and vice versa. The block number alone did not have a significant effect on magnitude of errors, but when looked at in conjunction with the amount of adaptive feedback presented, there was a relationship. A slopes analysis using Johnson-Neyman intervals was performed to investigate this relationship. The results showed that the number of adaptations triggered by participants was only a significant predictor (*p*<.05) of error magnitude when the block number was outside of the interval [4.62, 18.50]. Since the maximum block number is eight, the interval shows that the predictor is only significant when the block number is lower than 4.62, which is during the first half of the study. The covariates listed in Table 6.4 (teleporting interface, path length, and VE size) were necessary to build a robust model but are not the subject of this paper. These variables and their relationship to various types of errors and response time are discussed elsewhere (Kelly et al., n.d.).

Task Completion Time

The duration of every trial was recorded from the time the first marker appeared for the triangle completion task to the time the participant submitted their response. This was used as the response variable in an LMM to analyze how task completion time was affected by the number of times participants triggered adaptive feedback, and block number with teleporting interface, triangle path length, and VE size as covariates. Again, speed of trial completions were averaged across treatments, and a log transformation was used to eliminate skewness from the data. The result was a more normally distributed model with the similarities in variance necessary for the statistical test.

The results from the LMM using the Restricted Maximum Likelihood (REML) and t-tests using Satterthwaite's method can be found in Table 6.4. The results showed that the number of adaptations triggered, β=0.140, $t(269)$=7.295, p<.001, teleporting interface, β=-0.274, $t(264)$=-8.199, p<.001, and path length, β=0.206, $t(253)$=7.599, p<.001, had statistically significant impacts on task completion time. Lastly, there was a significant interaction between the number of adaptations received and block number when predicting task error magnitude, β=-0.019, $t(266)$=-5.150, p<.001. The block number and the VE size did not have a statistically significant impact on the task completion time.

Based on the positive beta coefficient, there is a direct relationship between the number of adaptations triggered by the participants and the magnitude of their errors in a given block of trials. This means that as the amount of adaptive feedback shown to the participant increased, the amount of time it took them to complete the task increased as well. Once again, the block number by itself did not have a significant impact on the response variable, but there was a significant interaction between the number of adaptations triggered, and the block number. A slopes analysis using Johnson-Neyman intervals was performed to investigate this relationship. The results showed that the number of adaptations triggered by participants was only a significant predictor (p<.05) of trial completion when the block number was outside of the interval [6.18, 9.94]. Again, the interval maximum is greater than the maximum block number. So, the results of the Johnson-Neyman interval imply that the predictor was only significant during the first 6 trial blocks.

Previous VR Experience

Subsequent LMMs were used to determine if VR and video game experience could explain performance and task completion time trends. Three self-reported metrics were used to describe VR and video game experience: average number of hours spent playing video games per week (M=32.68, SD=26.63), frequency of VR use (measured on a 5-point Likert scale and treated as an ordered factor), and average length of VR play session in minutes (M=79.08, SD=60.63). None of the VR or video game metrics were significant predictors of performance or task completion time. Similarly, an LMM analysis was conducted to see if the participants' previous VR and video game experience had an impact on the amount of adaptive feedback they triggered during the study. The LMM again showed that none of the VR or video game experience metrics were statistically significant predictors of adaptive feedback quantity.

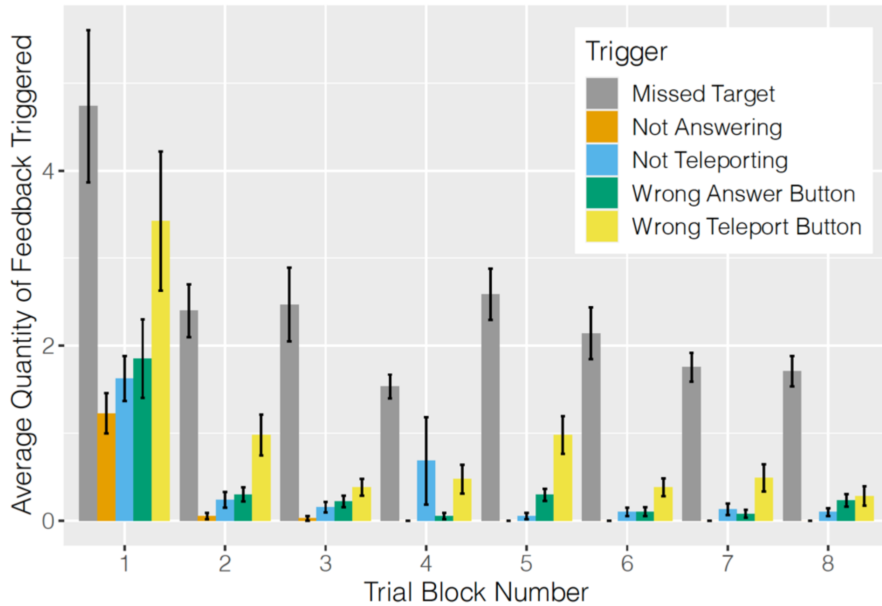


Figure 6.5. Average number of times feedback was triggered as a function of type of feedback and block number. Error bars represent ±1 standard error of the mean.

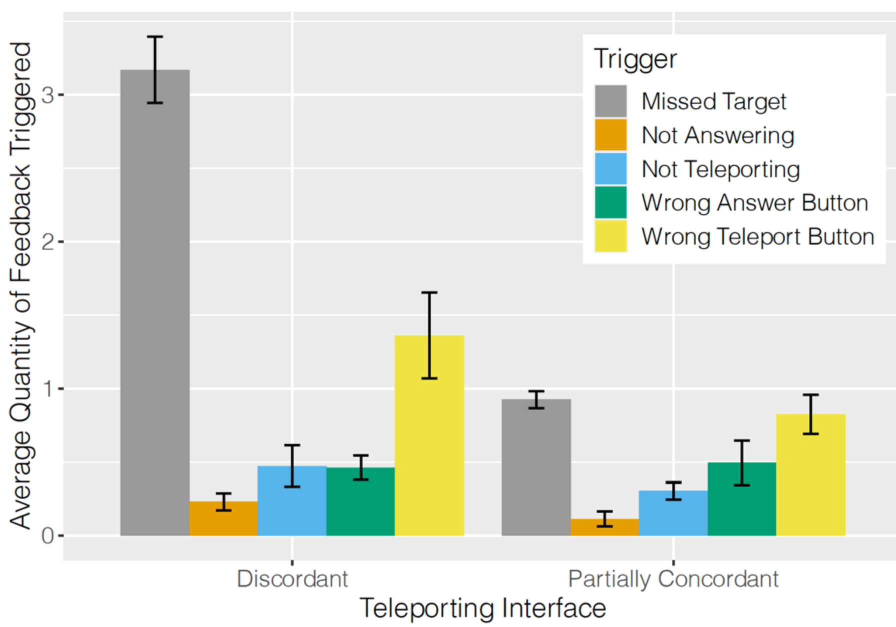


Figure 6.4. Average number of times feedback was triggered as a function of type of feedback and teleporting interface. Error bars represent ±1 standard error of the mean.

Table 6.5. Results of the LMM to test the effects of block number, teleporting interface, triangle path length, and VE size on the number and types of feedback received by the participant.

Independent Var.	Predictor (Reference Point)	Estimate (ß)	Std. Error	t-value	p-value	
Total Feedback	Intercept	3.756	0.282	13.323	< .001	***
	Block Number	-0.123	0.032	-3.852	< .001	***
	Teleporting Interface (part. concord.)	-1.764	0.146	-12.046	< .001	***
	Path Length (40ft)	0.257	0.146	1.755	.080	·
	VE Size (small)	0.199	0.147	0.811	.418	
Missed Target	Intercept	3.895	0.445	8.743	< .001	***
	Block Number	-0.283	0.068	-4.132	< .001	***
	Teleporting Interface (part. concord.)	-2.089	0.340	-6.136	< .001	***
	Path Length (40ft)	0.845	0.307	2.754	.006	**
	VE Size (small)	-0.306	0.306	-0.999	.318	
Not Answering	Intercept	0.656	0.100	6.553	< .001	***
	Block Number	-0.108	0.016	-6.921	< .001	***
	Teleporting Interface (part. concord.)	-0.095	0.071	-1.330	.184	
	Path Length (40ft)	0.041	0.071	0.571	.568	
	VE Size (small)	0.044	0.071	0.617	.538	
Not Teleporting	Intercept	1.016	0.207	4.899	< .001	***
	Block Number	-0.145	0.031	-4.616	< .001	***
	Teleporting Interface (part. concord.)	-0.130	0.145	-0.898	.370	
	Path Length (40ft)	0.172	0.144	1.193	.234	
	VE Size (small)	-0.013	0.144	-0.093	.926	
Wrong Answer Button	Intercept	0.908	0.236	3.852	< .001	***
	Block Number	-0.175	0.035	-4.991	< .001	***
	Teleporting Interface (part. concord.)	0.101	0.166	0.611	.542	
	Path Length (40ft)	0.311	0.166	1.880	.061	·
	VE Size (small)	0.181	0.166	1.088	.278	
Wrong Teleport Button	Intercept	2.438	0.436	5.593	< .001	***
	Block Number	-0.332	0.067	-4.924	< .001	***
	Teleporting Interface (part. concord.)	-0.452	0.316	-1.429	.154	
	Path Length (40ft)	0.571	0.316	1.808	.072	·
	VE Size (small)	-0.082	0.317	-0.258	.796	

Significance codes: < .001 = ***, .01 = **, .05 = *, .1 = ·

Types of Adaptive Feedback

Like in previous analyses, LMM was selected to account for multiple participant samples, and to measure the effects of multiple predictors simultaneously. Block number, teleporting interface, path length, and VR size were used as predictors in the model. For the response variable, the number of times each type of feedback was triggered was summed across all trials in the same treatment group. This resulted in eight measurements for each participant. The types of adaptive feedback being measured were labeled based on the user error that triggered it.

Therefore, the response variables will be referred to as: missed target, not answering, not teleporting, wrong answer button, and wrong teleport button. Additionally, the sum of all feedback received during the treatment was considered as a response variable and is referred to as "total feedback" in Table 6.5. Each of the models described in this section were tested for significant two- and three-way interactions. None of the interactions were significant, therefore they were removed from the final models.

The results of the LMM showed that block number, β=-0.123, $t(262)$=-3.852, p<.001, and teleporting interface, β=-1.764, $t(262)$=-12.046, p<.001, were statistically significant predictors of the total amount of feedback triggered by participants. Path length was almost a statistically significant predictor, β=0.257, $t(262)$=1.755, p<.080, and VE size was not a significant predictor.

By looking at the impact of block number on the response variable, one can understand the behavior of the response variable over time. The negative beta value shows that the amount of adaptive feedback triggered by the participants decreased as the block number increased (Figure 6.5), showing that fewer adaptations were triggered during later trials. The analysis also showed that participants received less feedback overall when using the partially concordant teleporting interface, rather than the discordant interface (Figure 6.4). This is represented by the negative beta value for teleporting interface with the partially concordant interface as the reference level (i.e., the level that was coded as zero for analysis purposes) in Table 6.5.

Missed Targets

A "Missed Target" error occurred when participants attempted to teleport to the next corner of the triangle (as marked by a colored post) but missed the mark. When comparing the impact of the predictors on the quantity of Missed Target adaptations, the LMM showed that several of the main effects were significant. Specifically, block number, β=-0.283, $t(700)$=-

4.132, p<.001, teleporting interface, β=-2.089, $t(701)$=-6.136, p<.001, and path length, β=0.845, $t(688)$=2.754, p=.006, were significant predictors of this type of feedback.

In practical terms, this means the number of missed targets decreased as block number increased, since the beta value was negative. Participants also had fewer target misses when they were using the partially concordant teleporting interface. Lastly, participants had more missed targets when they were navigating larger triangles, meaning targets that were farther away were harder to hit.

Not Answering and Teleporting

Participants could trigger two types of adaptive feedback due to inaction. Participants who failed to teleport to the next point on the triangle received adaptive feedback messages after 15 seconds. Adaptive feedback was also triggered when participants did not give their response at the end of their triangle completion task within 15 seconds. The results of the LMMs showed that these temporal triggers were activated less frequently as the participant progressed through the study. This is evident from the negative beta values associated with block number for both the "Not Answering" feedback, β=-0.108, $t(301)$=-6.921, p<.001, and "Not Teleporting" feedback, β=-0.145, $t(279)$=-4.616, p<.001. None of the other main effects had a significant impact on this type of adaptive feedback.

Wrong Button Presses

Incorrect button presses could happen in two different situations, each triggering their own feedback. First, the participant could press the trigger button instead of the thumb pad when trying to teleport ("Wrong Teleport Button" in Table 6.5). Alternatively, the participants could depress the thumb stick instead of the trigger when attempting to indicate their response at the end of the triangle completion task ("Wrong Answer Button" in Table 6.5). Other button presses, such as activating the grip button, did not trigger these adaptations.

The LMMs showed similarities in results for both types of incorrect button presses. First, when "Wrong Answer Button" triggers was the response variable, block number, β=-0.175, $t(313)$=-4.991, p<.001, was a statistically significant predictor, while path length, β=0.311, $t(313)$=1.880, p=.061, was nearly significant. The results for "Wrong Teleport Button" showed the same predictors with block number, β=-0.332, $t(336)$=-4.924, p<.001, again being a significant predictor, and path length, β=0.571, $t(331)$=1.808, p=.072, being nearly significant. In both cases, there was a decrease in the respective feedback type as participants progressed through the study exemplified by the negative beta values associated with block number.

User Opinions

After completing all trials of the online study with adaptive feedback, participants were asked to respond to three Likert scale questions about their views on the helpfulness, intrusiveness, and frequency of the feedback. The results of these surveys are visualized in Figure 6.6, where the y-axis is the number of participant responses, and the x-axis shows the ordered levels of the Likert scales.

The graphs of responses to the Likert scale questions show that participants tended to find the quantity of adaptive feedback they received was appropriate, since the majority of participants (66%) reported that it was "the right amount" n=24) on the frequency scale. Similarly, participants tended to rate the feedback low on the intrusiveness scale, represented by 76% of the responses labeled "Not at all annoying" n=12) or "Somewhat annoying" (n=17).

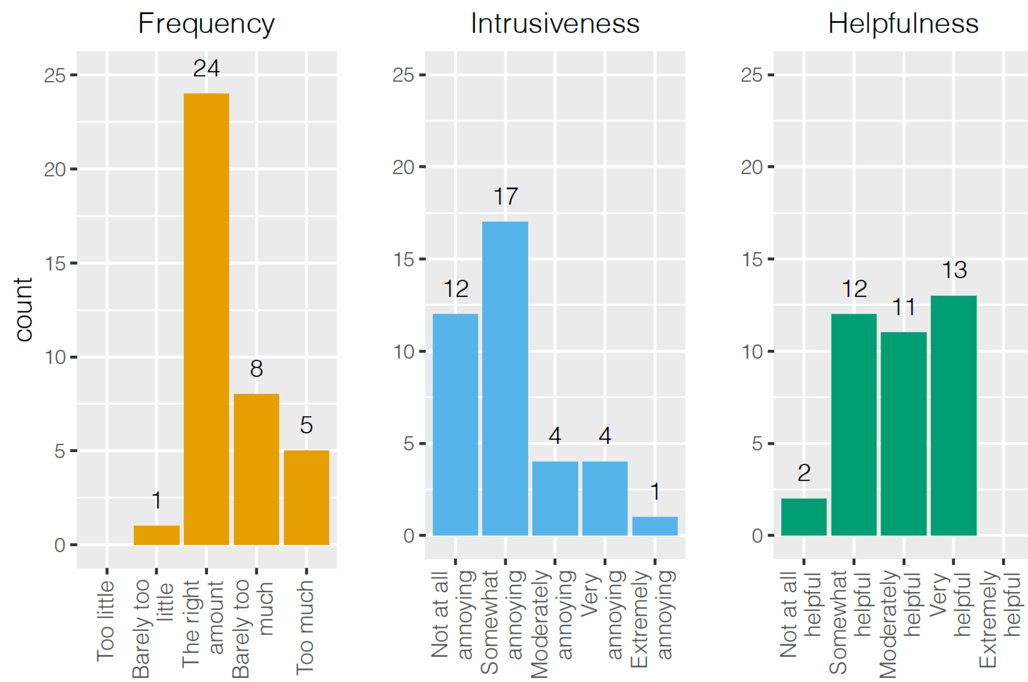


Figure 6.6. Visualizations of participants' opinions of the adaptive feedback.

Conversely, participants' opinions on the helpfulness of the adaptive feedback were more varied. Responses tended toward the middle of the range, represented by "Somewhat helpful" (n=12), "Moderately helpful" (n=11), and "Very helpful" (n=13). Only two participants described the adaptive feedback as "Not helpful at all," and none described it as "Extremely helpful."

Three separate Poisson regressions were conducted to predict the participants' ratings of feedback quantity, helpfulness, and intrusiveness based on the number of times they triggered each of the five types of feedback. The Poisson regression was chosen for its ability to handle response variables that represent count data, rather than continuous or factor data. The purpose of this analysis was to see if any individual type of feedback influenced these ratings more than other. The results of the Poisson regression showed that no one type of adaptation had a statistically significant impact on the way participants scored the helpfulness, intrusiveness, or quantity of the feedback.

Discussion

The purpose of this research was to validate the use of adaptive feedback in lieu of supervision from a researcher during online VR research and determine how the adaptive feedback could be improved. Participants completed a research study involving a locomotion task online using adaptive feedback to help them complete the study independently.

The first objective of this work was to determine if the adaptive feedback worked as designed, and to learn how it could be improved for future research. The first hypothesis (H1) was that the magnitude of errors produced by participants on the task would decrease over time. The results did not directly support this hypothesis because block number was not shown to be a significant driver of performance. The authors speculate that this was due to the nature of the adaptive feedback. The goal of the adaptive feedback was to help the participants *complete* the task, by giving instructions on which buttons to push, and reminding them what step they were currently working on. The purpose of the adaptive feedback was not to help them perform the triangle completion task with higher accuracy, which is what was tested by this hypothesis. With this in mind, the authors determined that this was not the best measure of the adaptive feedback's success.

The second hypothesis (H2) was that the time needed to complete the task would decrease toward the end of the study. Again, the block number did not have a significant impact on the speed with which participants completed the task. However, the results showed that participants took longer to complete task trials when they received adaptive feedback during the first half of the study. This is likely because participants spent some of that time reading the adaptive feedback. The authors also speculate that this relationship was not upheld in the second half of the study because participants triggered much less adaptive feedback near the end of the study. Additionally, if they triggered the same feedback more than once, they may not have spent as much time reading it during subsequent trials.

The third hypothesis (H3) addressed the potential impact of previous video game and VR experience on the participants' performance, task completion times, and the amount of adaptive feedback they received. The analysis showed that previous VR and video game experience had no relationship with performance, task completion times, nor the amount of feedback received by participants. This was counter to the hypothesis that participants with more VR and video game experience would trigger feedback less frequently. One theory that could explain this result is that participants who own their own VR hardware are more familiar with the controls, thereby reducing the learning curve typically associated with using new hardware. This theory is applicable because the VR participants recruited for this study had a high amount of video game experience, consistent with previous research on VR HMD owners (Kelly et al., 2021).

The fourth hypothesis (H4) was that adaptive feedback would be triggered by participants more frequently at the beginning of the experiment than at the end. This hypothesis was supported by the data, which showed an inverse relationship between block number and the number of times participants triggered feedback during the study. This result showed that online participants were able to learn from the adaptive feedback as they would from a live researcher and changed their behaviors, which resulted in fewer errors that trigger adaptive feedback later in the study. Additionally, participant tended to make more errors that triggered feedback when using the discordant teleporting condition. This result is supported by previous work that showed the discordant teleporting interface was inherently more difficult to use because of the added rotation control needed to teleport (Kelly et al., 2020). The increase in feedback when using the discordant interface was partly due to missing the target, since participants needed to align both the position and orientation of the teleport marker with that of the target in order to proceed. This was supported by the analysis because the number of missed target errors was also greater during the discordant teleporting trial blocks. The rest of the adaptive feedback triggers were not influenced by the teleporting condition but did follow the trend of decreasing frequency as the study progressed.

The fifth hypotheses (H5) suggested that the participants' ratings of the adaptive feedback in terms of helpfulness, intrusiveness, and quantity would not be different depending on the quantity and type of adaptations received. This hypothesis was partially supported by the data, since the participants' opinions about the feedback quantity and intrusiveness did not change based on the actual quantity of feedback they received. This result shows that the adaptive feedback worked as intended because participants who made more mistakes (and received more feedback), and participants who made fewer mistakes (resulting in less feedback) did not necessarily rate their experience with the feedback differently. Additionally, the majority of participants felt that the amount of feedback they received was "just the right amount" and that it was "not at all annoying" or only "somewhat annoying." This suggests that the design of the triggers was effective for this application. The perceived helpfulness of the feedback was also not affected by the amount of feedback the participant received. However, lower ratings of helpfulness overall indicated that perhaps some of the wording of the adaptive feedback should be changed, or adaptive feedback could be authored for other types of errors that were not initially accounted for.

To show that the remote VR study using adaptive feedback was successful in helping participants complete the study (H6), the authors looked at attrition rate of the online and lab studies. In this study 89% of participants completed the entire study successfully. The lab study conducted in 2019 (Kelly et al., 2020) had a similar success rate. In that study, 37 participants were recruited, and 4 of them either did not complete the study, or had technical difficulties during the study, resulting in an 89% success rate as well. This shows that the adaptive feedback was capable of ushering participants through the study at a similar rate as a researcher in the lab. Furthermore, the online study elicited similar results on the navigation task in terms of errors and response time, satisfying H7. These are the subject of another paper which focuses on the participants' use of navigational self-motion cues. For more information see the paper by Kelly et al., n.d. Overall, the results showed that conducting unsupervised remote research using adaptive feedback elicited similar performance and completion rates as the lab equivalent, making the use of unsupervised, remote studies with adaptive feedback a viable method for data collection in VR research.

One limitation of this study is that the authors did not compare the remote study with adaptive feedback to a control study in which remote participants did not experience adaptive feedback. This was not carried out because of limited access to online participants. However, this research should be conducted in the future and compared to the data from this study to ensure that the adaptive feedback in fact improves completion rates over traditional non-adaptive instructions.

Conclusion

Ultimately, the adaptive feedback aided participants in completing the unsupervised, remote VR study successfully. The adaptive feedback also worked as intended, triggering when the participants most needed it, which tended to be at the beginning of the experiment when they were learning how to interact with the VE. This suggests that adaptive feedback is an adequate replacement for live supervision from a researcher during VR studies, which is consistent with results from AR/VR training studies using adaptive feedback during lab-based studies (Fricoteaux et al., 2014; Westerfield et al., 2013). Additionally, the performance and task completion time of the online participants was not influenced by their previous gaming and VR experience, which could be tied to online recruitment, since online VR owners tend to have a lot of previous video game experience (Kelly et al., 2021). Furthermore, participants tended to trigger more adaptive feedback earlier in the study than at the end, showing that they learned from the adaptive feedback and changed their behavior. Future work in this area is to compare the remote study with adaptive feedback to the same remote study without adaptive feedback to further understand the difference made by the adaptive elements.

When evaluating the design of the adaptive feedback, it was found that the adaptive feedback triggered successfully when participants made mistakes during the VR task. Surveys conducted at the end of the study showed that most participants felt the amount of adaptive feedback they received was appropriate, and not too intrusive. This showed that the inputs used in the design of the triggers (operator, temporal, and spatial inputs) were effective. Some participants found that the adaptive feedback they received was not helpful, or only somewhat helpful. Based on this information, the adaptations applied to the system should be modified and further testing conducted to determine how to best present the adaptive feedback to the participants. Some options for future adaptations include changing the wording of text-based feedback to better fit the needs of the user, and trying new channels for feedback, such as haptics or audio. Further research is needed to test more varied channels for adaptive feedback and to understand what types of feedback and triggers are ideal for VR training in research studies.

References

Andrews, D., Nonnecke, B., & Preece, J. (2003). Conducting Research on the Internet: Online Survey Design, Development and Implementation Guidelines. International Journal of Human-Computer Interaction, 16(2), 185–210. https://auspace.athabascau.ca/handle/2149/1336

Barzilay, O., & Wolf, A. (2013). Adaptive rehabilitation games. Journal of Electromyography and Kinesiology, 23(1), 182–189. https://doi.org/10.1016/j.jelekin.2012.09.004

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspectives on Psychological Science, 6(1), 3–5.

Cherep, L. A., Lim, A. F., Kelly, J. W., Acharya, D., Velasco, A., Bustamante, E., Ostrander, A. G., & Gilbert, S. B. (2020). Spatial cognitive implications of teleporting through virtual environments. Journal of Experimental Psychology: Applied, 26(3), 480–492. https://doi.org/10.1037/xap0000263

Cherep, L., Kelly, J., Miller, A. J., Lim, A., & Gilbert, S. B. (2021). Individual Differences in Teleporting through Virtual Environments. Journal of Experimental Psychology: Applied. https://doi.org/10.31234/OSF.IO/B6CYD

Corbett, A. T., Koedinger, K. R., & Hadley, W. H. (2001). Cognitive Tutors: From the research classroom to all classrooms. Technology Enhanced Learning: Opportunities for Change, 235–263.

Crabtree, A., Chamberlain, A., Davies, M., Glover, K., Reeves, S., Rodden, T., Tolmie, P., & Jones, M. (2013). Doing innovation in the wild. ACM International Conference Proceeding Series. https://doi.org/10.1145/2499149.2499150

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. Human Factors, 54(6), 1008–1024. https://doi.org/10.1177/0018720812443983

Fricoteaux, L., Thouvenin, I., & Mestre, D. (2014). GULLIVER: A decision-making system based on user observation for an adaptive training in informed virtual environments. Engineering Applications of Artificial Intelligence, 33, 47–57. https://doi.org/10.1016/j.engappai.2014.03.005

Goodman, J. K., & Paolacci, G. (2017). Crowdsourcing consumer research. Journal of Consumer Research, 44(1), 196–210. https://doi.org/10.1093/jcr/ucx047

Kelley, C. R. (1969). What is Adaptive Training? Human Factors: The Journal of the Human Factors and Ergonomics Society, 11(6), 547–556. https://doi.org/10.1177/001872086901100602

Kelly, J. W., Cherep, L. A., Lim, A. F., Doty, T., & Gilber, S. B. (2021). Who are virtual reality headset owners? A survey and comparison of headset owners and non-owners. Proceedings - 2021 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 687–694. https://doi.org/10.1109/VR50410.2021.00095

Kelly, J. W., Hoover, M., Dotty, T. A., Renner, A., Zimmerman, M., Knuth, K., Cherep, L. A., & Gilbert, S. B. (n.d.). Remote research on locomotion interfaces for virtual reality: Replication of a lab-based study on teleporting interfaces.

Kelly, J. W., Ostrander, A. G., Lim, A. F., Cherep, L. A., & Gilbert, S. B. (2020). Teleporting through virtual environments: Effects of path scale and environment scale on spatial updating. IEEE Transactions on Visualization and Computer Graphics, 26(5), 1841–1850. https://doi.org/10.1109/TVCG.2020.2973051

Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk. Conference on Human Factors in Computing Systems - Proceedings, 453–456. https://doi.org/10.1145/1357054.1357127

Lagu, A. V., Landry, S. J., & Yoo, H. sang. (2013). Adaptive function allocation stabilization and a comparison of trigger types and adaptation strategies. International Journal of Industrial Ergonomics, 43(5), 439–449. https://doi.org/10.1016/j.ergon.2013.02.006

LaViola, J. J., Williamson, B. M., Brooks, C., Veazanchin, S., Sottilare, R. A., & Garrity, P. (2015). Using Augmented Reality to Tutor Military Tasks in the Wild. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), December, 1–10

Lim, A. F., Kelly, J. W., Sepich, N. C., Cherep, L. A., Freed, G. C., & Gilbert, S. B. (2020). Rotational Self-motion Cues Improve Spatial Learning when Teleporting in Virtual Environments. Proceedings - SUI 2020: ACM Symposium on Spatial User Interaction, 7. https://doi.org/10.1145/3385959.3418443

MacKenzie, I. S. (2012). Human-computer interaction: An empirical research perspective.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. Behavior Research Methods, 44(1), 1–23. https://doi.org/10.3758/s13428-011-0124-6

Moelson, A., & Hornbæk, K. (2017). Virtual reality studies outside the laboratory. Proceedings of the 23rd Acm Symposium on Virtual Reality Software and Technology, 1–10. https://doi.org/10.1145/3139131.3139141

Mottelson, A., Petersen, G. B., Lilija, K., & Makransky, G. (2021). Conducting Unsupervised Virtual Reality User Studies Online. Frontiers in Virtual Reality, 2, 66. https://doi.org/10.3389/frvir.2021.681482

Muñoz-Saavedra, L., Miró-Amarante, L., & Domínguez-Morales, M. (2020). Augmented and virtual reality evolution and future tendency. Applied Sciences, 10(1), 322. https://doi.org/10.3390/app10010322

Osborne, J. W., & Costello, A. (2008). Best Practices in Exploratory Factor Analysis BT - Best Practices in Quantitative Methods. In Best Practices in Quantitative Methods.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans., 30(3), 286–297. https://doi.org/10.1109/3468.844354

Radkowski, R., Herrema, J., & Oliver, J. (2015). Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty. International Journal of Human-Computer Interaction, 31(5), 337–349. https://doi.org/10.1080/10447318.2014.994194

Ratcliffe, J., Soave, F., Hoover, M., Ortega, F. R., Bryan-Kinns, N., Tokarchuk, L., & Farkhatdinov, I. (2021). Remote XR Studies: Exploring Three Key Challenges of Remote XR Experimentation. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems., 1–4. https://doi.org/10.1145/3411763.3442472

Rickel, J., & Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: perception, cognition, and motor control. Applied Artificial Intelligence, 13(4–5), 343–382. https://doi.org/10.1080/088395199117315

Rouse, W. B. (1976). Adaptive Allocation of Decision Making Responsibility between Supervisor and Computer. In Monitoring Behavior and Supervisory Control (pp. 295–306). Springer US. https://doi.org/10.1007/978-1-4684-2523-9_24

Saffo, D., Di Bartolomeo, S., Yildirim, C., & Dunne, C. (2020). Crowdsourcing Virtual Reality Experiments using VRChat. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, 1–8. https://doi.org/10.1145/3334480.3382829

Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators. https://apps.dtic.mil/sti/citations/ADA057655

Steed, A., Frlston, S., Lopez, M. M., Drummond, J., Pan, Y., & Swapp, D. (2016). An "In the Wild" Experiment on Presence and Embodiment using Consumer Virtual Reality Equipment. IEEE Transactions on Visualization and Computer Graphics, 22(4), 1406–1414. https://doi.org/10.1109/TVCG.2016.2518135

Van Welie, M., Van Der Veer, G. C., & Eliëns, A. (2001). Patterns as Tools for User Interface Design. In Tools for Working with Guidelines (pp. 313–324). Springer.

Westerfield, G., Mitrovic, A., & Billinghurst, M. (2013). Intelligent Augmented Reality Training for Assembly Tasks. International Conference on Artificial Intelligence in Education, 542–551. https://doi.org/10.1007/978-3-642-39112-5_55

Zahabi, M., & Abdul Razak, A. M. (2020). Adaptive virtual reality-based training: a systematic literature review and framework. Virtual Reality, 24(4), 725–752. https://doi.org/10.1007/s10055-020-00434-w

Appendix: IRB Approval Form

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office of Research Ethics
Vice President for Research
2420 Lincoln Way, Suite 202
Ames, Iowa 50014
515 294-4566

Date: 02/09/2021

To: Jonathan W Kelly

From: Office of Research Ethics

Title: **Online study of navigation in virtual reality**

IRB ID: **20-278**

Submission Type: Modification **Review Type**: Expedited

Approval Date: 02/09/2021 **Approval Expiration Date: 07/12/2021**

The project referenced above has received approval from the Institutional Review Board (IRB) at Iowa State University according to the dates shown above. Please refer to the IRB ID number shown above in all correspondence regarding this study.

To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- **Use only the approved study materials** in your research, including the **recruitment materials and informed consent documents that have the IRB approval stamp**.
- **Retain signed informed consent documents for 3 years after the close of the study,** when documented consent is required**.**
- **Obtain IRB approval prior to implementing any changes** to the study or study materials.
- **Promptly inform the IRB of any addition of or change in federal funding for this study.** Approval of the protocol referenced above applies only to funding sources that are specifically identified in the corresponding IRB application.
- **Inform the IRB if the Principal Investigator and/or Supervising Investigator end their role or involvement with the project** with sufficient time to allow an alternate PI/Supervising Investigator to assume oversight responsibility. Projects must have an eligible PI to remain open.
- **Immediately inform the IRB of** (1) **all serious and/or unexpected adverse experiences** involving risks to subjects or others; and (2) **any other unanticipated problems involving risks** to subjects or others.
- IRB approval means that you have met the requirements of federal regulations and ISU policies governing human subjects research. **Approval from other entities may also be needed.** For example, access to data from private records (e.g., student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of

CHAPTER 7. SUPPLEMENTAL RESULTS

This supplemental results section will address differences in performance between the online sample with adaptive feedback, and a lab sample without adaptive feedback. The newly collected online sample then compared to a sample group that completed the study in a lab without the use of adaptive feedback. More information about the results from the lab-based study can be found in the 2020 paper by Kelly et al. This sample used the same procedures, task, and treatments as the online study However, in the lab sample, participants completed 12 trials of each treatment (resulting in a total of 96 trials), while online participants only completed 6 (resulting in 48 trials.) The decrease in trials was designed to increase engagement and participation since online users would not have the additional motivation and social pressure resulting from a researcher's supervision. A total of 37 participants, who were mostly psychology undergraduates at Iowa State University, completed the study in the lab in 2019. Data from four participants were excluded from this data analysis because their data sets were incomplete, leaving 33 samples in the lab group. Eighteen of the participants identified as male, and 14 identified as female. All participants in this group completed the study using an HTC Vive.

The measure of performance used for the following analysis was absolute distance error between the participant's response and the correct answer in the triangle completion task. In the second section, analysis will be conducted to compare the task completion time between the lab and online groups.

Performance

During the triangle completion task, participants were asked to teleport along two legs of a triangle, and then point back at the ground to indicated where the path started (completing the triangle.) The absolute distance error was calculated to be the shortest distance between the given answer (response) and the correct path origin location along the ground plane measured in meters. Absolute distance errors were averaged across all trials in the same treatment group for each participant to account for the differences in number of trials completed by participants in each group, as well as to mitigate the effect of outliers. A natural log transformation was used to remove skewedness from the data (Osborne & Costello, 2008). The resulting data was more normally distributed and had more similar variances than the untransformed data. Although there were small departures from normality, the authors determined that the data was within reasonable limits to continue with the parametric test. A linear mixed-effects model (LMM) was found to best describe the relationship between the response and predictor variables, because it could account for multiple main effects, and incorporate a random effect to account for having multiple data points from each participant (repeated measures).

The main effects incorporated into the model were location (online or lab), condition order (1 to 8), teleporting interface (discordant or partially concordant), triangle path length (10 ft or 40 ft), and VE size (small or big). No interactions, only main effects were found, so the two- and three-way interactions were removed from the model. The LMM was fit using the Restricted Maximum Likelihood (REML) and t-tests using Satterthwaite's method. The results of the LMM showed that teleporting interface, β=-0.361, $t(477)$=-12.113, p<.001, path length, β=1.031, $t(477)$=34.563, p<.001, and VE size, β=-0.193, $t(477)$=-6.473, p<.001, had statistically significant impacts on absolute distance error, while location, β=-0.185, $t(67)$=-1.946, p=.056, had a nearly significant impact. Details about all the predictors used in this analysis can be found in Table 7.1. The order of conditions did not have a statistically significant impact on performance.

The negative beta value associated with location suggests online participants with adaptive feedback had less severe errors than participants in the lab (however not to a level that indicated statistical significance.) This analysis supported the hypothesis that the online study using adaptive feedback would result in similar performance levels as the lab study. And in some cases, the online group even performed better on the triangle completion task than the lab group. However, it is not clear from this test if the online group's performance level was a result of well-designed adaptive feedback, or other factors such as previous VR experience.

Task Completion Time

The duration of every trial was recorded and used as the response variable in an LMM to analyze how task completion time was affected by location, condition order, teleporting interface, triangle path length, and VE size. Again, speed of trial completions was averaged across treatments, and a log transformation was used to eliminate skewedness from the data. The result was a more normally distributed model with the similarities in variance necessary for the statistical test.

No interactions, only main effects were found, so the two-and three-way interactions were removed from the model. The results from the LMM using the Restricted Maximum Likelihood (REML) and t-tests using Satterthwaite's method can be found in Table 7.1. The results showed that the teleporting interface, β=-0.323, $t(477)$=-16.845, p<.001, path length, β=0.173, $t(477)$=9.007, p<.001, and condition order, β=-0.050, $t(477)$=-11.857, p<.001, had statistically significant impacts on task completion time, while location, β=0.125, $t(67)$=-1.990,

p=.051, did not have a statistically significant impact. The VE size did not have a statistically significant impact on the task completion time.

The analysis of the factors influencing task completion time showed that the online group performed the triangle completion task as fast as the lab group, with a trend toward faster mean completion times. This supports the hypothesis that the adaptive feedback provided in the online study is sufficient to replace the need for an instructor in VR research.

Table 7.1. Results of the LMM to test the effects of location, condition order, teleporting interface, triangle path length, and VE size on performance and task completion time.

Independent Var.	**Predictor (Reference Point)**	**Estimate (ß)**	**Std. Error**	**t-value**	**p-value**	
Performance	Intercept	0.875	0.080	10.973	< .001	***
	Location (online)	-0.185	0.095	-1.946	.056	·
	Teleporting Interface (part. Concord.)	-0.361	0.030	-12.113	< .001	***
	Path Length (40ft)	1.031	0.030	34.563	< .001	***
	VE Size (small)	-0.193	0.030	-6.473	< .001	***
	Trial Block Number	-0.001	0.007	-0.115	.908	
Completion Time	Intercept	3.057	0.052	58.435	< .001	***
	Location (online)	-0.125	0.063	-1.990	.051	·
	Teleporting Interface (part. Concord.)	-0.323	0.019	-16.845	< .001	***
	Path Length (40ft)	0.173	0.019	9.007	< .001	***
	VE Size (small)	-0.008	0.019	-0.433	.666	
	Trial Block Number	-0.050	0.004	-11.857	< .001	***

Significance codes: < .001 = ***, .01 = **, .05 = *, .1 = ·

Performance Improvement

The author also analyzed how participant's accuracy on the triangle completion task changed over time in the online study with adaptive feedback compared to the labs study with researcher supervision. To do this, the difference in the average error magnitude was calculated between the first and last trial blocks for each participant. Next, a regression analysis was conducted which used the study location (online or lab) to predict the difference in error magnitude. The analysis showed that the difference in error magnitude for the online group (M=-1.041, SE=0.696) and the lab group (M=0.109, SE=0.760) was not statistically significant (p=0.269) and had a small effect size (d=0.274). The marginal means for each group, along with

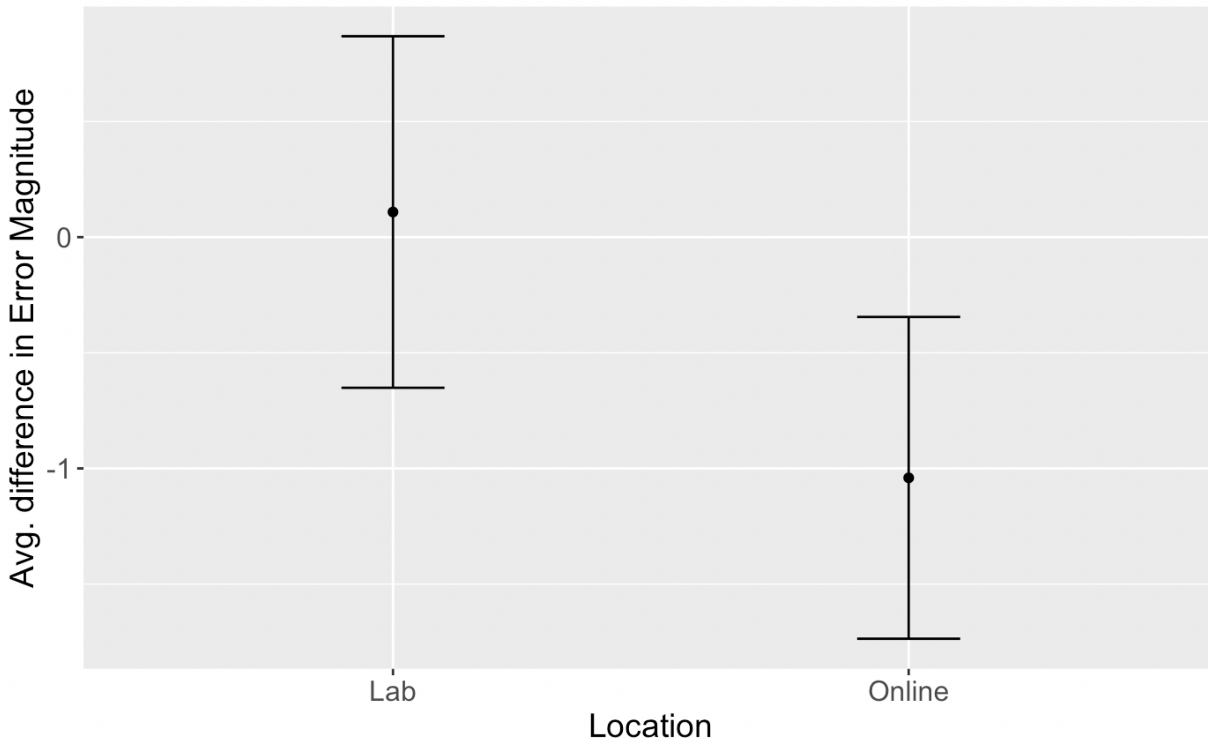


Figure 7.1. Mean difference in error magnitude from the first trial block to the last, grouped by study location (lab vs. online). Error bars represent ±1 standard error of the mean.

their respective standard errors are depicted in Figure 7.1. Although the difference was not statistically significant, the graphs shows that the average difference in error magnitude in the online group was negative, meaning participants in this group tended to improve their performance over time. In the lab group, the mean was positive and close to zero, showing that the lab group tended to maintain a similar level of performance between the beginning and the end of the study. This trend points toward the possibility that participants improved their performance more when using the adaptive feedback interface. This trend is especially interesting considering the online group with adaptive feedback completed half as many trials than the lab group, meaning that their performance improved just as much, or more than the lab group in half as many trials. More samples would need to be collected to see if this trend persists and becomes more significant with larger sample sizes.

Hardware Differences

New questions arose after analyzing the factors that influenced the number of adaptations triggered by participants in the previous chapter. The author theorizes that the “Missed Target” triggers, which happened the most frequently, were influenced by the hardware used by the participants. Since the participants use their own hardware, the types of HMD and controllers used varied greatly. The different types of controllers are particularly of interest because HTC Vive and HTC Vive Pro headsets used controllers with a capacitive touch pad for rotation during the discordant teleporting condition, while the other HMDs used thumb stick. Figure 7.2 shows how frequently “Missed Target” feedback was triggered when using each type of headset. A linear mixed effects model was used to determine whether the controller type, which was dictated by the HMD, and the teleporting interface influenced how often this type of feedback was triggered. The model showed a statistically significant impact of both teleporting interface, β=-2.099, $t(703)$=-6.088, p<.001, and controller type, β=-1.155, $t(51)$=-2.017, =.049. Since the reference points for these factors were the partially concordant interface and the capacitive touchpad controllers, respectively, the negative beta values mean that Missed Target triggers occurred more when using thumbs sticks and the discordant teleporting interface. This is also exemplified by looking at the bars representing the HTC Vive and HTC Vive Pro in Figure 7.2, which represent the HMDs with capacitive touch pads. This result was interesting, because the author originally believed the capacitive touch pad to be more difficult to use for teleporting through the VE. More research in a controlled environment is needed to determine why this occurred. Furthermore, more samples should be obtained to ensure that uneven sample sizes did not influence this result since only 9 participants used capacitive touch pad controllers and 29 used thumb sticks.

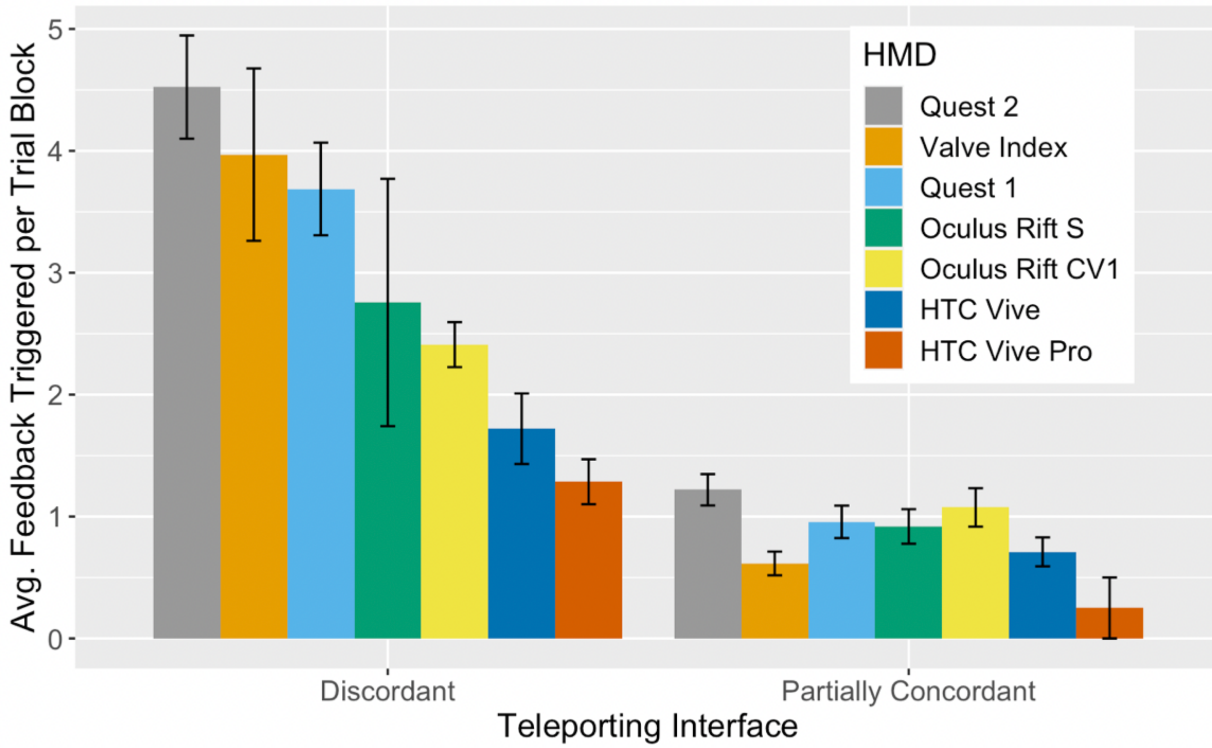


Figure 7.2. Average number of “Missed Target” triggers per trial block grouped by HMD type and teleporting interface. Error bars represent ±1 standard error of the mean.

References

Kelly, J. W., Ostrander, A. G., Lim, A. F., Cherep, L. A., & Gilbert, S. B. (2020). Teleporting through virtual environments: Effects of path scale and environment scale on spatial updating. IEEE Transactions on Visualization and Computer Graphics, 26(5), 1841–1850. https://doi.org/10.1109/TVCG.2020.2973051

Osborne, J. W., & Costello, A. (2008). Best Practices in Exploratory Factor Analysis BT - Best Practices in Quantitative Methods. In Best Practices in Quantitative Methods.

CHAPTER 8. GENERAL CONCLUSIONS

Adaptive automation is a powerful way to assist trainees in learning new skills. The use of adaptive automation can help learners by reallocating processes and information between a human and a computer, thereby reducing the burden on a learner when they are under stress. Learning to use equipment like a XR HMD can be a stressful undertaking, especially when the trainee must simultaneously learn to accomplish a training task with the complex hardware. Traditionally, a trainee would have the assistance of a live trainer to teach them how to use the XR hardware one on one. However, this is not an efficient ratio for training large numbers of people using XR. Additionally, with the advent of the COVID-19 pandemic, more research and training are moving online, making one on one support from a trainer more difficult. This research seeks to ease the burden on the learner by combining adaptive automation and virtual reality technologies to create adaptive training environments in XR that eliminate or reduce the need for training personnel. Specifically, this dissertation addressed three key research questions:

RQ 1. *When are adaptations needed by a leaner in an XR training task?*

RQ 2. *Can an automated adaptive feedback model supplant the need for a live instructor in an XR training system?*

RQ 3. *What specific automated adaptive feedback mechanisms/practices are required for XR training tasks?*

The work presented in this dissertation took initial steps towards answering these questions by first understanding how AR and VR expert trainers interact with trainees to mitigate confusion when learning to use new XR interfaces and hardware. Interviews were conducted with 12 experts to learn about how they identify and react to confusion exhibited by a trainee. An in-depth theme analysis was conducted on the resulting interviews and recommendations for the development of adaptive XR systems were presented. The recommendations based on expert interviews served to answer the first two research questions, and can be summarized as follows:

1. When it comes to adaptation triggers, verbal and physical triggers should be prioritized. This means triggers should use inputs such as verbalizations from the learner, as well as signals that indicate inaction, incorrect action, or incorrect body position.
2. When designing the effects adaptations have on an XR training system, the author recommends physical adaptations be replicated using models and animations in the VE to demonstrate physical tasks and audio or text instructions be used to provide information on procedures.
3. Additionally, the specificity of information provided by adaptations can be increased incrementally to allow the learner to think through a problem on their own and create varying levels of difficulty.

These recommendations were then used to inform the design of adaptations in an online user study. The user study evaluated the use of adaptations to replace the need for a supervisor in during a remote VR research task. The task in question was a triangle completion task, which is used to evaluate a person's ability to navigate through a virtual space. It was essential that this research be conducted remotely, rather than in a lab because of the risk of spreading COVID-19. By successfully implementing adaptive automation to replace researcher supervision in this task, the researchers would save significant time and effort, while participants completed the task without their aide.

The author designed the adaptive feedback for this task by using information from expert interviews, input from researchers, and extensive pilot testing. The resulting adaptations provided text-based feedback which was triggered by temporal, spatial, and operator inputs. The events which elicited adaptive feedback included attempting to teleport to the wrong location, pressing the wrong button when teleporting, pressing the wrong button when giving an answer at the end of the task, failing to teleport within 15 seconds, and failing to give a response to the triangle completion task after 15 seconds. Some of the adaptations also provided increasingly detailed instructions based on the number of times the participant made similar errors.

42 VR owners completed the remote, unsupervised study with adaptive feedback. The results showed that participants were able to complete the study at a rate of 89%, which was equal to the completion rate of a similar study conducted in a lab with supervision from a researcher. The study demonstrated that the adaptive feedback worked as intended because participants made more mistakes that triggered adaptations at the beginning of the study and were able to mitigate their mistakes in later trials. Additionally, participants reported that the feedback was not very intrusive, and that the quantity of feedback they received was appropriate. These results showed that the design of the adaptation triggers was successful. Some participants remarked that the feedback they received was only "somewhat helpful." This indicated that some of the adaptations could be authored differently, or that more adaptations should be authored for situations that were not considered during the initial design.

In response to RQ3, this study showed that the adaptive triggers designed using operator, spatial, and temporal inputs resulted in successful adaptation triggers, and could be extensible to other adaptive training scenarios. However, the results showed that text UI that was used to provide information to learners once an adaptation is triggered was not necessarily ideal. Further research is needed to determine the best way to author the adaptations, whether that means changing the language used, or trying other channels such as audio, haptic, or visual feedback using 3D models.

Overall, this paper addressed all three research questions which were presented at the beginning of the dissertation. The research discussed in this paper supported the use of adaptive automation as a method to supplant supervisors in XR training tasks and made recommendations for developers who seek to employ adaptive automation in future XR applications for training. Future research includes testing more of the recommendations from expert interviews in empirical user studies and testing the effectiveness of adaptive feedback in a variety of VR and AR applications.